Matrix Methods and Differential Equations

A Practical Introduction

Wynand S. Verwoerd

Wynand S. Verwoerd

Matrix Methods And Differential Equations

A Practical Introduction

Matrix Methods and Differential Equations: A Practical Introduction

ISBN 978-87-403-0251-6

Contents

Introduction Mathematical Modelling

1.1 What is a mathematical model?

A model is an artificial object that resembles something from the real world that we want to study. A model has both similarities with and differences from the real system that it represents. There have to be enough similarities that we can convincingly reach conclusions about the real system by studying the model. The differences, on the other hand, are usually necessary to make the model easier to manipulate than the real system.

For example, an architect might construct a miniature model of a building that he is planning. Its purpose would be to look like the real thing, but be small enough to be viewed from all angles and show how various components fit together. Similarly an engineer might make a small scale model of a large machine that he designs, but in this case it may have to be a working model with moving parts so that he can make sure that there is no interference between motions of various parts of the structure. So a model is built for a specific purpose, and cannot usually be expected to give information outside of its design parameters. For example, the architect might not be able to estimate the weight of his building by weighing the model because it is constructed from different materials, the thickness of the wall might not be the same, etc. In building a model, one therefore has to consider carefully just what aspects of the real world are to be included and which may be left out in order to make it easier to work with.

A **mathematical model** usually has at its core an equation or set of equations that represent the relationship between aspects of a real world system. As a simple example, a farmer who plans to buy new shoes for each of his x children, might use the following equation as a model to decide how many shoes he would have to fit into the boot of his car when returning from his shopping expedition:

$$N = 2x$$

This equation represents two relationships: 1) Each child has two feet; and 2) One set of shoes has to be stored in the boot for each child.

The same equation could just as well be used by a bus driver to decide how many passengers he can transport if x represents the number of double seats in his bus. But the underlying relationships in this case are obviously different from the two listed above.. That demonstrates that a mathematical model is more than just an equation; it includes the information about how the equation relates to the real world. On the other hand, the very strength of mathematical modelling is that it removes most of the complexity of the real world when relationships are reduced to equations. Once done, we have at our disposal all the accumulated wisdom of centuries of work by mathematicians, showing how the relationships can be combined and manipulated according to the principles of pure logic. This then leads to conclusions that can once more be applied to the real world.

Even though the shopping model is very simple, it describes not just a single situation such as that of a farmer with 3 children, but rather one that can be applied to different cases of a farmer with any number of children (or a bus with any number of seats). That is a common feature of most useful mathematical models (unlike the architect's building model!). On the other hand it does have its limitations; for example, the farmer would not be able to use it to calculate how many horseshoes he would need for x horses. Of course the model can be extended to cover that case as well, by introducing a new variable that represents the number of feet that each individual has. Whether that extension is sensible, will depend on the use that the model is to be put to. Once more, the equations in a model cannot be taken in isolation. An equation might be perfectly valid but just not applicable to the system that is modelled.

1.2 Using mathematical models

The shoe shopping model was so simple that we could immediately write down a formula that gives the answer which was required from the model. Usually, however, the situation is more complex and there are three distinct stages in using a mathematical model.

Stage 1: Constructing the model. The first step is usually to identify the quantities to be included in the model and define variables to represent them. This is done by considering both what information we have about the real system (the input to the model) and what it is that we need to calculate (the output required from the model). Next, we need to identify the relationships that exist between the inputs and outputs in the real system, and write down equations containing the variables, that accurately represent those relationships. Note that we do not at this stage worry very much about how these relationships will lead to answers to the questions that we put. The main emphasis is to encapsulate the information that we have about the real system into mathematical equations.

Stage 2: Solving the model. In this stage, we forget about the meaning of the variables. An equation is a logical statement that one thing is equal to another. The rules of mathematics tell us what we may legitimately do to combine and manipulate such statements, using just pure logic. The goal is usually to isolate the variable that we need to calculate, on the left hand side of an equation. If we can achieve that, we have found a solution.

Stage 3: Interpreting the solution. In some cases, the mathematics may deliver a single unique solution that tells us all that we need to know. However, it is usually necessary to connect this back to the real system at the very least by correctly identifying and assigning the units of measurement for the calculated quantities, as it is for example meaningless to give a number for a distance that we calculated, if we do not know if this represents millimetres or kilometres.

Moreover, there is often more than one possible solution. In some cases this may legitimately represent different outcomes that are possible in the real system. However, it can also happen that the mathematics delivers additional solutions that are not sensible in the real system; for example, if one of the solutions gives the number of shoes as a negative number. This does not mean that the mathematics is wrong, but merely that the equations that we set up did not include the information that only positive numbers are meaningful for our particular model (they may well be meaningful for another model, which uses the same set of equations). It is part of the interpretation stage to eliminate such superfluous information.

Also, in a complicated model, it often happens that the mathematical solution shows new relationships between variables that we were not aware of during the first stage. This allows us to learn something new about the system, and we then need to spend some effort to translate this back from a mathematical statement to some statement about the real system.

It is clear from this discussion that there is more to modelling than merely mathematics. It is true that in this book and most textbooks, most attention is given to the techniques and methods of mathematics. That is because those methods are universal – they apply to any model that is represented by equations of the type that are discussed. So you might get the impression that mathematical modelling is all about mathematics.

That would be a mistake. It is only the middle stage of the modelling process that is involved with mathematical manipulations! The other two stages often require just as much effort in practice. However, they are different in each particular model, so the only way to learn how to do those is practice. In this book we will work through some examples, but it is important that you try to work out as many problems yourself as you can manage.

Also, in assessment events such as test and exams, students are usually expected not merely to present the mathematical calculations, but also put them in context by clearly defining the variables, relationships, units of measurement and interpretation of your answers in terms of the real system. The same applies to anyone who is using modelling as part of a larger project in some other field of study such as physics, biology, ecology or economics.

1.3 Types of models

Typical modelling applications involve three types of mathematical models.

Algebraic models. The shoe shopping model is a trivial example of an algebraic model; in secondary school algebra you have presumably already dealt with much more complicated problems, including ones where you have to solve a quadratic or other polynomial equation. In this book, we will deal with the case that one has a set of linear equations, containing several variables. While solving small sets by eliminating variables should also be familiar, we will cover more powerful methods by introducing the concept of a matrix and using its properties e.g. to reach conclusions about whether solutions exist, and if so how many there are and how to find them all. Matrix methods can be applied to large systems, and as it turns out have other uses apart from solving linear equations as well. Part I of this book covers this type of model.

Differential equations. When dealing with processes that take place continually in a real system, it is not possible to pin them down in a single number. Instead, one can specify the rate at which they take place. For example, there is no sensible answer to the question "How much does it rain at this moment?" such as there is to the question "How many passengers fit into this bus?". Instead, one could say how much rain falls per time unit, and could then calculate the total for a specific interval. Specifying a rate means that we know the derivative, and if we know how this rate is determined by other factors in the real system, that relationship can be expressed as a differential equation. In this book you will learn how to solve such equations, either single ones or sets of them, in which case both matrices and calculus are used together. Once the differential equation(s) are solved, we are left with algebraic expressions, and so have reduced the problem to an algebraic model once more. Part II of the book deals with solution methods for differential equation models.

Models with uncertainty The outcome from either of the previous two types of model, is typically one or more formulas that could for example be implemented in a spreadsheet program to make predictions of what will or might happen in a real system. However, many real world systems contain uncertainties, either because we have limited knowledge of their properties, or because some quantities undergo random variations that we cannot control. In that case we can incorporate such uncertainties in a model to make predictions about probabilities even if we cannot predict actual numbers. To do this we would need to study the mathematical representation of probabilities and learn to use computer software that calculates the consequences of the uncertainties. That is a logical next step, but falls outside of the domain covered by this book.

1.4 How is this book useful for modelling?

This book is designed as a practical introduction, aimed at readers who are not primarily studying mathematics, but who need to apply mathematical models as tools in another field of study. In practice such readers will most likely use computer software packages to do their serious calculations. However, to understand and make intelligent use of the results, one does need to know where they come from.

A factory manager does not need to be an expert craftsman on every machine in his factory, but he can be much more efficient if he has at least tried to make a simple object on each machine. In that way he can learn what is possible and what is not; and this knowledge is essential when negotiating either with his clients or his workers. In a similar way an advanced computer package is better able to deal with the complications of a large model, but can only be managed successfully by a user who has worked out similar problems in a simpler context. This book should prepare the reader for that role.

In any university library there will be many textbooks that cover either linear algebra, or differential equations, in more detail. These can also be useful as a source of more example problems to work out, and the reader is invited to use this book in conjunction with such more formal mathematical textbooks.

Regarding computer software, three very well-known commercial packages that are often made available on university computer networks for general use, are:

1. *Maple* – see http://www.maplesoft.com for more details.
2. *Mathematica* – see http://www.wolfram.com for details.
3. *MATLAB* – see http://www.mathworks.com for details.

The first two of these are particularly designed as tools for symbolic mathematics on a computer, and very suitable for trying the methods and examples discussed in this book. To help with this, the actual Maple and/or Mathematica instruction that implements a step as discussed in the text, is often given in the book. The syntax of instructions in both programs are similar, but not the same. To avoid confusing duplication, the *Mathematica* syntax is given in the linear algebra section, and the *Maple* syntax in the differential equation section. Users of the other program, or *MATLAB*, will be able to convert to their syntax with a little practice using the online help functions.

Sometimes the computer package actually contains a more powerful instruction that will execute many steps discussed in this book automatically, but for instruction purposes it may be better to follow the steps we suggest. This book is not intended as an instruction manual for the software, but it is hoped that the reader will familiarize him/herself with the use of the software through these examples.

Also, it should serve as a starting point for further exploration – compared to the tedium of trying out ideas by manual calculation, it is so easy to do the same with only a few keystrokes, that it should become part of one's workflow to keep a session of the software system open in one window while you are working through this book in another window. Then one can test your understanding of each statement you read, by immediately constructing a test example in *Maple* or *Mathematica*. One often learns as much from such trials that fail, as from the ones that do work as you expect!

A useful strategy in such trials, is to start from an example that is so simple that you know the exact answer, and first confirm that the syntax you chose for the instruction you type, does give the correct answer. For example, when trying to find the roots of a complicated quadratic equation, one might first enter something like **Solve[x^2-1, x]** (Mathematica syntax) and if this correctly yields $x=\pm 1$, one can then replace the simple quadratic by the one you are really interested in.

The three software packages listed above by no means exhausts the possibilities. Not only are there may other commercial packages, but there are also freeware packages available that can be downloaded from the internet. A fairly comprehensive listing can be obtained by searching the Wikipedia for "comparison of symbolic algebra systems".

The material covered in this book should extend your ability to apply mathematics to practical situations, even though it by no means exhausts the wide range of useful mathematical knowledge. If you enjoy what is offered here, it may well be worth your while to follow this up with more advanced courses as well.

Part 1
Linear Algebra

2 Simultaneous Linear Equations

2.1 Introduction

Consider a simple set of simultaneous equations

$$x + y = 3$$
$$2x + 3y = 1$$

We can use the usual way of elimination to get a solution, **if one exists,** of this set. Firstly, multiply the first equation by –2 and then add them together to get

$$-2x - 2y = -6$$
$$2x + 3y = 1$$

which gives

$$y = -5$$
$$x = 8$$

The solution could also be found by entering into *Mathematica*,

```
Solve[{x + y == 3, 2 x + 3 y == 1}, {x, y}]
```

2.1.1 General remarks

- The equations are **linear** in the variables *x,y*. What this means is that the equations respond proportionately or linearly to changes of *x,y*. It would be more difficult to solve something like

$$x^y = 3$$
$$\cos(x) + e^{-y} = 1$$

- The two equations are **independent**. We would be unable to find a unique solution if we had equations that depended on one another, like

$$x + y = 3$$
$$2x + 2y = 6$$

Instead, we would have many solutions – even infinitely many in this case: for any x there is a corresponding y that would solve the pair of equations.

- The equations are **consistent**. We would be unable to find any solution if we had

$$x + y = 3$$
$$x + y = 1$$

The three cases above are demonstrated graphically by Figure 1 below.

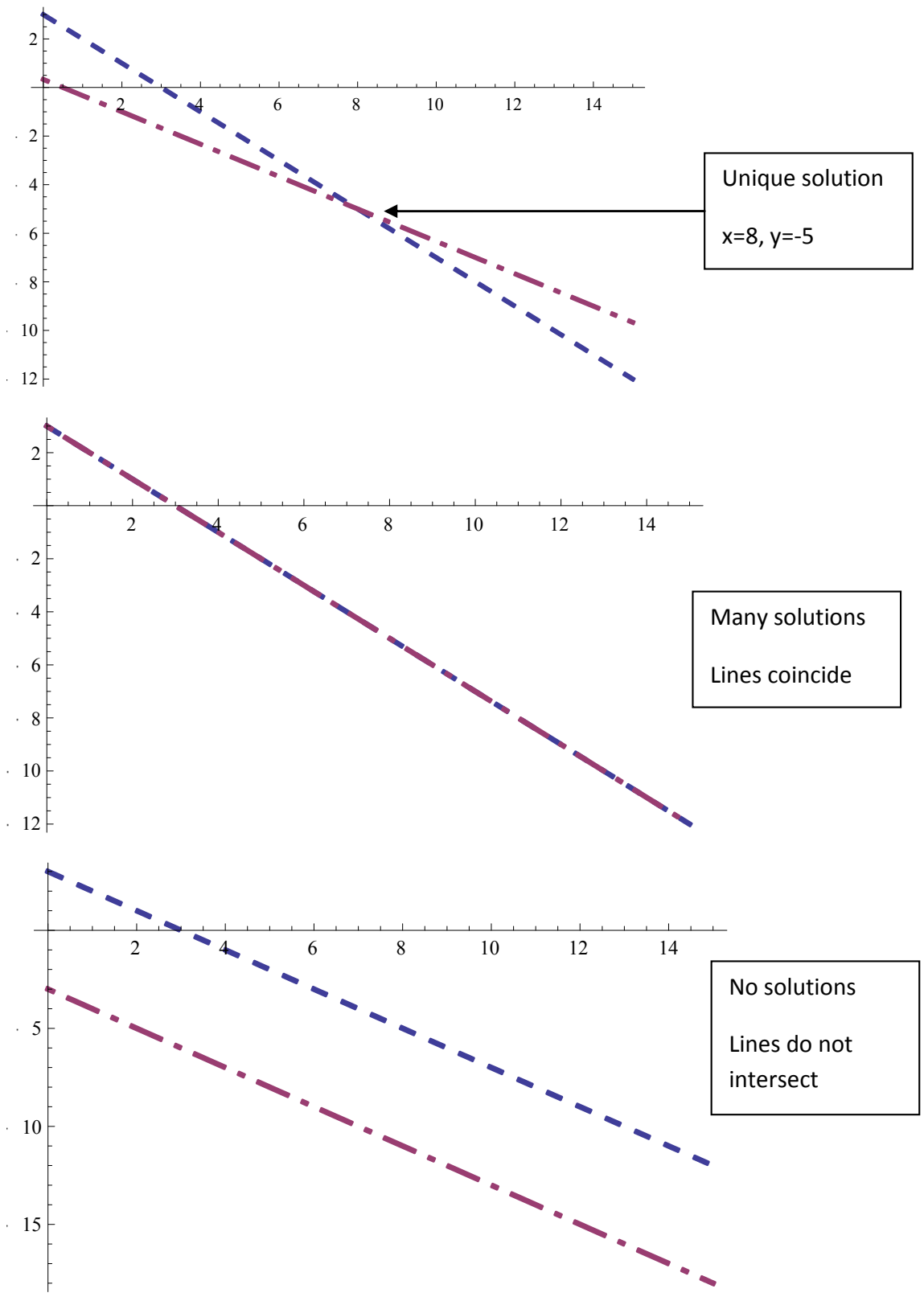


Figure 1: The 3 typical cases of a single solution, multiple solutions, and no solution, illustrated by line plots.

In this case where we had only two equations with two variables, it was easy to find which type of solution we get in each case. But if we have 50 equations with 50 unknown variables, how could we tell if there is one solution, many, or none at all?

To do that we first invent a new way of writing the set of equations, in which we separate the coefficients, which are known numbers, from the variables, that are not known. Each of these sets is collected together in a new mathematical object which we call a **matrix**.

2.2 Matrices

A **matrix** is a square or rectangular array of values or **elements**, written in one of two ways

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \end{bmatrix} \quad \text{or} \quad \left\{ a_{ij} \right\}$$

A $p * q$ matrix has p rows and q columns. p and q are also called the **dimensions** of the matrix. For a square matrix, $p = q$ and their common value is referred to as the dimension of the matrix. The **element**, a_{ij} is that one in the i^{th} row and the j^{th} column.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2q} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{iq} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pj} & \dots & a_{pq} \end{bmatrix}$$

A **diagonal** matrix is square with all non-diagonal elements zero:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5.6 \end{bmatrix}$$

An **identity** matrix I is square with ones on the diagonal and zeros elsewhere. It is also called a **unit** matrix, often shown as I_n to indicate the **dimension** n of the matrix:.

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Similarly we have the **zero** matrix, written as **0**, the matrix where **all** elements are zero.

An **upper triangular** matrix has all elements below the diagonal element equal to zero

$$U = \begin{bmatrix} 1 & 3 & 5.7 \\ 0 & 2 & 4.5 \\ 0 & 0 & -5 \end{bmatrix}$$

There are also **lower triangular** matrices.

Matrices of **one** column are called **column matrices** or **column vectors**. Likewise, those of one row are **row vectors** or **row matrices**. Sometimes a special notation is used to distinguish vectors from other matrices, such as an underlined symbol $\underline{b}$, but usually we do not bother.

We may **transpose** matrices or vectors. That means that the 1st row becomes the 1st column, the 2nd row the 2nd column, etc. The symbol to indicate a transpose is usually a capital superscript T or a prime ‘.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \end{bmatrix}; \qquad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 7 \\ 3 & 6 \end{bmatrix}$$

$$\underline{\upsilon} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}; \qquad \underline{\upsilon}' = \begin{bmatrix} 1 & 5 \end{bmatrix}$$

2.2.1 Rules of arithmetic for matrices

A matrix is a generalisation of the concept of a number. In other words, an ordinary number is just the special case of a matrix with 1 row and 1 column. So, just as we do arithmetic with numbers, we can do arithmetic with matrices if we define the rules for their arithmetic as follows below. Because of this similarity, it is useful to distinguish between numbers and matrices in the way that we write symbols for them. A common method, that is also used in this book, is to represent numbers by lower case letters (a, b, x, y) and matrices by upper case (capital) letters such as A, B, X, Y.

We can **multiply** a matrix **by a scalar** (i.e., by an ordinary number) by multiplying each element in the matrix by that number:

$$A = \begin{bmatrix} 1 & 5 \\ -6 & 9 \end{bmatrix}; \quad -3.7A = \begin{bmatrix} -3.7 & -18.5 \\ 22.2 & -33.3 \end{bmatrix}$$

Addition (or subtraction) of matrices : The matrices must **conform;** that is, they both must have the same number of rows and the same number of columns. (We must distinguish between “conform for addition” and “conform for multiplication”, but more about this later). To add matrices we just add corresponding elements:

$$A = \begin{bmatrix} 1 & 5 & 3 \\ -6 & 9 & 1 \end{bmatrix}; \quad B = \begin{bmatrix} -2 & 1 & 0 \\ 4 & 10 & -11 \end{bmatrix}; \quad A + B = \begin{bmatrix} -1 & 6 & 3 \\ -2 & 19 & -10 \end{bmatrix}$$

The following shows how to enter matrices into Mathematica; the instruction “MatrixForm” (note the capitals) in the second line just displays the matrix in the block form shown above.

```
A = {{1, 5, 3}, {-6, 9, 1}}; B = {{-2, 1, 0}, {4, 10, -11}};
MatrixForm[A + B]
```

Commutative Law $A + B = B + A$

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

For the zero matrix **0** we have

$$A + \mathbf{0} = A$$

For b,c, scalars: $(bc)\,A = b(cA) = c(bA)$

$$1.A = A$$

$$0.A = \mathbf{0}$$

$$(b+c)A = bA + cA$$

$$c(A+B) = cA + cB$$

2.2.2 Multiplying Matrices

There is a special rule for multiplying matrices, constructed in a way that is designed so that we can use it to represent simultaneous equations using matrices. How that happens is shown below:

$$\begin{matrix} A & \cdot & B & = & C \end{matrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 4 \end{bmatrix}\begin{bmatrix} 2 & 3 & -1 & 4 \\ -1 & 0 & -1 & 3 \\ 1 & 2 & 1 & 1 \end{bmatrix}=\begin{bmatrix} 3 & 9 & 0 & 13 \\ 2 & 5 & 5 & 0 \end{bmatrix}$$

The first element of the product, *C*, is the **sum** of the **products** of each element of row 1 of *A*, by the corresponding element of column 1 of *B*:

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}=1\cdot 2+2\cdot(-1)+3\cdot 1=3$$

The elements of the first **row** of *C* are the sums of the products of the first row of *A* and consecutive **columns** of *B*. Similarly, the second row of *C* is obtained by multiplying and summing the second row of *A* with each column of *B*, etc. To remember which way to take the rows and columns, it is just like reading: first from left to right, then top to bottom.

You will see that the number of columns of *A* **must equal** the number of rows of *B*, otherwise they cannot be multiplied.

If the dimensions of *A* is p^*q and *B* is q^*r, then *C* is p^*r.

This product is also called the **dot product** and sometimes represented by putting the symbol "·" between the two matrices, or by just writing the two matrix symbols next to each other. However, do not use the "×" symbol to indicate this matrix product, because there is also another type of matrix product called the "cross product" for which the "×" is used. We will not study cross products in this book.

To perform the multiplication above in *Mathematica*, the instruction is (note the dot!)

```
A = {{1, 2, 3}, {-1, 0, 4}};
B = {{2, 3, -1, 4}, {-1, 0, -1, 3}, {1, 2, 1, 1}};
MatrixForm[A.B]
```

2.2.3 Rules of multiplication

A (BC)	=	(AB) C			
A (B+C)	=	AB + AC			
(B+C) A	=	BA + CA			
k (AB)	=	(kA) B	=	A (kB)	where k is a scalar
A **0**	=	**0**	(note this is not the same as A 0, which is just a scalar multiplication)		
0 A	=	**0**			

All the rules above work just as they would for ordinary numbers.

But in general A B ≠ B A

So, the **order** in which we write a matrix product is vital! Also

$A\ B$	=	0	does NOT imply that either A = **0** or B = **0**
$A\ B$	=	$A\ C$	does NOT imply that $B = C$

Finally, we can show that a product of matrices is transposed as follows:

$(A\ B)^{\mathrm{T}} = B^{\mathrm{T}}\ A^{\mathrm{T}}$

2.3 Applying matrices to simultaneous equations

We can use matrix multiplication to re-express our simultaneous equations:-

$$x + y = 3$$
$$2x + 3y = 1$$

This is just what we get from the following matrix expression by applying our special multiplication rule:

$$\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

In other words, identifying the matrices in this equation as A, X and B respectively, the set of equations becomes just a single simple equation:

$$\mathrm{A\ X = B}$$

The way that multiplying matrices was defined in the previous section, may have appeared rather strange at the time – but we can now see that it makes sense, exactly because it allows a whole set of linear equations to be written as a single equation, containing matrices instead of numbers.

If the symbols in this equation had represented numbers, we could easily solve it by dividing out the A. So do we need to define a division operation for matrices as well? That is not really necessary. Even with numbers, we can avoid using division. We just need to recognize that every number except 0, say x = 4, has an associated number, in this case y = 0.25, called its reciprocal and sometimes written as $x^{-1} = y = 0.25$. Instead of dividing by x, we can just as well multiply by x^{-1}. The two numbers are related by the equation $y\ x = 1$.

Applying the same idea to matrices, we would still be able to solve the matrix equation above if for the known matrix A we are able to find another matrix Q that we call its **inverse**, that satisfies the equation

$$Q\ A = 1$$

If we can find such a matrix, we can just multiply each side of the matrix equation by Q to solve it. The matrix Q has a special name and is called the **inverse** of A and is written A^{-1}. Note that the superscript "-1" is just a notation, it does not mean "take the reciprocal" as it would for a simple number. In particular,

Important: A^{-1} does **not** mean $\left\{a^{-1}{}_{ij}\right\}$ i.e., taking reciprocals of the elements!

In general the inverse is not easy to calculate. It may not even be possible to find an inverse, just as there is a number – zero – which does not have a reciprocal. It turns out that one can only find an inverse if A is a square matrix, and even then not always.

But in the simple case above, $A=\begin{bmatrix}1 & 1\\ 2 & 3\end{bmatrix}$ and it turns out that $A^{-1} = \begin{bmatrix}3 & -1\\ -2 & 1\end{bmatrix}$

We could use *Mathematica* to determine this by entering

```
A = {{1, 1}, {2, 3}}; MatrixForm[Inverse[A]]
```

Check for yourself by manual multiplication that in this case $AA^{-1} = I$ and that $A^{-1}A = I$. The inverse is **unique**, if it exists, and can be used equally well to multiply from either side.

We can now use the inverse above to calculate the values of x and y directly

$$\begin{bmatrix}x\\ y\end{bmatrix}=\begin{bmatrix}3 & -1\\ -2 & 1\end{bmatrix}\begin{bmatrix}3\\ 1\end{bmatrix}=\begin{bmatrix}8\\ -5\end{bmatrix}$$

Now even though inverses in general are difficult to calculate there is a quick method for obtaining an inverse for a **2 x 2** matrix. This is a special case of Cramer's rule used to solve sets of equations.

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}^{-1}=\frac{1}{ad-bc}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}$$

This formula means is that there are three steps to invert a 2x2 matrix:

1. *Swap the diagonal elements*
2. *Change the sign of the off-diagonal elements*
3. *Divide each element by ad-bc*

So for our example the procedure is as follows:

$$\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}^{-1} = \frac{1}{1.3 - 2.1}\begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} = \frac{1}{1}\begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -2 & 1 \end{bmatrix}$$

Now what happens if $ad = bc$? Then we would be attempting to divide by zero and consequently the inverse would not exist. In this case we define the original matrix A to be a **singular** matrix. If the inverse **does** exist we say that the matrix is **non-singular**.

One way that we can get $ad = bc$ is for the second row of the matrix to be a multiple of the first. This occurs when the equations are **not independent** (remember the second case discussed in section 2.1.1?). In this case we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ ka & kb \end{bmatrix}$$

ad – bc = akb – kab = 0

We see that even without actually calculating the inverse matrix, we can make a decision whether an inverse exists by just calculating a single number, the denominator in the formula.

This denominator is called the **determinant**.

Cramers's rule also exists for larger matrices but is computationally very inefficient. Therefore it is helpful especially for large matrices if we can determine before starting, whether the inverse exists. This can be done by defining also for large matrices a single number that characterises the matrix – again, it is called the **determinant** of the matrix.

2.4 Determinants

The determinant of a **2 x 2** matrix is calculated as

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (ad - bc)$$

Note the alternative notation used to indicate a determinant – vertical bars instead of brackets as used for a matrix.

For a 3 x 3 matrix the determinant is found by expanding it in terms of three 2 x 2 determinants. One takes the elements of any row, multiply each by the determinant that remains if you delete the row and column to which the element belongs, and add these up while alternating the arithmetic signs:

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} = aei - ahf - bdi + bgf + cdh - gec$$

You get the same result if you calculate the determinant column-wise rather than row-wise. Thus

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - d\begin{vmatrix} b & c \\ h & i \end{vmatrix} + g\begin{vmatrix} b & c \\ e & f \end{vmatrix} = aei - ahf - dbi + dhc + gbf - gec$$

To check this with *Mathematica*, type **Det[{{a, b, c}, {d, e, f}, {g, h, i}}]**

For a 4 x 4 matrix the same idea is applied once again:

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{vmatrix} = a\begin{vmatrix} f & g & h \\ j & k & l \\ n & o & p \end{vmatrix} - e\begin{vmatrix} b & c & d \\ j & k & l \\ n & o & p \end{vmatrix} + i\begin{vmatrix} b & c & d \\ f & g & h \\ n & o & p \end{vmatrix} - m\begin{vmatrix} b & c & d \\ f & g & h \\ j & k & l \end{vmatrix}$$

The subdeterminants, three 2 x 2 ones for the 3 x 3 main determinant, or four 3 x 3 determinants for the 4 x 4 original, are known as **minors**. For example, the first of these is called the "minor of a". Note how the signs put in front of each term alternate between positive and negative, always starting with a positive sign for the a_{11} element.

2.4.1 Properties of Determinants

We saw that a 2x2 determinant is a sum of twofold products, and a 3x3 determinant a sum of threefold products. Generally, when we have simplified all the minors in working out a large determinant, repeating as many times as necessary as the subdeteminants become smaller in each round, until all determinants have been eliminated, we are left with a sum of terms. For an *n*-dimensional determinant, each term in the sum consists of a product of *n* elements of the matrix. Each of these comes from a different (row,column) combination.

A number of properties follow from fully expanding determinants in that way:

- Interchanging 2 rows (or columns) results in a change of sign of the determinant. *viz.*

$$\det\begin{bmatrix} a & b & c \\ g & h & i \\ d & e & f \end{bmatrix} = ahf - aei - bgf + bdi + cge - cdh$$

  Compare that with the expressions above where the last two rows were interchanged.

- Multiplying any row or column by a scalar constant is the same as multiplying the determinant by that constant; for example

$$\det\begin{bmatrix} a & b & c \\ kd & ke & kf \\ g & h & i \end{bmatrix} = a\begin{vmatrix} ke & kf \\ h & i \end{vmatrix} - b\begin{vmatrix} kd & kf \\ g & i \end{vmatrix} + c\begin{vmatrix} kd & ke \\ g & h \end{vmatrix} = akei - ahkf - bkdi + bgkf + ckdh - cgke$$

Important: Note that this is different from multiplying a matrix by a constant!

- From which it follows that if any two rows or columns are equal, the determinant is **zero**.

  That is because swopping two equal rows (or columns) changes nothing in the determinant, but should change its sign according to the previous bullet point; and the only number that stays the same when you change its sign, is zero.

- Adding a multiple of one row (column) to another row (column) does not change the determinant.

  Test that for yourself on any small determinant. If you want to test it using *Mathematica*, the instruction **Det[A]** will calculate the determinant of matrix A.

To see why this is so, first prove for yourself that the following statement is true:

$$\begin{vmatrix} a_{11} & a_{12}+b_{12} & a_{13} \\ a_{21} & a_{22}+b_{22} & a_{23} \\ a_{31} & a_{32}+b_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & b_{12} & a_{13} \\ a_{21} & b_{22} & a_{23} \\ a_{31} & b_{32} & a_{33} \end{vmatrix}$$

This is easily done by expanding the determinant on the left by taking minors of its second column. Then, consider what happens when the b's are equal to (or multiples of) the corresponding a's, using the previous bullet points.

We saw above that to judge whether a matrix is singular, we need to know whether its determinant is zero. From the above properties of determinants it follows that

The value of a determinant is zero if any of the following is true:

- *All elements of one row (or column) are zero*
- *Two rows (or columns) are identical*
- *Two rows (or columns) are proportional*

Set up some examples, and test these rules for yourself by using *Mathematica* or other software!

For a triangular matrix the determinant is the product of the diagonal elements. You can see this in the **3 x 3** case as follows:

$$\det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei - ahf - dbi + dhc + gbf - gec$$

In the case of a triangular matrix the elements *d,g,h* are equal to zero, then all the terms of the determinant vanish **except** *aei.*

2.4.2 Product properties of determinants

- *The determinant of a product is the product of the determinants:*

$$\det(A \cdot B) = \det(A)\det(B) = \det(B \cdot A)$$

Note that this rule can be applied many times in sequence and we do so later in the similar matrices section.

- *The determinant of the inverse of A is the reciprocal of the determinant of A*

$$\det(A) = \frac{1}{\det(A^{-1})},$$

provided that A^{-1} exists

Notice, from the first property, that although the order of multiplying matrices matters and $A.B \neq B.A$, when taking the determinant this distinction is removed.

2.4.3 Relation of determinants to matrices and equations

The determinant of a matrix is a single number that acts in some ways like a measure of the "magnitude" of the matrix. This comparison should not be taken too literally – a determinant can for example be negative.

But having det(A) = 0 implies that A^{-1} does not exist (just as the reciprocal of a number does not exist if the magnitude of the number is zero) and then A is called **singular**. Conversely, if A has a non-zero determinant then A^{-1} does exist and A is termed **non-singular**.

Consider what happens if we try to solve the equation $A\,X = 0$ (note the zero vector on the right hand side). If A has an inverse, we can multiply the equation by A^{-1} and find that $X = 0$. i.e. the whole set of variables are all zero. This is called the **trivial** solution and is usually not meaningful or interesting in a mathematical model. So

If, and only if, det(A) = 0, the homogeneous equation

$$A\,X = 0$$

has a solution other than X = 0.

This is rather like the situation where we have two numbers, p and q, multiplied together and the result is zero:

$$p\,q = 0$$

What this tells us is that either p is zero, or q is zero, or possibly both p and q are zero. Suppose that (like A in the matrix equation) p is known, but we do not know the value of q. If we know that p is not zero, then q *has* to be zero. But if we know that p is zero, then it is possible (although not certain) that q is non-zero. The value of p here plays a similar role to that of det(A) in the matrix equation.

2.5 Inverting a Matrix by Elementary Row Operations

The recipe given above for calculating a determinant is in principle straightforward, but can become very tedious for large matrices, and it becomes worse for finding the inverse. A more efficient method is based on what is called **elementary row operations**.

There are three types of elementary row operations on a matrix, corresponding to the operations that we apply to sets of equations if we want to eliminate variables:

- Add a multiple of one row to another row
- Multiply a row by a non-zero scalar
- Interchange two rows

Each of these operations can be done "manually", but can also be performed by multiplying a given matrix by some special invertible matrix, as given below.

Adding a multiple of one row to another:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ kd+g & ke+h & kf+i \end{bmatrix}$$

So, putting k in row 3, column 2 of the identity matrix, and multiplying, has added k times row 2 to row 3 of the matrix.

The other operations are similarly produced from a modified identity matrix.

Multiplying a row by a non-zero scalar:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ kd & ke & kf \\ g & h & i \end{bmatrix}$$

Interchanging two rows:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ g & h & i \\ d & e & f \end{bmatrix}$$

The appropriate matrices shown above for each operation each has an inverse. To show that they do, all we have to do is to show that they have a non-zero determinant. Try it!

For matrix dimensions larger than 3, we can similarly modify the appropriate larger identity matrices to produce all elementary row operations.

Suppose we can multiply a given matrix progressively by additional ones of these elementary row operation matrices, until it is reduced to a unit matrix:

$$\underbrace{\begin{bmatrix\end{bmatrix}\begin{bmatrix\end{bmatrix}\begin{bmatrix\end{bmatrix}\cdots\begin{bmatrix\end{bmatrix}}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Then we see that collecting all of those row operation matrices together and multiplying them, **they form the inverse of the starting matrix**. Since each one individually is invertible, so is their product. That is because the determinant of the product equals the product of the determinants, so it will not be zero if none of the determinants of the individual matrices are zero.

Calculating the inverse of a matrix in this way, relies on whether a set of row operations can be found to reduce it to a unit matrix. The next goal is to design a systematic way to find this set of row operations for a given matrix or set of equations.

2.6 Solving Equations by Elementary Row Operations

Consider the set of equations

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

Using the collective row operation matrix described above, we can multiply on both sides from the left:

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}\begin{bmatrix} p \\ q \\ r \end{bmatrix}$$

But the left-hand side would become the identity matrix (remember $A.A^{-1}=I$) so we would be left with the vector of unknowns on the left hand side and the right hand side would be the solutions, which are obtained by **premultiplying** (order is important!) the right-hand side by the inverse of A. That is the method of solution that was already shown above in section 2.3, except that here we now have a better method to calculate the inverse.

But instead of actually multiplying the inverse on the right, we can just as well simply perform the same elementary row operations on the right.

A shortcut to do this is to set up a new matrix where we can do it all together. We **augment** the original matrix A with the original column vector of constants to create an augmented matrix

$$\begin{bmatrix} a & b & c & p \\ d & e & f & q \\ g & h & i & r \end{bmatrix}$$

and then we perform elementary row operations on the whole (now non-square) augmented matrix to get the identity matrix (**3 x 3)** in the left hand part of the matrix and the solutions in the right hand side of the matrix.

$$\begin{bmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \end{bmatrix}$$

These ideas show that if the matrix of coefficients is non-singular, we can use elementary row operations to invert it and hence to solve equations that have non-zero terms p,q and r on the right hand side. First have a look at some examples.

Example

If we have the equations

$$\begin{aligned} y + 2z &= 3 \\ x + 2y + z &= 1 \\ x + y &= 0 \end{aligned}$$

then from this we get the augmented matrix

$$\begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

To solve this we perform the following row operations. The strategy is to work systematically towards a unit matrix, starting from the top left. The first stage aims to get the lower triangle of the matrix in shape; i.e., 1's on the diagonal and 0's below it. We do that by working through the columns one by one from the left. We first get a 1 in the top left corner; then use row subtraction to get all the other elements in the first column to zero.

Interchange row 3 and row 1 which gives

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Subtract row 1 from row 2 (ie. row 2 = row 2 – row 1)

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}$$

Column one is now in order, so we work on the second column. Make its diagonal element 1 (e.g., by dividing the 2nd row by the value currently in that position, if it is not 1 already). Then, we can use rows 2 and 3 to make other values in column 2 zero by subtraction. Note that we cannot use row 1 for that purpose, because if we do the zeroes we have already produced in column 1 will be destroyed. In this case, let us subtract row 2 from row 3 (ie. row 3 = row 3 – row 2)

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

That completes the first stage: the lower triangle is done. In the second stage we produce zeroes in the upper triangle elements, column by column. To correct column 2 subtract row 2 from row 1 (ie. row 1 = row 1 – row 2)

$$\begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Moving on to column 3, add row 3 to row 1 (ie. row 1 = row 1 + row 3)

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Finally, subtracting row 3 from row 2 removes the last offending non-zero element:

$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Notice how, in clearing the top triangle, we selected row 2 to clear values from column 2, and row 3 to clear values from column 3. Doing this ensures that the preceding columns already cleared will not be affected by the row addition or subtraction because the selected row has only zeroes in those columns. That was ensured in the first stage when the lower triangle was cleared.

The final form of the augmented matrix gives us the solution for the three unknown *x, y* and *z* as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Check the answer by entering the following into *Mathematica*; note the double equal signs that are needed to enter an equation (as opposed to setting the value of a variable with a single =)

```
Solve[{y + 2 z == 3, x + 2 y + z == 1, x + y == 0}, {x, y, z}]
```

Another example – finding a determinant at the same time

Start with the set of equations:

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 3 \\ -2x_1 + x_3 &= -8 \\ x_1 + x_2 + 2x_3 &= 0 \end{aligned}$$

From this we get the augmented matrix

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ -2 & 0 & 1 & -8 \\ 1 & 1 & 2 & 0 \end{bmatrix}$$

To fix column 1, subtract row 1 from row 3 & add row 1 twice to row 2 (no change to determinant)

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 4 & 3 & -2 \\ 0 & -1 & 1 & -3 \end{bmatrix}$$

For column 2, notice that there is a convenient -1 in row 3. So switch rows 2 & 3 (which changes the sign of the determinant)

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & -1 & 1 & -3 \\ 0 & 4 & 3 & -2 \end{bmatrix}$$

Change sign of row 2 (ie multiply row by –1 ∴ multiply determinant by –1)

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 4 & 3 & -2 \end{bmatrix}$$

Subtract 4 times row 2 from row 3 (determinant unchanged)

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 7 & -14 \end{bmatrix}$$

Divide row 3 by 7, thereby dividing the determinant by 7

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

So far, the lower triangle is 0, and the diagonal 1.

Now get the upper triangle 0 as well. To simplify the writing, the row operations used are indicated in a short-hand way next to each matrix as shown below:

$$\begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \begin{matrix} R1-R3 \\ R2+R3 \end{matrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \quad R1-2*R2$$

The determinant does not change in any of these row operations. The resulting equation is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}$$

and the determinant of the resultant matrix (equivalent to the set of transformations we carried out) is (from steps 2, 4 and 5 respectively)

(new det) = (-1) (-1) (1/7) (old det) = 1/7 (old det)

which is 1, because the new matrix that we formed is just the identity matrix, which has a determinant of 1. From this we conclude that determinant of the original matrix is 7. Check this out for yourself by calculating it directly from the original matrix. So if we did not know the original determinant, we could find it by just taking the reciprocal of the simple product of changes produced by the elementary row operations.

2.6.1 Comments

In these examples, we have not bothered to explicitly calculate the inverse, but as was shown above it could in principle be done by simply writing down the matrix that produces each row operation and multiplying them together. That is tedious by hand, but easily performed by a computer program.

If, in the course of our systematic process the augmented matrix ever assumes a form where a row or column of the coefficient part is all zeroes, that means that the determinant is zero. We then have to stop; it proves that the coefficient matrix is singular, so the method of solution described so far does not apply. The same is true if any of the other tests for a zero determinant listed in the box of section2.4.1 are satisfied. The next section explains how to handle such cases. To summarise:

Non-homogeneous equations with a non-singular coefficient matrix can be solved by applying the following row operations to the augmented matrix, in such a way that the coefficient part is reduced to the identity matrix:

- *Add a multiple of one row to another row*
- *Multiply a row by a non-zero scalar*
- *Interchange two rows*

Row operations of the last two kinds multiply the determinant by a factor. The determinant of the coefficient matrix is calculated by multiplying all these factors together, and taking the reciprocal.

2.7 Homogeneous and Non-homogeneous equations

We have seen that it is important to know if the coefficient matrix A in a matrix equation of the form

$$A\,X = \mathbf{0}$$

is singular, because only if it is, will there actually be a non-trivial solution where all x's are not zero. This form of equation, is called a **homogeneous equation** because the right hand side vector is the zero vector. But we have not yet shown how to get the solution or multiple solutions, if they exist.

The examples above have shown on the other hand, how to obtain the solution of a **non-homogeneous equation** (i.e., one with a non-zero right hand side vector)

$$A\,X = B$$

in the case that A is non-singular, by using row operations that in effect find the inverse of A. Because the inverse is a unique matrix, in this case we also only have a single solution. In this case the matrix method that we have introduced is only a streamlined version of the variable elimination method covered in more elementary algebra courses.

So far, the situations are just the reverse of each other: for homogeneous equations we want the coefficient matrix A to be singular, and for the non-homogenous case we want it to be non-singular. But suppose we are given a non-homogenous set, and it happens that A is singular?

In this case, there are multiple solutions to the set of equations. To find these, we need to solve both the non-homogeneous equation and the homogenous equation that has the same coefficient matrix.

The reason for this is as follows. Suppose we have a vector X that solves the non-homogeneous equation, and another one Y that solves the homogeneous equation:

$$A\,X = B \quad ; \quad A\,Y = 0$$

Then we can add any multiple of Y to X, and it will also be a solution of the non-homogenous equation, because

$$A(X + c\,Y) = A\,X + A(c\,Y) = A\,X + c\,A\,Y = A\,X + 0 = B$$

And similarly, if the homogenous equation has multiple solutions, we can obviously add multiples of all of them to X and will still get a solution to the non-homogenous equation.

It can be proven that all possible solutions of the non-homogeneous equation can be obtained in this way. The procedure to find all non-homogenous solutions reduces to the following logical steps:

- *If the coefficient matrix is non-singular, there is only one solution and it is found using elementary row operations.*
- *If the coefficient matrix is singular, find just one solution (any one) – this is called the particular solution.*
- *In this case, also solve the homogeneous equation with the same coefficient matrix – all solutions must be found.*
- *The complete solution of the non-homogeneous equation is then obtained by adding multiples of all the homogenous solution vectors to the particular solution vector.*

The implication of the last step is that the complete solution will still contain unknown coefficients in front of the homogenous solution vectors, and so in this case no unique solution has been found.

However, the solution found in this way will represent all of the information that was present in the set of equations that we started with. We might for example start with 100 equations in 100 unknowns, and because of dependencies between the equations of which we were unaware, end up with a final solution that consists of the particular solution plus two homogeneous solution vectors. Then we have reduced the number of unknowns from 100 to 2, which is a big improvement in our knowledge about the system that we are modelling, even if it is not completely solved. In some cases there may be other constraints, such as that all the variables must be positive, that can be used to narrow down the solution further.

It all boils down to the fact that we still need to find a way to solve a homogeneous matrix equation. To do this, we use the method of reducing the coefficient matrix to **row echelon form**, which is simply a generalisation of the reduction to a unit matrix that was illustrated above for non-homogenous equations.

2.7.1 Row Echelon Form

If by row operations you reduce a matrix to upper triangular form such that

- *Any zero rows are below any non-zero rows*
- *In each non-zero row the first value is unity –"leading 1"*
- *In each non-zero row (after the first) the leading 1 is to the right of the leading 1 above it*

Then you have row echelon form abbreviated as ***REF****.*

For a non-homogenous matrix equation this reduction is performed on the augmented matrix, for a homogenous equation only the coefficient matrix needs to be included.

The identity matrix to which we reduced the left hand side previously, is a special case of this. Check for yourself that any identity matrix indeed has REF according to the definition above. In cases where the coefficient matrix is non-singular, obtaining REF is just like using elementary row operations to invert the matrix, only we stop halfway (after completing stage one) because we do not bother to make elements in the upper triangle zero.

However, when the matrix is singular, it is impossible to reach a unit matrix form, but it is still possible to reach REF by elementary row operations. Some properties of the solutions can be recognized directly from the form of the REF matrix:

- *If there are zero rows in the REF, their number gives the number of* ***row-dependencies*** *and leads to multiple solutions.*
- *Inconsistent equations are also displayed, i.e. a row which is zero in the coefficient matrix part but has a non-zero element in the augmented column.*

These remarks become clearer by looking at some examples.

Examples

Two simple examples of REF are the unit matrix I and the zero matrix.

But all the following matrices also have REF:

$$\begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 & 0 & 1 & 3 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2.7.2 Strategy for obtaining REF

We start with the first row of the given matrix and work our way **downwards**. In the first row we usually just divide as needed to get the (1,1) element equal to 1. If this element happens to be 0, we first swap the row 1 with any other row that has a non-zero first element.

For each row, the first step is to get all elements below the leading 1 in the current row, zero (by subtracting a multiple of the current row).

Then we move to the next row and make the leading value a 1 by dividing the row by the appropriate number.

In getting elements zeroed, we always use the current row to subtract from the others below it. In that way, because of the developing pattern of zeroes, it is usually easy to systematically cancel the unwanted non-zero values without disturbing the ones that have already been done.

Throughout the process, we watch out for opportunities to speed things up by exchanging rows that will fit into the desired pattern, especially moving rows full of zeroes downwards.

As with any strategy game, it is best learnt by playing – the examples below should give you a start.

2.7.3 Constructing solutions from the REF

Even though not all matrices can be inverted, REF can always be achieved.

For a non-singular matrix, the REF form will have zeroes in the lower triangle, 1's on the diagonal, and (mostly) non-zero elements in the upper triangle. In this case one could continue working to reduce the upper triangle elements to zero as well, but it is a waste of time since the solution can easily be constructed from the REF form itself.

A key idea in constructing solutions is to realise that when the solution is not unique, one can **freely choose** the values of some of the unknown variables, while the remaining ones will be determined by the equations. The strategy described below shows how to decide which variables can be chosen. Also, it simplifies our work by choosing variables to be 0 wherever possible, and if not we take them as 1. Apart from reducing work this helps to avoid arithmetic errors.

A non-homogenous set of equations.

Once you have the REF, the solution to the set of equations is constructed by substituting the values that you choose for variables, always starting at the last row that is non-zero and working your way **upwards**.

In the non-singular case, the last equation will have only the single non-zero coefficient 1 and there is no choice, the last unknown will have to be equal to whatever value occupies the corresponding augmented element.

However, in a singular case one ignores any zero equations at the bottom, and the first non-zero one encountered will then have two or more non-zero coefficients. In this case all but one of the variables that belong to those non-zero coefficients can be freely chosen.

The strategy to follow is that we start from the last unknown, and put all but the first one that occurs in the equation equal to 0. That way one variable can be solved. Next, we move up to the previous equation, and put that and all the zeroes that were chosen into it. This gives the next variable value, and continuing upwards in this fashion all the variables will then be solved.

In this way, one will have constructed either the **only** solution for the non-singular case, or the **particular** solution for the singular case.

A homogenous set of equations.

The strategy for the homogenous equation is pretty similar, with one difference. In the non-homogenous case the strategy was to choose as many variables to be 0 as you can. That will not work in the homogenous case – because the right hand sides are zero anyway, you will simply end up with the trivial solution. So here, among those variables that can be freely chosen, one chooses one at a time to be 1, and the rest 0. If there are several variables to be freely chosen, giving each of them a chance to be 1 will give you one more solution. So in this way all the multiple solutions of the homogeneous equation are constructed.

Example

Consider the non-homogeneous equations:

$$\begin{aligned} x_1 + x_2 - x_3 + 2x_4 &= 3 \\ 3x_1 + 2x_2 + x_3 - x_4 &= 5 \\ 2x_1 + x_2 + 2x_3 - 3x_4 &= 2 \\ 4x_1 + 3x_2 + x_4 &= 8 \end{aligned}$$

The augmented matrix is given by

$$\left[\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 3 \\ 3 & 2 & 1 & -1 & 5 \\ 2 & 1 & 2 & -3 & 2 \\ 4 & 3 & 0 & 1 & 8 \end{array}\right]$$

Element (1,1) is already 1 so we do not need to do anything to the first row. The operations needed for each row to zero elements in the first column are shown on the right:

$$\left[\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 3 \\ 0 & -1 & 4 & -7 & -4 \\ 0 & -1 & 4 & -7 & -4 \\ 0 & -1 & 4 & -7 & -4 \end{array}\right] \qquad \begin{array}{l} \text{R2} - 3*\text{R1} \\ \text{R3} - 2*\text{R1} \\ \text{R4} - 4*\text{R1} \end{array}$$

Because the last 3 rows have become equal, REF is obtained by simply subtracting row 2 from rows 3 and 4:

$$\left[\begin{array}{cccc|c} 1 & 1 & -1 & 2 & 3 \\ 0 & 1 & -4 & 7 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{array}{l} \text{R3} - \text{R2} \\ \text{R4} - \text{R2} \end{array}$$

The *Mathematica* instruction for reduction to row echelon form, is RowReduce. To perform this on the example matrix above, enter

```
MatrixForm[ RowReduce[{{1, 1, -1, 2, 3}, {3, 2, 1, -1, 5}, {2, 1, 2, -3, 2}, {4, 3, 0, 1, 8}}]]
```

The result returned, is $\left[\begin{array}{ccccc} 1 & 0 & 3 & -5 & -1 \\ 0 & 1 & -4 & 7 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

Can you explain how that relates to the result obtained above?

The non-homogenous solution: The last two rows in the REF matrix are zero rows showing that the coefficient matrix is singular and there are multiple solutions. Ignoring the last two rows, in row 2, variables x_3 and x_4 can be freely chosen, and we make them 0. That gives $x_2 = 4$ as the solution for the equation corresponding to row 2. Putting these values into row 1 yields $x_1 = -1$. So the particular solution is

$$X_p = \begin{bmatrix} -1 \\ 4 \\ 0 \\ 0 \end{bmatrix}$$

The homogenous solution: Now the augmentation column is taken as zero. Taking row 2 again, it is clear that we cannot take $x_3 = 0$ and $x_4 = 0$ in this case because that will lead to $x_2 = 0$ and hence from the first row $x_1 = 0$ as well.

So our first option is to take $x_4 = 1$, $x_3 = 0$, which gives $x_2 = -7$ when we substitute those in row 2. Then putting those values into row 1 gives $x_1 = 5$.

The other possibility is to take $x_4 = 0$, $x_3 = 1$, which gives $x_2 = 4$ when we substitute those in row 2. Then putting those values into row 1 gives $x_1 = -3$.

These two solutions of the homogenous equation are combined with the particular solution to give the complete solution

$$X = \begin{bmatrix} -1 \\ 4 \\ 0 \\ 0 \end{bmatrix} + c_1 \begin{bmatrix} -3 \\ 4 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 5 \\ -7 \\ 0 \\ 1 \end{bmatrix}$$

This is the best we can do to give a solution of the given equations. It still contains two unknowns; but we are guaranteed that whatever values are taken for c_1 and c_2, the expression above will satisfy all the equations in the set.

If more information is known about the solution, for example that all the x's have to be positive, the possible values for the *c*'s can be narrowed down further. For example, looking at the last two elements of X we see that $c_1 > 0$ and $c_2 > 0$ in this case, otherwise x_3 or x_4 would be negative. Additional constraints on the *c*'s can be obtained from the first two elements of X – try to obtain those yourself!

3 Matrices in Geometry

So far, we have considered matrices and vectors merely as collections of numbers, derived from sets of linear equations, and convenient to manipulate for solving the equations.

However, there is far more to the concept than that. In geometry, vectors and matrices have a very direct physical interpretation, and understanding that is helpful in forming meaningful mental pictures of many of the more abstract ideas that are used.

The most basic example of a geometrical vector is the position vector. The position of a point on a plane is specified by defining a coordinate system, i.e. mutually perpendicular X- and Y-axes. The pair of numbers that specify the position of the point relative to these axes, i.e. the x- and y-coordinates of the point, is its **position vector**. The same idea can be applied to a point in 3 dimensional space, where the position vector has 3 elements.

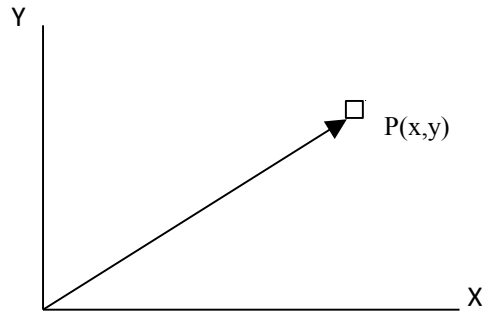
There are many other examples of physically measurable quantities that are also mathematically represented as vectors, like velocity, acceleration, electric and magnetic fields, etc.

So how do matrices fit into the picture? We already saw that elementary row operations on a matrix (or vector) can be performed by multiplying it by an appropriate special matrix. This is an example where we take any given matrix as input, multiply it with a specially chosen matrix, and obtain a new matrix as output. This is expressed by saying that the input matrix is **transformed** to the output matrix. The same idea applies in geometry: operations that have a physical meaning, like rotating a position vector by a certain angle around some axis, can be achieved by multiplying the vector by an appropriate matrix. We now look at some examples to see how such matrices are constructed.

3.1 Reflection

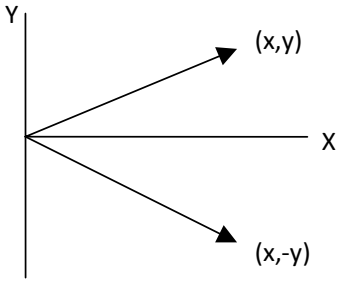
If P is a point in 3-space with coordinates (x,y,z) then $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is the vector of coordinates.

In the two dimensional plane or 2-space we can draw this as



The matrix that produces a reflection of the vector across the X-axis is

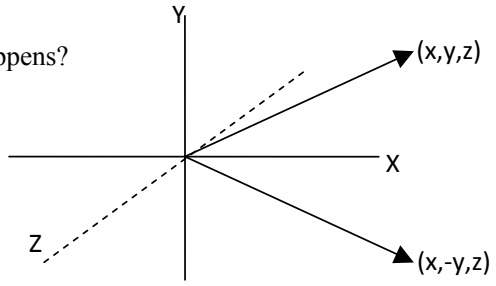
$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$



because

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}$$

In 3 dimensions, if we had $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ what happens?



We get a reflection in the plane $y = 0$.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ -y \\ z \end{bmatrix}$$

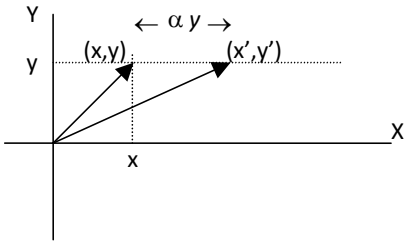
To produce reflections in the X- or Z-planes, we similarly only need to put a negative sign on the corresponding diagonal elements of the unit matrix.

Check for yourselves that these operations produce the correct reflections as for the above example.

3.2 Shear

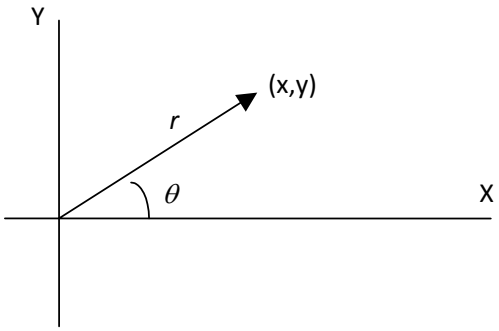
Putting a non-zero element into an off-diagonal position of the unit matrix, produces a shear distortion of the position vector, such as would result if you exert a sideways force on a cubical blob of jelly that is fixed at its bottom:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + \alpha y \\ y \end{bmatrix}$$



3.3 Plane Rotation

In standard polar coordinates we have that



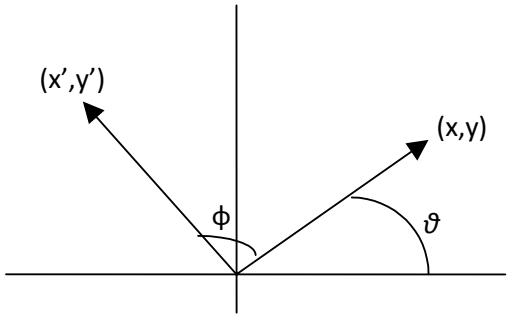
$$r = \sqrt{x^2 + y^2}\ , \ \& \ \ \tan\theta = \frac{y}{x}$$

We can describe the point in terms of $(r,\ \theta)$ just as easily as (x,y). From the diagram, where r is the hypotenuse of a right-angled triangle:

$$x = r\cos\theta$$
$$y = r\sin\theta$$

Consider now the following situation where we have rotated this point around the origin by a further angle φ so that now we have

$$x' = r\cos(\theta + \phi)$$
$$y' = r\sin(\theta + \phi)$$



From trigonometry we have that

$$\cos(\theta + \phi) = \cos\theta.\cos\phi - \sin\theta.\sin\phi$$
$$\sin(\theta + \phi) = \sin\theta.\cos\phi + \cos\theta.\sin\phi$$

so we can write

$$x' = r\cos\theta.\cos\phi - r\sin\theta.\sin\phi$$
$$y' = r\sin\theta.\cos\phi + r\cos\theta.\sin\phi$$

which becomes

$$x' = x.\cos\phi - y.\sin\phi$$
$$y' = y.\cos\phi + x.\sin\phi$$

This we can write in matrix form as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

So if we wish to rotate the vector through an angle of φ=30° the matrix becomes

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

And the position vector for the point (1,0) becomes

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

Draw the triangle yourself, formed by rotating the line on the X-axis to (1,0), through 30° anti-clockwise and check that this is correct.

In 3-space, a rotation counter clockwise through the angle φ about the z-axis is given by

$$\begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Consider now what happens if we perform a sequence of rotations, say firstly an angle of φ and then again by the same angle φ.

The first rotation is specified by

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

and the second one is

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$= \begin{bmatrix} \cos^2\phi - \sin^2\phi & -2\sin\phi.\cos\phi \\ 2\sin\phi.\cos\phi & \cos^2\phi - \sin^2\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos 2\phi & -\sin 2\phi \\ \sin 2\phi & \cos 2\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

So a rotation by angle ϕ, followed by another rotation through the same angle ϕ, is equivalent to a single rotation through an angle (2ϕ). Of course this is what common sense tells us, but we see how this followed from the mathematics of matrix products and trigonometrical formulas as well.

In general we have for n rotations each by an angle φ

$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}^n = \begin{bmatrix} \cos n\phi & -\sin n\phi \\ \sin n\phi & \cos n\phi \end{bmatrix}$$

To test these ideas in *Mathematica*, note that you do not have to enter the matrices manually, they come predefined in the instruction **RotationMatrix[θ]** for 2 dimensions.

For 3 dimensions, one also needs to specify the axis around which the rotation takes place, by giving a vector *a* pointing along the axis direction, e.g. *a* = (0,1,0) for a rotation around the y-axis. The syntax is **RotationMatrix[θ,{0,1,0}]** – try it for yourself.

When multiplying matrices, remember to use the dot symbol (.) in *Mathematica*. Also, to raise a matrix *A* to a power *n* using matrix products, one needs to use **MatrixPower[A,n]** because just entering *A*^*n* will simply raise each element of *A* to the power *n*, which is not the same thing!

As an example, we can test if the equation above works for *n* = 3. We enter

```
Simplify[MatrixPower[RotationMatrix[\[Theta]], 3]] // MatrixForm
```

and get the result $\begin{pmatrix} \text{Cos}[3\theta] & -\text{Sin}[3\theta] \\ \text{Sin}[3\theta] & \text{Cos}[3\theta] \end{pmatrix}$.

Here, the "//MatrixForm" at the end is just another way to get the answer formatted as a matrix. The "Simplify" instruction performs the function of applying trigonometric simplifications; without it we would just get polynomials in cos θ and sin θ, as was shown above for the case of $n = 2$. If you try this for other values of n, you will notice that sometimes *Mathematica* thinks another expression is simpler than cos $n\theta$ or sin $n\theta$; you can always confirm that the identity holds by e.g. entering the instruction **Simplify[MatrixPower[RotationMatrix[θ], 5] = = RotationMatrix[5 θ]]**. Note the double equal sign to signify a logical equality, which *Mathematica* will test and return "True" if both sides are equal.

Question: What happens when we raise the shear matrix $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ to the n-th power?

3.4 Orthogonal and orthonormal vectors

If we have two vectors that are at right angles to each other, i.e. they have an angle between them equal to 90° or $\pi/2$ radians, these vectors are said to be **orthogonal**. In this case the matrix that would produce an orthogonal vector starting from an arbitrary position vector, is just the special case where it produces a rotation by $\pi/2$:

$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos\frac{\pi}{2} & -\sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & \cos\frac{\pi}{2} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

so that the point (x,y) becomes the point $(-y,x)$ when multiplied by this matrix.

To multiply these vectors together, using the rules of matrix multiplication, the first vector needs to be converted to a row vector (i.e., we take its transpose) and we get

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} -y \\ x \end{bmatrix} = -xy + yx = 0$$

From this we get an important point. Any two non-trivial vectors whose product is zero are **orthogonal.**

This is an example of how matrices that are each individually non-zero, can still multiply to get zero – something that cannot happen with numbers!

In two dimensions, we can only have 2 vectors at a time that are orthogonal to each other (e.g., ones pointing along the X and Y axes) but in three dimensions we could have a set of 3 vectors that are mutually orthogonal, like the ones along the X,Y and Z axes.

Geometrically, orthogonality has to do with the mutual directions of a set of vectors. But what about their lengths?

For a vector in 2 or 3 dimensions it is obvious what the length is – by the Pythagoras theorem the length of (x,y) is just $\sqrt{x^2 + y^2}$. We can extend this idea to any number of dimensions:

- *The length of a vector* $X = (x_1, x_2, x_3, \ldots, x_n)$ *is* $\sqrt{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}$.
- *A vector with a length equal to 1, is called a* ***unit vector*** *or alternatively we say that the vector is* ***normal****.*

Given any vector (x,y, …), we can find the unit vector pointing in the same direction simply by dividing it by its length. This is called **normalising** the vector.

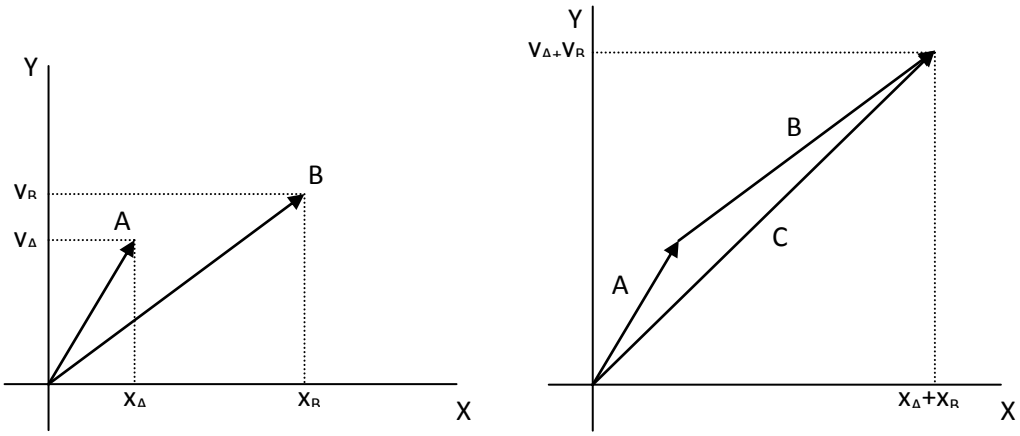
When we have a set of orthogonal vectors, it is sometimes convenient to work with the set of unit vectors obtained when we normalise each of the orthogonal vectors. Such a set, consisting of unit vectors that are orthogonal to each other, is called an **orthonormal** set. An example is the pair of unit vectors that point along the X and Y axes respectively.

When a set of vectors are orthonormal, that means that when multiplying two vectors belonging to this set we will always get either zero or one: 1 if the we multiply any vector by itself, and 0 if we multiply different vectors. Formally, we say $\underline{a}_i^T \underline{a}_j = 1$ while $\underline{a}_i^T \underline{a}_j = 0$ when i ≠ j.

In a later section we will extend these concepts, by showing that square matrices can also be orthogonal, and that orthogonal matrices have some useful special properties.

3.5 Geometric addition of vectors

Just like for any other matrices, when vectors are added we simply add the corresponding components. As was seen in the discussion above, the elements of a vector (such as a position vector) represent the components of the vector in a specific coordinate system. Consequently, the new vector $C = A + B$ obtained by adding components of vectors A and B is geometrically represented by the third side of the triangle formed by moving B so that its starting point coincides with the endpoint of A, as shown in the figure below.



Reversing the argument, we can see from this that the vector C could be formed by adding many alternative pairs of vectors different from A and B, corresponding to all other triangles that have one side equal to C.

However, suppose that we have a fixed set of vectors, like the A and B of our example, would we be able to form any other vector by just adding multiples of A and B? (Bear in mind that geometrically, taking a multiple of a vector means to change its length but not its direction). The answer to this question is the subject of section 4.1.

3.6 Matrices and vectors as objects

There is another general lesson that can be learnt from the geometric applications of matrices. It is that vectors and matrices should be considered as individual **objects**, not merely collections of numbers. The numbers are just a way to represent them.

The same vector will, for example, be represented by a different pair, or triplet, of numbers if we choose a different set of coordinate axes. The same is true of a matrix – e.g., if we have coordinate axes with a different orientation, the angles used to describe a rotation will be different and the individual elements of the rotation matrix will be different. Nevertheless, the matrix will still describe the same physical action, and so is an entity with its own meaning quite independently from the axes that were chosen for convenience to represent it.

4 Eigenvalues and Diagonalization

4.1 Linear superpositions of vectors

4.1.1 Linearly independent vectors

Orthogonal vectors are examples of vectors that are independent. To explain what this means, let us first consider what is meant by saying that one vector depends on others.

A vector X_1 depends on a collection of other vectors $X_2 \ldots X_n$, if it can be constructed out of them, in other words if we can find coefficients $c_2 \ldots c_n$ such that

$$X_1 = c_2 X_2 + c_3 X_3 + \ldots . c_n X_n$$

The 3 vectors A,B and C in a flat plane, discussed in section 3.5, is a simple example of this. One can see from the geometry that any vector in a plane can always be built up out of the other two, by forming a triangle where they form the sides, with their lengths adjusted as necessary but without changing their directions – that is what the coefficients take care of. But that would not, for example, be true in 3 dimensions when A,B and C are not in the same plane.

The expression on the right hand side of the equation above is called a **linear superposition** of vectors. So X_1 **depends** linearly on $X_2 \ldots X_n$, if it can be written as a linear superposition of them.

Then clearly X_1 is **linearly independent** of $X_2 \ldots X_n$, if it **cannot** be written as a linear superposition of them. The simplest example of this is two orthogonal vectors, e.g. mutually perpendicular vectors in a plane; obviously one of them cannot be made a multiple of the other (which is the simplest kind of linear superposition) no matter what coefficient is tried.

More formally, we say that a set of column vectors, $X_1,X_2,\ldots,X_n$ are (mutually) linearly independent if, and only if, reduction of their linear combination to zero

$$c_1X_1 + c_2X_2 + \ldots\ldots c_nX_n = 0$$

(where the $c_1\ldots..c_n$ are scalars), **only** occurs when

$$c_1 = c_2 = c_3 = \ldots\ldots = c_n = 0$$

It is not hard to find out if this happens or not for a given set of vectors. We just put the *X*-vectors together as the columns of a matrix, and multiply that with the column vector formed from the *c*'s. That is exactly the left hand side of the linear combination equation above. Since the X-vectors are known here, the resulting matrix equation is an homogeneous equation to be solved for the vector of unknown *c*'s. The *X*'s are linearly independent if the equation only has the trivial solution. That happens if the ***X*** matrix is not singular, i.e. its determinant is non-zero.

So we have a simple test for linear independence:

> *A set of vectors is linearly independent if the determinant of the matrix formed by combining the vectors into a matrix, is not zero.*

All of the vectors in the linear combination have to have the same dimension – otherwise we cannot add them up. And it can be shown that the maximum number of vectors that can be linearly independent is equal to this dimension. The example of 3 vectors in 2 dimensions mentioned above illustrates this.

The concept of one vector being written as a linear superposition of others is an extremely important one, which comes up in many branches and applications of mathematics. For example, many functions can be written as superpositions of other functions (think of Fourier sums). In some ways we can think of a function as a vector with an infinite number of dimensions. We will also see that superposition of functions plays a role in solving differential equations; more about that in Part II of this book.

Example

Suppose we have the following three vectors. We wish to find if they are independent.

$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad x_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \qquad x_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Now if we set up the test for independence. Notice in the following equation how we can think of a matrix as a row vector, in which each element is a column vector:

$$c_1x_1 + c_2x_2 + c_3x_3 = 0 \quad \Leftrightarrow \quad \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0$$

When we put back the individual vectors into this, we get

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -2 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = 0$$

Here we have three equations in three unknowns and we could use *Maple* or *Mathematica* to find the inverse. Or we could reduce the matrix to row echelon form. But the test above gives us a shortcut: we find the determinant, because if the determinant is non-zero then an inverse exists and we know from above that there is only the trivial solution available.

We expand the determinant by elements in the first column:

$$\det = 1\begin{vmatrix} 0 & -2 \\ -1 & 1 \end{vmatrix} - 2\begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} + 1\begin{vmatrix} 1 & 1 \\ 0 & -2 \end{vmatrix} = 1(-2) - 2(2) + 1(-2) = -8$$

It is not zero, so from this test we know that the three vectors x_1, x_2 & x_3 **are** linearly independent.

4.1.2 Eigenvectors of a matrix

We have seen before that in general, when a matrix is multiplied into a vector, the result is another vector, which is usually quite different (e.g., pointing in another direction).

However, it can and does happen that sometimes a matrix just changes the size of the vector and nothing else:

$$A X = \lambda X$$

In other words, the complicated matrix multiplication procedure produces the same result as merely multiplying the vector by a **scalar** number λ.

It turns out, as we shall see, that for any matrix this only happens for a small set of very special vectors. The particular vectors for which it happens, are different for each matrix, and is called its set of **eigenvectors**. Usually the number of independent eigenvectors is the same as the dimension of the matrix, but in exceptional cases there may be even fewer. And for each of these, the scalar value λ which we can substitute for the matrix, can be different – this is called the **eigenvalue** that belongs with that eigenvector.

Once we know what the eigenvectors and eigenvalues of a given matrix are, it becomes simple to predict what the matrix will do to any other vector. Starting with an N-dimensional matrix A, we denote its set of eigenvectors as $\underline{a}_i$ where $i = 1, 2, \ldots N$ and its eigenvalues as λ_i. Note that we use a slightly different notation than before – to show that the eigenvectors belong to matrix A, we use the same small (lower case) letter, and then underline it to show that it is a vector and not a scalar.

As one cannot have more than N linearly independent vectors in N dimensions, it must be possible to write any other vector $\underline{x}$ as a linear combination of them:

$$\underline{x} = c_1 \underline{a}_1 + c_2 \underline{a}_2 + \ldots + c_N \underline{a}_N$$

Making use of our knowledge that multiplying A by its eigenvector is just like multiplying it by the eigenvalue, the effect of applying A to the unknown vector $\underline{x}$ is easily obtained from this equation:

$$A\,\underline{x} = c_1\,\lambda_1\,\underline{a}_1 + c_2\,\lambda_2\,\underline{a}_2 + \ldots + c_N\,\lambda_N\,\underline{a}_N$$

For large matrices, the right hand side of this equation is much easier to calculate than the left hand side. Applying the rules of matrix multiplication, the left hand side requires of the order of N^3 arithmetic operations, but the right hand side only needs multiplying each vector by a scalar and adding them up, i.e. of the order of N^2 operations. This can be important, especially in applications like computer graphics where we may need to multiply a large number of vectors by the same matrix, e.g. to rotate a 3-dimensional image on the screen.

For this to work, we still need to calculate the coefficients c_i for a given vector $\underline{x}$. However, this is easy, because as we will see, the eigenvectors are (or can be made to be) an orthonormal set. By this we mean (see 4.6) that $\underline{a}_i^T\,\underline{a}_i = 1$ and $\underline{a}_i^T\,\underline{a}_j = 0$ So if we multiply the linear superposition for $\underline{x}$ through by $\underline{a}_i^T$ most of the terms on the right just fall away and we are left with the simple result

$$c_i = \underline{a}_i^T\,\underline{x}$$

The discussion above shows that knowing the eigenvectors and eigenvalues of a matrix can be very useful; we will come across other uses for them in what follows as well. So the next step is to find a way to calculate them.

4.2 Calculating Eigenvalues and Eigenvectors

Suppose we have a given matrix A. Then the equation that defines its eigenvalues and eigenvectors can be written as

$$A\,\underline{x} - \lambda\,\underline{x} = 0$$

This looks rather like the matrix form of a homogeneous set of equations. To make it even more so, we would like to factor out the $\underline{x}$ – but need to be careful, because the temptation is to write (A-λ) for the remaining factor. That would be wrong, because A is a matrix while λ is a number and we can only subtract objects that conform. That is corrected by inserting a unit matrix, i.e. $\lambda\,\underline{x} = \lambda\,I\,\underline{x}$ so that we get

$$\left(A - \lambda\,I\right)\underline{x} = 0$$

Now the equation really does look like a homogeneous set, which we know how to solve – but there is one difference. Not only is the vector $\underline{x}$ unknown, as before, but in addition there is an unknown λ in the coefficient matrix! But this value (the **eigenvalue**) can be determined beforehand, because we know that in order to have non-trivial solutions for the $\underline{x}$, the determinant of the coefficient matrix must be zero.

For every solution that we find for the λ, we have a different set of equations to solve to get the **eigenvectors**; but each of them is just a homogeneous equation problem that we solve as before by using reduction to REF.

Example 1.

If we use the *A* matrix we had above

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & -2 \\ 1 & -1 & 1 \end{bmatrix}$$

then to find eigenvalues we solve

$$\det(A - \lambda I) = 0 \quad \Leftrightarrow \quad \begin{vmatrix} 1-\lambda & 1 & 1 \\ 2 & -\lambda & -2 \\ 1 & -1 & 1-\lambda \end{vmatrix} = 0$$

This gives by calculating the determinant using the first column

$$\begin{aligned} &(1-\lambda)[-\lambda(1-\lambda)-(-1)(-2)]-2[1(1-\lambda)-(-1)1]+1[-2+\lambda]=0 \\ &-\lambda+\lambda^2-2+\lambda^2-\lambda^3+2\lambda+\lambda-2=0 \\ &-\lambda^3+2\lambda^2+4\lambda-8=0 \\ &\lambda^3-2\lambda^2-4\lambda+8=0 \end{aligned}$$

Cubic equations are usually hard to solve, but here we notice that λ = 2 solves it so to get the other solutions, take (λ-2) as a factor and divide it out:

$$\lambda-2\overline{)\lambda^3-2\lambda^2-4\lambda+8}\;\;\lambda^2-4$$

The remaining part (λ^2-4) factors into (λ-2)(λ+2). So we have for this particular cubic

$$(\lambda-2)^2(\lambda+2)=0 \Leftrightarrow \lambda=2,2,-2$$

So, in this case where we had a 3-dimensional matrix we obtained 3 eigenvalues. Looking back you can see that the reason is that we put an additional λ on each row of the matrix; so if the matrix has *N* dimensions, the determinant that we need to solve will be an *N*-degree polynomial in the variable λ and will have *N* solutions. The polynomial is often called the **characteristic polynomial** of the matrix. *Mathematica* provides the instruction **CharacteristicPolynomial[M,x]** to calculate it as a polynomial in the variable *x*, for a matrix *M*.

Finding eigenvalues of matrices is the same as finding roots of polynomials. Closed form solution formulas as used routinely for quadratic expressions don't exist for polynomials above quartics (λ^4 + …..) and even those for quartics and cubics are very complicated. Numerical methods, such as Newton's method, or other are available from computer programs, but complex roots can be an additional complication.

There are properties of some matrices which help; e.g. symmetric matrices have non-negative real roots.

Returning to the calculation for matrix *A*, we proceed to calculate the **eigenvectors** for each of the 3 eigenvalues.

For $\lambda = -2$

$$(A-\lambda I)=\begin{bmatrix}3 & 1 & 1\\ 2 & 2 & -2\\ 1 & -1 & 3\end{bmatrix}$$

We can see that row 3 = row 1 – row 2 $\therefore$ det =0, as it should be because we calculated λ that way.

First reduce this to row echelon form

$$\begin{bmatrix} 1 & -1 & 3 \\ 2 & 2 & -2 \\ 3 & 1 & 1 \end{bmatrix}$$ by swapping row 1 and row 3

$$\begin{bmatrix} 1 & -1 & 3 \\ 0 & 4 & -8 \\ 0 & 4 & -8 \end{bmatrix}$$ row 2 = row 2 –2*row 1; row 3 = row 3 –3* row 1

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$ row 3 = row 3 – row 2; row 2 = row 2/4; row 1 = row 1 + row 2

In the standard way, we ignore the last row; the 2nd row allows us to choose the arbitrary value 1 for x_3 and this leads to $x_2 = 2$ and then from the first row $x_3 = -1$. So the eigenvector belonging with $\lambda = -2$ is

$$\underline{x}_1 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

The arbitrary choice that we made for x_3, means that other choices would also have given us an eigenvector. But all such choices are proportional to each other; one can easily see from the eigenvector definition that if $\underline{x}$ is an eigenvector, then any multiple of $\underline{x}$ will also be an eigenvector with the same eigenvalue. This means that we are free to choose the length of the eigenvector, and as we will see below it is usually convenient to choose the length as 1. To do this, we just divide the vector that we obtained by its length – i.e., we **normalise** the eigenvector.

The case of the other eigenvalue, $\lambda = 2$, is slightly more complicated because it occurred more than once (a repeated root). Let us see what happens in that case.

Putting $\lambda = 2$, then when we form A-λI we get

$$A - 2I = \begin{bmatrix} -1 & 1 & 1 \\ 2 & -2 & -2 \\ 1 & -1 & -1 \end{bmatrix}$$

and if we reduce this to *ref* (row echelon form) we get

$$\begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

In this case the first non-trivial row contains all three unknowns and we have both x_2 and x_3 arbitrary. This is as it should be, so that we can get two solutions belonging to the two roots. We must consider each option separately.

Firstly, put $x_3 = 0$, and $x_2 = 1$ then this gives $x_1 = 1$. Thus the eigenvector becomes $\underline{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$.

Secondly, put $x_2 = 0$ and $x_3 = 1$, then this gives $x_1 = 1$. Here the eigenvector is $\underline{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$.

Example 2

$A = \begin{bmatrix} 3 & 2 & -2 \\ 2 & 0 & -1 \\ -2 & -1 & 0 \end{bmatrix}$, then $\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 2 & -2 \\ 2 & -\lambda & -1 \\ -2 & -1 & -\lambda \end{vmatrix} = 0$ gives the polynomial

$(3-\lambda)(\lambda^2-1) - 2(-2\lambda-2) - 2(-2-2\lambda) = 0$

$-\lambda^3 + 3l\,\lambda^2 + 9\lambda + 5 = 0$

Substitution shows that λ = -1 solves this equation, so we can factor this by long division as we did before to get (λ-5)(λ+1)(λ+1) and we have eigenvalues of +5, -1 & -1. Take each of these in turn to find eigenvectors:

$$(A-5I)=\begin{bmatrix}-2 & 2 & -2\\ 2 & -5 & -1\\ -2 & -1 & -5\end{bmatrix} \Rightarrow \text{REF} \Rightarrow \begin{bmatrix}1 & 0 & 2\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$$ which gives the eigenvector $\underline{x}_1=\begin{bmatrix}-2\\ -1\\ 1\end{bmatrix}$.

For the repeated eigenvalue λ = -1

$$(A+1I)=\begin{bmatrix}4 & 2 & -2\\ 2 & 1 & -1\\ -2 & -1 & 1\end{bmatrix} \Rightarrow \text{REF} \Rightarrow \begin{bmatrix}2 & 1 & -1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix} \Rightarrow \text{eigenvectors } \underline{x}_2=\begin{bmatrix}\frac{-1}{2}\\ 1\\ 0\end{bmatrix} \text{ and } \underline{x}_3=\begin{bmatrix}\frac{1}{2}\\ 0\\ 1\end{bmatrix}$$

Again, a single eigenvalue has multiple eigenvectors if the eigenvalue is a multiple root of the polynomial.

Example 3

$$A=\begin{bmatrix}1 & 4 & -2\\ 4 & 1 & 2\\ -2 & 2 & 4\end{bmatrix}$$

gives rise to the **characteristic** polynomial $\lambda^3 - 6\lambda^2 - 15\lambda + 100$, which has roots λ = 5, 5, -4.

For λ = 5 we get two eigenvectors $\begin{bmatrix}-1\\ 0\\ 2\end{bmatrix}$ & $\begin{bmatrix}1\\ 1\\ 0\end{bmatrix}$ and for l = -4 we have an eigenvector $\begin{bmatrix}2\\ -2\\ 1\end{bmatrix}$.

To solve this problem using *Mathematica*, we could enter

Eigenvalues[{{1, 4, -2}, {4, 1, 2}, {-2, 2, 4}}]

and this gives the eigenvalues as a list {5, 5, -4}. Then we can find eigenvectors by the instruction

Eigenvectors[{{1, 4, -2}, {4, 1, 2}, {-2, 2, 4}}]

which gives the result as a list of vectors {{-1, 0, 2}, {1, 1, 0}, {2, -2, 1}}. Both steps can be accomplished together by the instruction

Eigensystem[{{1, 4, -2}, {4, 1, 2}, {-2, 2, 4}}]

The result from this is {{5, 5, -4}, {{-1, 0, 2}, {1, 1, 0}, {2, -2, 1}}} which is slightly harder to read, but has the advantage that it is clear which eigenvalue in the first sublist belongs to which eigenvector in the second sublist.

4.3 Similar matrices and diagonalisation

We start with a definition:

> *Two matrices A and B are termed similar if for some non-singular matrix Q,*
> $B = Q^{-1}A\,Q$ *or* $QB=AQ$.

We would also call this a **similarity transformation**, or say A has been transformed into B by applying a similarity transformation to A.

This has a close connection with eigenvectors. So far we have considered the equation that connects an eigenvalue and corresponding eigenvector to A separately for each eigenvector.

However, we can collect all these equations together by using the **eigenvector matrix** P. This is simply a matrix that is made up by taking the eigenvectors as its columns. E.g, for example 3 in 0 above, P is given by

$$P = \begin{bmatrix} -1 & 1 & 2 \\ 0 & 1 & -2 \\ 2 & 0 & 1 \end{bmatrix}$$

We saw before how the matrix form can collect a set of separate algebraic equations together into a single matrix equation. Similarly, we can collect the equations that are satisfied by the separate eigenvectors into the following single matrix equation:

$$A\,P = P \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_N \end{bmatrix}$$

This expression is easily confirmed by just multiplying out the right hand side, from which you should see that in effect the diagonal matrix containing the eigenvalues on its diagonal, just multiplies each column by its appropriate eigenvalue. So putting each column of $A\,P$ equal to the corresponding column on the right, is just the "ordinary" eigenvalue equation for that single eigenvector.

The following shorthand is often used to represent this equation more compactly:

$$A\,P = P\,\Lambda$$

The equation just defines a new notation – we use the capital Λ to indicate a **diagonal** matrix with the eigenvalues on its diagonal, as written out in the previous version of the eigenvector matrix equation. It looks superficially very much like the original equation for a single eigenvector, but notice that because P and Λ are matrices their order is important and in fact we have to write the eigenvalue matrix Λ last, not first.

Now comparing this equation with the similarity definition above we see that they have the same form, and can also be rewritten

$$P^{-1}AP = \Lambda$$

In other words, if we use the eigenvector matrix of A to apply a similarity transformation to it, a very special result is obtained – the original non-diagonal matrix is transformed into a much simpler diagonal one, and as a bonus the values that appear on its diagonal are just the eigenvalues!

So if we can find any other way to determine such a special similarity transformation, we would also have solved the eigenvalue problem for the matrix. In practice it usually works the other way round – we solve the eigenvalue problem in the way described in the previous section, and we can then **diagonalise** the matrix by using its eigenvector matrix. The term "diagonalise matrix A" is often used as shorthand for "find the eigenvalues and eigenvectors of matrix A".

Why are matrices that are connected by the formula above called **similar**? That is because they share many important values and properties, namely :

- *Similar matrices share eigenvalues*
- *If z is an eigenvector of $B = Q^{-1}A\,Q$, then $Q\,z$ is an eigenvector of A.*
- *Similar matrices have the same determinant value.*
- *The trace (i.e., the sum of diagonal elements) of similar matrices are the same*

To prove the first statement, calculate the eigenvalues of B by evaluating

$$\det(B-\lambda I)=\det(Q^{-1}AQ-\lambda I)=\det(Q^{-1}AQ-\lambda Q^{-1}Q)=\det[(Q^{-1})(A-\lambda I)Q]$$
$$=\det(Q^{-1}).\det(A-\lambda I).\det(Q)=\det(Q^{-1}).\det(Q).\det(A-\lambda I)=\det(A-\lambda I)$$

That is, the two similar matrices have the same characteristic polynomial and thus have the same eigenvalues. In this calculation we made use of the determinant properties that were listed at the start of section 2.4.2.

If z is an eigenvector of $B = Q^{-1}A\,Q$, then $Q\,z$ is an eigenvector of A.

For proving the second statement, consider that for $\underline{z}$ to be an eigenvector of B we must have

$$B\,\underline{z}=\lambda\,\underline{z}$$

If we insert the expression that defines B as a similarity transformation of A and multiply through by Q, we get

$$Q^{-1}AQ\,\underline{z}=\lambda\,\underline{z}$$
$$AQ\,\underline{z}=\lambda Q\,\underline{z}$$

Thus $Q\,z$ is an eigenvector of A.

The 3rd property you can prove for yourself, in a very similar way as the first. The proof of the last is also similar, but relies on the fact that trace($A.B$) = trace($B.A$) just like the analogous property of determinants. If you consider 2x2 matrices A and B and write out the formulas for the diagonal elements of $A.B$ and $B.A$, it should be easy to see why the trace also follows such a simple rule for matrix products.

From these properties, a lot can be learned about the eigenvalues and eigenvectors of A by studying those of any other matrix that is similar to it. And we now know that A is similar to Λ, so what can we tell about the eigenvalues and eigenvectors of Λ?

The answer is that the eigenvectors of a diagonal matrix is simply the set of unit vectors

$$\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix},\begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix},\ldots\ldots\begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$$

and moreover, the corresponding eigenvalues are just the values on the diagonal. You should test this for yourself by simply multiplying out the eigenvalue equation for a diagonal matrix.

The conclusions that can be drawn from all the connections that we studied, are summarised in a later section.

4.4 How eigenvalues relate to determinants

For a diagonal matrix,the determinant is pretty easy to calculate; it is just the product of all the diagonal elements. Similarly the trace is just the sum of the diagonal elements. Applying this to the Λ matrix constructed out of eigenvalues, it follows that det(Λ) = product of eigenvalues and trace(Λ) = sum of eigenvalues. But because the Λ matrix is **similar** (in the strict mathematical sense defined in the previous section) to the original matrix such as A above, and they share eigenvalues, we can reach the following conclusions:

- *The determinant of any matrix A is equal to the product of its eigenvalues*
- *The trace of any matrix A is equal to the sum of its eigenvalues*

4.5 Using diagonalisation to decouple linear equations

In the usual form of a set of linear equations, the unknowns x_i are all coupled (i.e., we cannot solve for them one at a time, but all have to be determined together:

$$A\,\underline{x} = \underline{b}$$

Now taking any invertible matrix Q, we can multiply this equation by Q^{-1} and write it in the form:

$$Q^{-1} A Q Q^{-1} \underline{x} = Q^{-1} \underline{b}$$

By grouping together the factors in the following way, we can recognise the form of a similarity transformation:

$$\underbrace{Q^{-1} A Q}_{B} \quad \underbrace{Q^{-1} \underline{x}}_{\underline{y}} = \underbrace{Q^{-1} \underline{b}}_{\underline{z}}$$

We saw previously that, at least for 2- and 3-dimensional vectors, multiplying a vector by a matrix Q can be interpreted as rotating the vector in space. The same effect is achieved if we rotate the system of axes used to describe the vector, in the opposite way. Therefore, we can interpret the action of Q^{-1} as just rotating the axes. $\underline{y}$ would then just be the same vector described in this new set of axes. Then the equation above shows that the similarity transformation using Q just gives the new form of *A* as it would be in the new set of axes.

Now suppose that we make a special choice for Q by taking it to be the eigenvector matrix *P* that belongs to *A*. Then the similarity transformation causes *B* to reduce to the diagonal matrix, i.e. the equation above becomes

$$\Lambda \underline{y} = \underline{z} = P^{-1}\underline{b}$$

Now this is again a set of linear equations, for the unknown components of the vector $\underline{y}$ this time, but it is a far simpler than the one before for $\underline{x}$ – because the equations are now **decoupled**. By that is meant that if we multiply it out, we see that each equation now contains only one variable, and so the equations can each be solved separately. The solution for the *i*'th equation is simply

$$y_i = \frac{z_i}{\lambda_i}$$

Then we can put these together into a solution vector $\underline{y}$ and then since $\underline{y} = Q^{-1}\underline{x}$ we see that when the special choice for Q is made, the solution of the original, coupled set of equations is given by

$$\underline{x} = P\underline{y}$$

So solving the eigenvalue problem for the original coefficient matrix has enabled us to simplify the difficult **coupled** set of equations to an almost trivial **uncoupled** set. This type of decoupling is another use for eigenvalue calculations, and we will see that it can be used similarly also when solving differential equations.

Notice that in using an eigenvector matrix to perform a similarity transformation, and hence to solve equations by the decoupling method,, we need to find its inverse. This would normally be hard, but there are special matrix properties that can make this task a lot easier. That will be studied in the next section.

4.6 Orthonormal eigenvectors

In section 3.4 the geometric meaning of orthogonal and orthonormal sets of vectors were discussed. We will now see that the set of eigenvectors of a matrix often is, or can be made to be, an orthonormal set. But first we need to extend these concepts to matrices.

A **matrix** is called orthogonal when all its column vectors are orthogonal. For example, using the two vectors considered in section 3.4, U defined as follows is orthogonal:

$$U = \begin{bmatrix} x & -y \\ y & x \end{bmatrix}$$

If we now form the product U^TU we get

$$U^TU = \begin{bmatrix} x & y \\ -y & x \end{bmatrix}\begin{bmatrix} x & -y \\ y & x \end{bmatrix} = \begin{bmatrix} x^2+y^2 & 0 \\ 0 & x^2+y^2 \end{bmatrix}$$

The fact that the off-diagonal elements are zero, is no coincidence. Remember that the rows of U^T are just the columns of U, and so an off-diagonal element in U^TU is just the product of two different columns of U and this is zero because the columns are orthogonal. So the same will happen for any orthogonal matrix.

A particularly useful kind of orthogonal matrix is one where in addition to this, the diagonal elements are 1 so that then $U^TU=I$, the identity matrix. In this case we call U an **orthonormal** matrix.

The diagonal elements in our example were each equal to the square of the length of the corresponding column vector of U. They also happened to be equal, but that was just because our second vector $(-y, x)$ was obtained by just rotating the first. We could have used a vector $(-3y, 3x)$ instead; the matrix would still be orthogonal, but its diagonal elements would have been different.

So a matrix is orthonormal if the column vectors are not only orthogonal, but in addition each of them individually have a unit length. We can also say an orthonormal matrix is one where its column vectors form an orthonormal set.

The reason that this is so useful is that if we compare the orthonormal property $U^TU=I$ with the definition of an inverse, $U^{-1}U=I$, we see that

*The **inverse** of an orthonormal matrix U is exactly the same as its **transpose**.*

In other words, if we can somehow make sure that the matrix that we work with is orthonormal, life becomes so much easier because instead of having to laboriously calculate its inverse, we can take the shortcut of merely writing down its transpose – no calculation required.

This is not always possible – e.g., in solving equations we have to make do with the coefficient matrix that we are given – but we will see below that when working with eigenvectors, applying a little foresight can give us this wonderful simplification.

It is helpful to have a geometrical interpretation of what it means that a matrix is orthonormal. It turns out that this means that such a matrix transforms a vector in such a way that its length is not changed. For example, rotations and reflections do not change the length of vectors – and you should check for yourself that the matrices given before for these transformations are indeed orthonormal. On the other hand, the matrix for the shear transformation is not orthonormal, and it is clear from the figure illustrating a shear transformation that the length of the vector does change here.

When we use a set of vectors to construct another vector (a so-called **basis** set), as we do with eigenvectors as discussed at the end of section 4.1.2, we can classify their desirable properties into a hierarchy as follows:

1. At least we want the set to be linearly independent. Adding another dependent vector to a set, won't make it any better, because anything the new vector contributes can be equally well represented by the existing vectors in the set.
2. Having an orthogonal set is better, because then the contribution made by each vector in the set is unique. An orthogonal set is necessarily linearly independent, but not vice versa.

3. The best is an orthonormal set, because then all vectors in the set contribute equally, and that simplifies a lot of the arithmetic. An orthonormal set is both orthogonal and linearly independent.

If we look back at the examples of eigenvectors calculated for three different matrices that were shown in section 4.2, some interesting observations can be made. In all three cases, the eigenvectors that were calculated were linearly independent.

However, in the last two examples, we can see that the eigenvectors that belong with **different** eigenvalues, were in fact **orthogonal.** The ones that belonged to the same eigenvalue were not orthogonal to each other but merely linearly independent.

In the first example, on the other hand, there was no orthogonality at all.

The reason that the relationship was different in the first example, is that in the last two cases the matrix was **symmetric**, i.e. equal to its own transpose. In the first example the matrix was not symmetric. It is generally found that a symmetric matrix gives orthogonal eigenvectors for all those belonging to different eigenvalues.

A symmetric, real matrix also has the nice property that its eigenvalues and eigenvectors always come out as real numbers. If the matrix is not symmetric, one can get complex eigenvalues and/or complex eigenvectors, although as the first example shows that does not necessarily happen. More details about complex numbers appear in the Appendix.

We can check on the orthogonality of the eigenvectors by seeing if the definition given above for an orthonormal matrix U works for it. Take, for example, the eigenvector matrix of example 3:

$$P = \begin{bmatrix} -1 & 1 & 2 \\ 0 & 1 & -2 \\ 2 & 0 & 1 \end{bmatrix}$$

We can check this by calculating $P^{\mathrm{T}}P$ as follows

$$P^T P = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 1 & 0 \\ 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 2 \\ 0 & 1 & -2 \\ 2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Now if we look carefully at P we find that the first two columns of P are made up of the two eigenvectors associated with $\lambda = 5$ and the third column comes from the eigenvector associated with $\lambda = -4$. From the resultant $P^{\mathrm{T}}P$ we can see that the product of the eigenvectors from different eigenvalues are zero ie. the eigenvectors are orthogonal, but not so for the repeated eigenvalues.

However, with a little extra effort we can **make** P orthonormal. We need to change two things. First, we have to make the two eigenvectors that belong to the same eigenvalue orthogonal to each other.

To do that, we leave one of them – here we choose to take $\underline{x_2} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ - the same, but replace the other by a new one constructed as a linear superposition of them both:

$$\underline{x_1}' = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} s-1 \\ s \\ 2 \end{bmatrix}$$

Now we choose the coefficient s in the linear superposition in such a way that it makes the new vector orthogonal to $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ which means that $\begin{bmatrix} s-1 & s & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0$. This means that s-1+s = 0, which gives s=½.

That gives the new orthogonal eigenvector as

$$\underline{x_1}' = \begin{bmatrix} \frac{-1}{2} \\ \frac{1}{2} \\ 2 \end{bmatrix}$$

The second step is to normalise all the new columns. We can **normalise** a vector by dividing each of its components by its length, the sum of the squared components. In other words, if we have a vector

$\underline{z} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ then the normalized vector is $\hat{z} = \dfrac{1}{\sqrt{a^2 + b^2 + c^2}} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$

The hat on top of the z is a notation that is often used to indicate a unit vector.

Thus the normalized vector from the above P matrix for l = -4 would be

$$\frac{1}{\sqrt{9}} \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

and for the newly orthogonalised vector that we found, it is $\begin{bmatrix} \frac{-1}{3\sqrt{2}} \\ \frac{1}{3\sqrt{2}} \\ \frac{4}{3\sqrt{2}} \end{bmatrix}$.

Repeating the normalisation for the third vector as well, we obtain the orthonormal form of the eigenvector matrix:

$$P' = \begin{bmatrix} \frac{-1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{-2}{3} \\ \frac{4}{3\sqrt{2}} & 0 & \frac{1}{3} \end{bmatrix}$$

For practice, you should check that each column of this matrix is still an eigenvector of A, and that P' is now an orthonormal matrix.

The orthonormalisation process described above is called the Gram-Schmidt method, and is relatively simple, but there are more sophisticated methods available. To avoid the tedious manual work, one could have used Mathematica for the purpose, with the instruction:

```
Orthogonalize[{{-1, 0, 2}, {1, 1, 0}, {2, -2, 1}}]
```

Despite the name, this instruction includes the normalisation step as well and yields the output $\{\{-\frac{1}{\sqrt{5}}, 0, \frac{2}{\sqrt{5}}\}, \{\frac{4}{3\sqrt{5}}, \frac{\sqrt{5}}{3}, \frac{2}{3\sqrt{5}}\}, \{\frac{2}{3}, -\frac{2}{3}, \frac{1}{3}\}\}$. The last vector is the same as we obtained, but the first two are different. They are, nonetheless an orthonormal set, and could equally well be used in the analysis that follows. We learn from this that orthogonalization is not a unique process; in fact *Mathematica* allows one to choose different algorithms through optional arguments not shown the instruction reproduced above, and these may (or may not) give different answers.

Having changed *P* into an orthonormal matrix, we are now able to find its inverse by simply transposing it. That makes it easy to use for decoupling equations as we saw in section 4.4.

The trick of orthogonalising the eigenvectors to each other by making a linear superposition of them, only worked because they belong to the same eigenvalue. If they belonged to different eigenvalues, a linear superposition would not be an eigenvector any more – prove that yourself by applying the definition of an eigenvector!

If the original matrix was symmetric, that would never be necessary, because any eigenvectors that belong to different eigenvalues are automatically orthogonal. On the other hand an asymmetric matrix does not have that nice property and so it is not possible to reduce its eigenvector matrix to an orthonormal form.

To summarise: For a symmetric matrix, we can always reduce its eigenvector matrix to an orthonormal form and thus create a shortcut to get the inverse of the eigenvector matrix. If the matrix is not symmetric, we are stuck with the eigenvector matrix that we get in the first place and have to calculate its inverse the hard way e.g. by using a REF transformation.

4.7 Summary: eigenvalues, eigenvectors and diagonalisation.

4.7.1 Eigenvalues

- *An N-dimensional matrix A has N eigenvalues, some of which may be identical.*
- *The product of the eigenvalues is the determinant of the original matrix.*
- *This means that if we have a zero eigenvalue, $\lambda = 0$, then $\det(A) = 0$.*
- *The number of **non-zero** eigenvalues is the **rank** of the matrix i.e. the number of independent columns.*
- *The sum of diagonal elements of A, which is called the **trace** of the matrix, is equal to the sum of the eigenvalues.*
- *If A is a matrix of real elements, then this leads to an eigenvalue polynomial with real coefficients, but the roots (i.e. eigenvalues) may be complex.*
- *If the elements of A is not only real but also symmetric, the eigenvalues will be real.*

4.7.2 Eigenvectors

- *There is one eigenvector per eigenvalue.*
- *When an eigenvalue is repeated, there are usually as many distinct eigenvectors for that eigenvalue as the repeat count. However, in exceptional cases the same eigenvector may be found more than once for the same eigenvalue and then we will in total have fewer non-trivial eigenvectors than the dimension of the matrix.*
- *The eigenvectors for **distinct** eigenvalues are linearly independent.*
- *If the matrix is symmetric, those eigenvectors are orthogonal as well.*
- *Eigenvectors belonging to the **same** eigenvalue are linearly independent but not necessarily orthogonal. However, as any linear combination of these is also an eigenvector with the same eigenvalue, they can be made orthogonal.*
- *Eigenvectors are solutions apart from an arbitrary scaling factor. The scaling factor can be chosen to normalise the eigenvectors, and produce an orthonormal eigenvector matrix.*

4.7.3 Steps to perform diagonalisation

- *Calculate the characteristic polynomial $\det(A - \lambda I)$ of an $\underline{n} \times \underline{n}$ matrix A*
- *Put this to 0 and solve to get the eigenvalues $(\lambda_1 \ldots\ldots \lambda_n)$*
- *Substitute each λ in $(A - \lambda I)$ to calculate the corresponding eigenvector; call them $\underline{x}_1, \underline{x}_2, \ldots\ldots, \underline{x}_n$*
- *Optional: If A was symmetric but had repeated eigenvalues, orthogonalise and normalise the eigenvectors.*
- *Create a matrix P made up from the eigenvectors of A: $[\underline{x}_1, \underline{x}_2, \ldots\ldots, \underline{x}_n]$*
- *Calculate P^{-1}, by transposing if orthonormalisation was carried out, otherwise by REF.*
- *$P^{-1}AP$ is the diagonal form of A; it should be a diagonal matrix Λ whose elements down the leading diagonal will be $(\lambda_1 \ldots\ldots \lambda_n)$*

Example

$A=\begin{bmatrix}2&1&1\\2&3&2\\1&1&2\end{bmatrix}$ has a characteristic polynomial given by $\begin{vmatrix}2-\lambda&1&1\\2&3-\lambda&2\\1&1&2-\lambda\end{vmatrix}$ which is

$$=(2-\lambda)\begin{vmatrix}3-\lambda&2\\1&2-\lambda\end{vmatrix}-(1)\begin{vmatrix}2&2\\1&2-\lambda\end{vmatrix}+1\begin{vmatrix}2&3-\lambda\\1&1\end{vmatrix}$$

$$=(2-\lambda)[(3-\lambda)(2-\lambda)-2]-[(2)(2-\lambda)-2]+[2-(3-\lambda)]$$

$$=5-11\lambda+7\lambda^2-\lambda^3$$

$$=(1-\lambda)(1-\lambda)(5-\lambda)$$

If we choose the eigenvalue λ=5 then the matrix A-5λI gives $\begin{bmatrix}-3&1&1\\2&-2&2\\1&1&-3\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=0$

Reduce the matrix to *ref* (row echelon form) to get $\begin{bmatrix}1&0&-1\\0&1&-2\\0&0&0\end{bmatrix}$. This yields $\begin{matrix}x_1-x_3=0\\x_2-2x_3=0\end{matrix}$

Choose x_3 = 1 in this to obtain the eigenvector $\begin{bmatrix}1\\2\\1\end{bmatrix}$.

If we take λ = 1 then A- λI becomes $\begin{bmatrix}1&1&1\\2&2&2\\1&1&1\end{bmatrix}$ which reduces by *ref* to $\begin{bmatrix}1&1&1\\0&0&0\\0&0&0\end{bmatrix}$. This is a matrix

with a rank =1 so there are two linearly independent solutions to $x_1+x_2+x_3=0$

Choosing x_3=0 gives x_1 = -x_2 and hence the eigenvector $\begin{bmatrix}-1\\1\\0\end{bmatrix}$

The alternative choice x_2=0 gives x_1 = -x_3 which gives the eigenvector $\begin{bmatrix}-1\\0\\1\end{bmatrix}$.

The matrix of eigenvectors is $P=\begin{bmatrix}1&-1&-1\\2&1&0\\1&0&1\end{bmatrix}$.

Because we did not take the trouble to orthonormalise, we might have to resort to a computer

calculation such as using *Mathematica*, to get $P^{-1}=\begin{bmatrix}1/4&1/4&1/4\\-1/2&1/2&-1/2\\-1/4&-1/4&3/4\end{bmatrix}$.

This gives

$$P^{-1}A = \begin{bmatrix} 5/4 & 5/4 & 5/4 \\ -1/2 & 1/2 & -1/2 \\ -1/4 & -1/4 & 3/4 \end{bmatrix} \text{ and } \quad D = P^{-1}AP = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ as we would expect.}$$

For practice, you should repeat the last part by first doing the orthonormalisation and check that the same result is obtained.

Part 2
Differential Equations

5 Revision: Calculus Results

5.1 Differentiation formulas

We show the Maple syntax in red, and the output produced in blue:

Diff(x^n,x);

$$\frac{\partial}{\partial x} x^n = n\, x^{n-1}$$

Diff(exp(kx),x);

$$\frac{\partial}{\partial x} \mathrm{e}^{kx} = k\, \mathrm{e}^{(k\,x)}$$

Diff(ln(x),x);

$$\frac{\partial}{\partial x} \ln(x) = \frac{1}{x}$$

Diff(sin(x),x);

$$\frac{\partial}{\partial x} \sin(x) = \cos(x)$$

Diff(cos(x),x);

$$\frac{\partial}{\partial x} \cos(x) = -\sin(x)$$

Diff(tan(x),x);

$$\frac{\partial}{\partial x} \tan(x) = \sec^2(x)$$

5.2 Rules of Differentiation

Sum **Diff(`(f+g)`,x);**

$$\frac{\partial}{\partial x}(f+g) = f' + g'$$

Product **Diff(`(fg)`,x);**

$$\frac{\partial}{\partial x}(fg) = f'\,g + g'\,f$$

Quotient **Diff(`(f/g)`,x);**

$$\frac{\partial}{\partial x}(f/g) = \frac{f' g - g' f}{g^2}$$

Chain **Diff(`f(g(x))`,x);**

$$\frac{\partial}{\partial x} f(g(x)) = f'(g(x))\, g'(x)$$

5.3 Integration Formulas

Int(x^n,x);

$$\int x^n \, dx = \frac{x^{n+1}}{n+1} + c$$

Int(1/x,x);

$$\int \frac{1}{x} \, dx = \ln(x) + c$$

Int(g^`'`*`(x)`/`g(x)`,x);

$$\int \frac{g'(x)}{g(x)} \, dx = \ln(\mathrm{g}(x)) + c$$

Int(exp(k*x),x);

$$\int \mathrm{e}^{kx} \, dx = \frac{\mathrm{e}^{kx}}{k} + c$$

Int(sin(x),x);

$$\int \sin(x) \, dx = \text{-}\cos(x) + c$$

Int(cos(x),x);

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

Int(sec^2*x,x);

$$\int \sec^2 x \, dx = \tan(x) + c$$

5.4 Integration Methods

5.4.1 Integration by Substitution

$$\int \frac{\cos(x)}{1+\sin(x)} dx$$

Use the substitution $u = 1 + \sin(x)$ and differentiate: $du = \cos(x).dx$

The integral then becomes $\int \frac{du}{u} = \ln(u) + c$

$$\int \frac{\cos(x)}{1+\sin(x)} dx = \ln(1+\sin(x)) + c$$

Check using Maple : **int(cos(x)/(1+sin(x)),x);**

$$\ln(1+\sin(x))$$

In general when using this technique one attempts to choose a substitution to get the integral into one of the standard forms given above, or in more extensive integration tables.

5.4.2 Integration by parts

Remember the product rule

$$\frac{d}{dx}(f\,g) = f'g + g'f$$

Integrate both sides

$$f\,g = \int f'\,g + \int g'\,f$$

and rearrange

$$\int f'\,g = f\,g - \int g'\,f$$

Example $\int xe^x dx$

When doing integration by parts it is a good idea to develop a consistent way of setting them out. By using a 2 x 2 template for the functions and derivatives and filling it in, the process becomes easier.

Choose

$f = x$	$f' = 1$
$g = e^x$	$g' = e^x$

Thus

$$\int x\,e^x dx = x\,e^x - \int e^x\,1\,dx$$
$$= x\,e^x - e^x + c$$

This can be done using the Maple command

```
int(x*exp(x),x);
```

$$x\,e^x - e^x$$

5.4.3 Use of Partial Fractions

This is used to integrate rational expressions consisting of a polynomial in both the numerator and denominator.

Remember how it is possible to combine fractions together over a common denominator:

$$\frac{2}{(x-1)} + \frac{4}{(x+3)} = \frac{2(x+3)+4(x-1)}{(x-1)(x+3)} = \frac{6x+2}{(x-1)(x+3)}$$

Partial Fractions is the reverse of this process.

$$\frac{6x+2}{(x-1)(x+3)} = \frac{A}{(x-1)} + \frac{B}{(x+3)} = \frac{A(x+3)+B(x-1)}{(x-1)(x+3)}$$

Gathering coefficients of x^1 and x^0 (=1) and equating numerators on both sides gives

$$A + B = 6$$
$$3A - B = 2$$

There are two equations and two unknowns. Solve for A and B. In complicated cases we might use REF; but here just add

$$4A = 8; \quad \Rightarrow \quad A = 2, \quad B = 4$$

as expected, giving

$$\frac{6x+2}{(x-1)(x+3)} = \frac{2}{(x-1)} + \frac{4}{(x+3)}$$

```
convert((6*x+2)/((x-1)*(x+3)),parfrac,x);
```

$$2\frac{1}{x-1} + 4\frac{1}{x+3}$$

Thus if we want to evaluate the following integral, we first split it into partial fractions and then integrate each one separately.

$$\int \frac{6x+2}{(x-1)(x+3)}dx = \int \frac{2}{(x-1)}dx + \int \frac{4}{(x+3)}dx$$
$$= 2\ln(x-1) + 4\ln(x+3) + c$$

```
int((6*x+2)/((x-1)*(x+3)),x);
```

$$2\ln(x-1) + 4\ln(x+3)$$

Some things to remember:

The partial fraction method is used directly, if the degree (i.e., power) of the numerator polynomial is less than that of the denominator.

When the polynomial in the numerator has a power equal or higher than the denominator, use long division to split it into a part where the denominator can be divided out and a remainder which is further simplified by the use of partial fractions.

Example:

$$\frac{3x^2-11}{x^2+2x-3}=?$$

$$\begin{array}{r} 3 \\ x^2+2x-3\overline{)\,3x^2-11} \\ \underline{3x^2+6x-9} \\ -6x-2 \end{array}$$

$$\therefore \frac{3x^2-11}{x^2+2x-3}=3-\frac{6x+2}{x^2+2x-3}=3-\frac{6x+2}{(x-1)(x+3)}$$

$$=3-\frac{2}{(x-1)}-\frac{4}{(x+3)}$$

When one of the denominators in the partial fraction terms on the RHS (right hand side) is a polynomial of degree higher than one, which you cannot factor into simpler terms, the corresponding denominator is taken as a polynomial of degree one less, with unknown coefficients.

Example:

$$\frac{2x^2+3x+1}{\left(x^4-1\right)\left(x^4+1\right)}=\frac{A}{x-1}+\frac{B}{x+1}+\frac{Cx+D}{x^2+1}+\frac{Ex^3+Fx^2+Gx+H}{x^4+1}$$

If there is a repeated factor, such as $(x\text{-}3)^3$ in the denominator, you need to include several terms on the RHS, with increasing powers of that factor up to the highest one in the LHS, and with a simple numeric numerator in each.

Example:

$$\frac{4x^2+5x}{(x+3)^3}=\frac{A}{(x+3)^3}+\frac{B}{(x+3)^2}+\frac{C}{(x+3)}$$

6 First Order Differential Equations

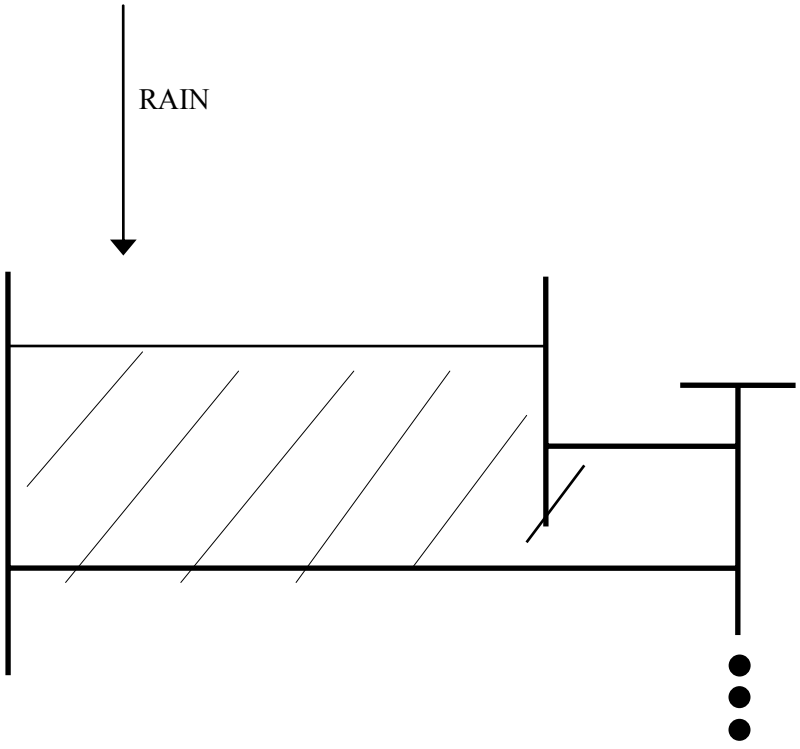
6.1 Introduction

A differential equation (DE) is an equation in which derivatives of an unknown function occur e.g.

$$\frac{dy}{dt} = 6t + 3$$

Why do we have differential equations? Derivatives measure **rates** and in a large number of processes we are interested in the accumulation of some quantity.

Example



The quantity Q in tank changes because of rain flowing in and water flowing out:

$$\frac{dQ}{dt} = \text{rain input rate - outflow rate}$$

Let us take some concrete numbers:

A 200 litre tank contains brine with a concentration of 2 grams/litre. Brine is input into the tank at a rate of 8 litres/minute with [I] (the concentration of brine being put into the tank) 4 grams/litre. The resultant liquor is drawn off at a rate of 8 litres/minute. Write the differential equation for this system.

Let $u(t)$ = quantity of salt in tank in grams

$$\text{Rate in} = \frac{8l}{\min} * \frac{4g}{l} = 32\ g/\min$$

$$\text{Rate out} = \frac{8l}{\min} * \frac{u(t)g}{200l} = 0.04u(t)\ g/\min$$

Thus $\frac{du}{dt} = 32 - 0.04u(t)$ for $t > 0$

We can solve this using Maple directly:

```
dsolve({diff(u(t),t)=32-4/100*u(t)},u(t));
```

$$u(t) = 800 + e^{(-1/25\,t)}\_C1$$

The solution contains an unknown constant C1. We try C1 = 70:

$$u(t) = 800 + 70e^{-0.04t}$$

We can establish that this is a solution by evaluating the left and right hand sides of the differential equation separately and confirming that they are equal. Firstly let us evaluate the LHS

$$\frac{du}{dt} = -0.04.70.e^{-0.04t} = -2.8e^{-0.04t}$$

And now to evaluate the RHS we substitute our solution into the differential equation

$$\begin{aligned} 32 - 0.04\overbrace{(800 + 70e^{-0.04t})}^{u(t)} &= 32 - 32 - 2.8e^{-0.04t} \\ &= -2.8e^{-0.04t} \\ &= \frac{du}{dt} \end{aligned}$$

as expected. But what about a different value for C1?

Let us try $u(t) = 800 + ce^{-0.04t}$ $\frac{du}{dt} = -0.04ce^{-0.04t}$

and substituting gives $32 - 0.04(800 + ce^{-0.04t}) = 32 - 32 - 0.04ce^{-0.04t} = -0.04ce^{-0.04t}$

So there are an infinite number of solutions depending on which value of c we choose.

6.2 Initial value problems

Why are there so many solutions? Because we have not used the initial conditions. We know that the initial concentration is 2 grams/litre, so this means that at t=0 there are $(0) = 200(2) = 400$ grams of salt in the tank. But

$$u(0) = 800 + ce^{-0.04.0} = 800 + c$$

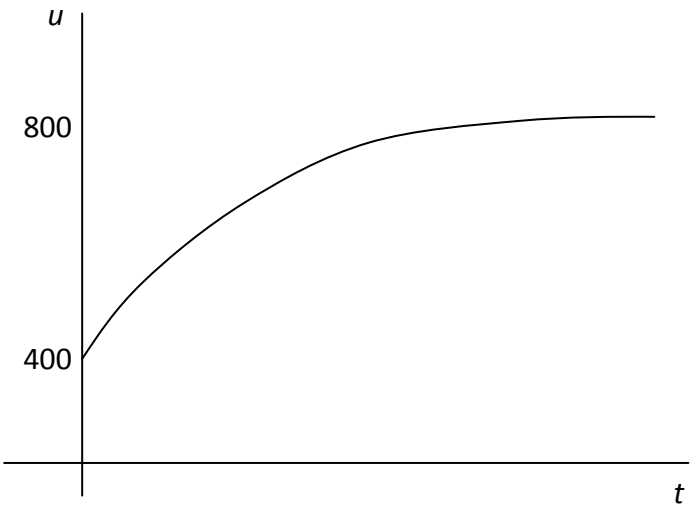
So c = -400 using the initial conditions. And thus we have the unique solution for this problem as

$$u(t) = 800 - 400e^{-0.04t}$$

Check using Maple, and also plot the solution:

```
dsolve({diff(u(t),t)=32-4/100*u(t),u(0)=400},u(t));
```

$$u(t) = 800 - 400\, e^{(-1/25\, t)}$$



- *An **initial value problem** consisting of a **first order differential equation** (DE), i.e. containing only derivatives of the form du/dt, together with an **initial condition** (IC) has a unique solution.*

$$\left.\frac{du}{dt} = F(t,u)\right\} \text{ differential equation}$$

$$\left.u(t_o) = q\right\} \text{ initial condition}$$

6.3 Classifying First Order Differential Equations

The simplest kind of 1st order differential equation has the form

$$\frac{du}{dt} = f(t)$$

To solve these types of equations (or initial value problems) we can sometimes just integrate it; e.g.

$$\frac{du}{dt} = e^{-2t} \quad ; \quad t > 0 \quad ; \quad u(0) = 5$$

Integrate each side of the differential equation to get

$$u = \int e^{-2t}\,dt = -\frac{e^{-2t}}{2} + c$$

Use the initial condition

$$5 = -\tfrac{1}{2} + c \qquad \Rightarrow \qquad c = \tfrac{11}{2}$$

So the solution is

$$u = \frac{11}{2} - \frac{e^{-2t}}{2}$$

dsolve({diff(u(t),t)=exp(-2*t),u(0)=5},u(t));

$$u(t) = -\frac{1}{2}e^{(-2t)} + \frac{11}{2}$$

But not all differential equations are that simple. Sometimes we **cannot** just integrate the RHS, for example because it contains the function *u* that is unknown until we have solved the equation. Consider our salt tank model

$$\frac{du}{dt} = 32 - 0.04u$$

This is an example of a **linear first order differential equation** (LFODE) which in general we can write as

$$\frac{du}{dt} + a(t).u = f(t)$$

It is **linear** because it does not contain squares or higher powers of u and u'; and it is **first order** because it does not contain second or higher order derivatives of u. We can further classify a LFODE to be homogeneous if $f(t) \equiv 0$. A general homogeneous LFODE is thus

$$\frac{du}{dt} + a(t).u = 0$$

In the general form, when $f(t) \neq 0$ then the equation is **nonhomogeneous** and $f(t)$ is called the **inhomogeneity**.

- *A **linear first order differential equation** (LFODE), has the form*

  $$\frac{du}{dt} + a(t).u = f(t)$$

  where $a(t)$ is a known function of t, while u is unknown.
- *If $f(t) = 0$ the equation is homogenenous*

Next, we discuss methods to solve LFODE's.

6.4 Separation of variables

6.4.1 Homogeneous 1st order DE

We solve equations of this type by **separating** the variables to get all the terms involving u on the LHS and all the terms involving t on the RHS

$$\frac{du}{dt} = -a(t).u$$
$$\frac{1}{u}.du = -a(t).dt$$

then **integrating** both sides $\ln(u) = A(t) + c$

where $A(t) = -\int a(t).dt$

and then **exponentiating** both sides

$$e^{\ln u} = u = e^{A(t)+c} = e^{c}.e^{A(t)}$$
$$\therefore u = k\,e^{A(t)}$$

Here c is just an integration constant; since it is still undetermined, e^c is just another undetermined constant and so in the last step we replaced it with a k to be determined e.g. from an IC.

For use in the next section, notice that we can write the equation for the homogenous solution in the following form:

$$u(t)\,e^{\int \alpha(t)dt} = k$$

As an example of this, we consider a homogeneous equation that we make from the inhomogeneous tank model system. We had

$$\frac{du}{dt} + 0.04u = 32$$
$$f(t) = 32 \neq 0$$

If we omit the nonhomogeneity $f(t)$ we have the following homogenous equation:

$$\frac{du}{dt} = -0.04u$$

Now applying the above methods (Separate, Integrate & Exponentiate) gives

$$\frac{du}{u} = -0.04.dt$$
$$\therefore \ln u = -0.04t + c$$
$$\therefore u = k\,e^{-0.04t}$$

This is OK thus far but we now need to know how to solve the total equation **including** $f(t)$.

6.4.2 Inhomogeneous 1st order D.E.

Remember the product rule

$$\frac{d}{dx}(fg) = f\frac{dg}{dx} + g\frac{df}{dx}$$

Now look at the left hand side of the LFODE:

$$\frac{du}{dt} + a(t).u$$

The function u appears in the LHS in a similar way as the g in the product rule. So can we make the LHS look like the derivative of a product? If we can, it would become easy to integrate the LHS that contains the unknown u, and that will help us solve the equation.

One thing missing at present is a factor in front of the du/dt. Any desired factor can be inserted by multiplying the LFODE by the factor. The idea is to choose this multiplying factor, φ(t), in such a way that the LHS becomes the derivative of some product u(t) Q(t). Both φ and Q are unknown functions, but we can determine them by setting

$$\phi\frac{du}{dt}+\phi a u=\frac{d}{dt}\left(u Q\right)=u\frac{dQ}{dt}+Q\frac{du}{dt}$$

This equation will be satisfied if we set the coefficients of u and du/dt in the first and last expressions to be equal, giving us two equations for the two unknowns φ and Q:

$$\varphi=Q;\quad \varphi\alpha=\frac{dQ}{dt}=\frac{d\varphi}{dt}$$

Eliminating Q we now have a new homogeneous LFODE for φ which we can solve as above to find

$$\varphi(t)=e^{\int\alpha(t)dt}$$

If we compare that with the solution that was found above for the general homogenous LFODE, we see that we can write

$$u(t)\phi(t)=k$$

In other words, the factor by which we can multiply the inhomogeneous LFODE so that we can integrate it, is the same as the factor appearing in the solution of the homogenous LFODE. It is no coincidence that these two solutions are related in this way. Because the LHS of the homogenous and inhomogeneous equations are the same, if we multiply the homogenous equation by the factor φ it becomes the derivative of the product (uQ) and because we have found φ = Q we can write the homogenous equation in the form

$$\frac{d}{dt}\left(u(t)\phi(t)\right)=0$$

and the solution to this is obviously just $u(t)\phi(t)=k$ as we found previously without explicitly using an integrating factor.

The implication of this is that when solving an inhomogenous LFODE, the first step is to solve the honmogenous one; then, we can identify an integrating factor from its solution and use this to multiply the inhomogenous equation so that it can be integrated too. The situation is analogous to what we encountered with matrix equations, where solution of the homogenous case was also the first step to solving the inhomogenous case.

6.4.3 Integrating factors

The method that we developed so far can be described by the following steps:

1. *Solve corresponding homogeneous equation to find the* ***integrating factor*** $\phi(t)$ *where the solution of the homogeneous equation is* $\phi(t)\cdot u(t)=c$
2. *Multiply the nonhomogeneous equation by the integrating factor. This makes the LHS the derivative of the product* $\phi(t)\cdot u(t)$
3. *Integrate both sides, getting on the LHS* $\phi(t)\cdot u(t)$
4. *Solve for u(t) by dividing through by* $\phi(t)$

Example 1

Solve $$x\frac{dy}{dx}+y=e^x \qquad ; \qquad x>0$$

1. Write the equation in the form
$$\frac{dy}{dx}+\frac{y}{x}=\frac{e^x}{x}$$

2. The homogeneous equation is thus
$$\frac{dy}{dx}+\frac{y}{x}=0$$
$$\therefore \frac{dy}{dx}=-\frac{y}{x}$$
$$\therefore \frac{dy}{y}=-\frac{dx}{x}$$

Integrate

$$\ln(y) = k - \ln(x)$$

Exponentiate

$$y = e^{k-\ln(x)} = e^k\, e^{-\ln(x)} = \frac{e^k}{x}$$

$$\therefore xy = e^k$$

From this we identify the integrating factor as x. From step 1 multiplying through by the integrating factor

$$x\frac{dy}{dx} + y = e^x$$

This has now turned the LHS into the derivative of a product **which is always the integrating factor multiplied by the variable of interest,** which in this case is y

$$\frac{d}{dx}(xy) = e^x$$

Thus

$$xy = \int e^x$$

$$\therefore xy = e^x + c$$

$$\therefore y = \frac{e^x + c}{x}$$

dsolve(x*diff(y(x),x)+y(x)=exp(x),y(x));

$$y(x) = \frac{e^x + \_C1}{x}$$

Check

$\frac{dy}{dx}$ = derivative of quotient

Set $f = e^x + c \quad ; \quad f' = e^x$

$g = x \quad ; \quad g' = 1$

Use $\frac{d}{dx}(\frac{f}{g}) = \frac{g f' - f g'}{g^2}$

gives $\frac{x.e^x - (e^x + c).1}{x^2}$

Thus $\frac{dy}{dx} = \frac{xe^x - e^x - c}{x^2}$

The LHS of the original equation is

$$x\frac{dy}{dx} + y$$

$$\therefore \frac{xe^x - e^x - c}{x} + \frac{e^x + c}{x} = \frac{1}{x}[xe^x - e^x - c + e^x + c]$$

$$= e^x$$

= RHS of the original

We can conclude that the solution is OK.

Example 2

Solve

$$\frac{dy}{dx}+3y=5$$

Put into homogeneous form and then **separate** and **integrate**

$$\frac{dy}{dx}+3y=0$$
$$\frac{dy}{dx}=-3y$$
$$\frac{dy}{y}=-3.dx$$
$$\ln(y)=-3x+c$$

Exponentiate

$$y=e^{-3x}.e^{c}$$
$$e^{+3x}.y=k$$
$$k=e^{c}$$

This gives us an **integrating factor** of e^{3x} and we now multiply the original equation by this integrating factor

$$e^{3x}\frac{dy}{dx}+3e^{3x}.y=5e^{3x}$$

We now know that the left hand side is the derivative of a product, made up of the variable, *y*, multiplied by the integrating factor e^{3x}, which gives the first equation below which then becomes the second equation after integration.

$$\frac{d}{dx}(ye^{3x})=5e^{3x}$$
$$ye^{3x}=\frac{5e^{3x}}{3}+c$$
$$y=\frac{5}{3}+ce^{-3x}$$

dsolve(diff(y(x),x)+3*y(x)=5,y(x));

$$y(x)=\frac{5}{3}+e^{(-3x)}\_C1$$

Example 3

Solve $$\frac{dy}{dx} - \frac{3y}{x} = 5x$$

Firstly, we consider only the LHS (the homogeneous bit). We then separate and integrate to get the integrating factor.

$$\frac{dy}{3y} = \frac{dx}{x}$$
$$\frac{dy}{y} = \frac{3.dx}{x}$$
$$\ln(y) = 3\ln(x) + c$$
$$\ln(y) = \ln(x^3) + c$$
$$y = x^3.e^c$$

The integrating factor is $\frac{1}{x^3}$

$$\frac{1}{x^3}.\frac{dy}{dx} - 3.\frac{y}{x}.\frac{1}{x^3} = 5x.\frac{1}{x^3}$$
$$\frac{d}{dx}(\frac{1}{x^3}.y) = \frac{5}{x^2}$$

Now integrate both sides

$$\frac{y}{x^3} = \int 5x^{-2}dx = \frac{5x^{-1}}{-1} + c = \frac{-5}{x} + c$$

and solve for y

$$y = -5x^2 + cx^3$$

dsolve(diff(y(x),x)-(3*y(x))/x=5*x,y(x));

$$y(x) = -5\,x^2 + x^3\_C1$$

Example 4

Solve $$\frac{dy}{dx} + y.\tan(x) = \sec(x)$$

Check for yourself that in this case the integrating factor is sec(x), so the equation reduces to

$$\frac{d}{dx}(\sec(x).y) = \sec^2 x$$

which is easily integrated using the integration formulas. Can you obtain the same solution as *Maple*?

```
dsolve(diff(y(x),x)+y(x)*tan(x)=sec(x),y(x));
```

$$y(x) = \sin(x) + \cos(x)\_C1$$

6.5 General Method for solving LFODE's.

Consider a general LFODE $\frac{dy}{dx} + P(x).y = Q(x)$

We do not really have to work through the steps of finding the integrating factor and then using it, on an ad hoc basis for each new equation. Instead, we can work with the equation in its general form and go through the steps once and for all to get a solution that can be applied directly to any specific problem that we encounter.

For the general equation, we want to find an **integrating factor** $R(x)$ so that when we multiply through we get the derivative of a product on the LHS.

$$R(x)\left(\frac{dy}{dx} + P(x).y\right) = R(x).Q(x)$$

Omitting the argument x for brevity we can write this as

$$R\frac{dy}{dx} + R.P.y = R.Q \tag{1}$$

For the LHS to be the derivative of a product, then the product must be $R\,y$ because the first term on LHS is R (dy/dx) and the derivative of $R\,y$ is

$$\frac{d}{dx}(R.y) = R\frac{dy}{dx} + y\frac{dR}{dx} \tag{2}$$

Thus by inspecting the second term of equations (1) & (2) we have that

$$y\frac{dR}{dx} = R\,P\,y$$

Thus after cancelling the y we have

$$\frac{dR}{dx} = R.P$$

Now by **separating** variables we get

$$\frac{dR}{R} = P.dx$$

which we can **integrate** to get

$$\ln R = \int P dx$$

and **exponentiating** to get R gives

$$R = e^{\int P.dx} = e^{\int P(x)dx}$$

Thus we have

$$\frac{d}{dx}\big(R(x).y\big) = R(x).Q(x)$$

Integrate this

$$e^{\int P(x).dx}.y = \int e^{\int P(x).dx}.Q(x).dx$$

So finally

$$y(x) = e^{-\int P(x).dx}\left[\int e^{\int P(x).dx} Q(x)\,dx + C\right]$$

Check that Maple gets the same:

```
dsolve(diff(y(x),x)+P(x)*y(x)=Q(x),y(x));
```

$$y(x) = e^{\left(-\int P(x)\,dx\right)} \int e^{\left(\int P(x)\,dx\right)} Q(x)\,dx + e^{\left(-\int P(x)\,dx\right)} \_C1$$

This gives us a very general method of solving first order ODE's. The formula above allows us to write down the solution for any functions *P(x)* and *Q (x)*, provided that we can integrate them as needed. This is quite an exceptional situation – for most more complicated differential equations, one can only find solutions if you know the specific functions that they contain. To summarise:

- *For the general inhomogenous LFODE*

$$\frac{dy}{dx} + P(x).y = Q(x)$$

the solution is

$$y(x) = e^{-\int P(x).dx}\left[\int e^{\int P(x).dx} Q(x)\,dx + C\right]$$

- *For the homogenous case Q(x) = 0 and this reduces to*

$$y(x) = C e^{-\int P(x).dx}$$

There is one simple special case of the general formula that occurs often enough that it is worthwhile to look at it specifically:

6.5.1 General Equation with Constant Coefficients

$$\frac{dy}{dx} + ay = b$$

The integrating factor (I.F.) is obtained by using the techniques of the previous page to get

$$e^{\int a.dx} = e^{ax}$$

The LHS of the DE becomes $e^{ax}.\frac{dy}{dx} + ae^{ax}.y = be^{ax}$

which we know is the derivative of a product, and the RHS is also multipled by the integrating factor to give

$$\frac{d}{dx}\left(e^{ax}.y\right)=b.e^{ax}$$

and upon integrating this it becomes

$$e^{ax}.y=\frac{be^{ax}}{a}+c$$

which if we divide through by e^{ax} (which we can always do because it is never = 0 for **any** value of x, gives

$$y=\frac{b}{a}+ce^{-ax}$$

dsolve(diff(y(x),x)+a*y(x)=b,y(x));

$$y(x)=\frac{b+e^{(-a\,x)}\_C1\,a}{a}$$

Example

Compare this with our salt tank model $\frac{du}{dt} + 0.04u = 32$

where $a = 0.04$, $b = 32$. Applying the formula above, the solution is

$$u = \frac{32}{0.04} + ce^{-0.04t} = 800 + ce^{-0.04t}$$

This is just the same as we found before when we calculated the integrating factor directly, and as done there the initial condition can be used to find c.

6.6 Applications to modelling real world problems

Note in the following examples that the first step to solving a real world problem is to formulate it as one or more equations; the examples discussed are typical of situations where the appropriate mathematical formulation is as a differential equation. Often, this step is the hardest one in the modelling process!

6.6.1 Radioactive decay

How does the number of radioactive atoms in a sample change as a result of radioactive decay?
In radioactive decay, a fixed proportion l of the radioactive atoms that remain in a sample decay into non-radioactive atoms of other elements per time unit. Let C represent the number of radioactive atoms; then this decay law is mathematically expressed as

$$\frac{dC}{dt} = -\lambda C \quad ; \quad t > 0 \quad ; \quad C(0) = C_0$$

This is a linear first order DE with constant coefficients; write it in the standard form to identify *a* and *b*:

$$\frac{dC}{dt} + \lambda.C = 0 \quad \therefore a = \lambda \quad ; \quad b = 0$$

Applying the general formula the solution is

$$C = \frac{0}{\lambda} + C_0 e^{-\lambda.t}$$

dsolve({diff(C(t),t)=-lambda*C(t), C(0)=C[0]},C(t));

$$C(t) = e^{(-\lambda t)} C_0$$

6.6.2 Population Growth

Year	Population	Immigration
1790	3.93	-
1800	5.31	-
1810	7.24	-
1820	9.64	-
1830	12.9	0.1
1840	17.1	0.6
1850	23.2	1.7
1860	31.4	2.6
1870	38.6	2.3
1880	50.2	2.8
1890	63.0	5.2
1900	76.0	3.7
1910	92.0	8.8
1920	106	6
1930	123	4
1940	132	1
1950	151	1
1960	179	3
1970	203	3
1980	227	4

From historical records, population figures in the USA are given in the table above.

Studies of other countries have found that the typical population growth in the first half of the 19th century due to natural birth, was 2.5%. Is that rate compatible with the figures for the USA over the period 1830 to 1860?

Constructing the model

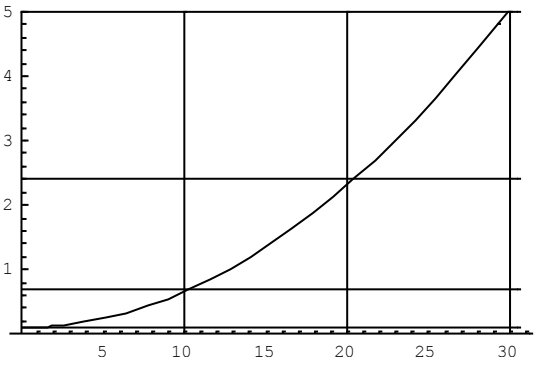
For the changing population $n(t)$, the birth rate contribution is easy to write down as $0.025\,n$.

However, the immigrants also have children, and to include that we need determine the **rate** at which immigrants entered the country. From the table, we can calculate that the total number of immigrants that were present over that period is as follows:

Year	Total Immigrants
1830	0.1
1840	0.7
1850	2.4
1860	5.0

If we plot this, taking 1830 as year 0, we get the figure shown, where the grid lines show the actual number of immigrants, I, from the table. This is clearly did not increase at a constant rate but as shown the numbers are fitted quite well by a parabola with the equation

$$I = 0.1 + 0.0055t^2$$



We can differentiate this to get the immigration rate as 0.011 t, and use that to write down the differential equation for our model that includes both contributions to the population growth rate:

$$\frac{dn}{dt} = 0.025n + 0.011t$$
$$n(0) = 12.9$$

Rewrite this in the standard form:

$$\frac{dn}{dt} - 0.025n = 0.011t$$

$\therefore$ The integrating factor is $e^{-0.025t}$ which gives

$$\frac{d}{dt}\left(e^{-0.025t}.n\right) = 0.011t.e^{-0.025t}$$

We can integrate the RHS using integration by parts to give

$$e^{-0.025.t}\,0.011\left(\frac{t}{-0.025} - \frac{1}{(0.025)^2}\right) + c$$

which gives

$$n(t) = -17.6 - 0.44t + ce^{0.025\,t}$$

Using $n(0)$ = 12.9 gives c = 30.5.

$$\therefore n(t) = -17.6 - 0.44t + 30.5e^{0.025t}$$

Now when we test this solution by evaluating this equation for t=20 we get

$$n(20) = 23.9$$

which compares favourably with the tabulated value for 1850 of 23.2.

See for yourself if you can get the same solution using the Maple instruction

dsolve({diff(n(t),t)=25/1000*n(t)+11/1000*t,n(0)=12.9},n(t));

6.6.3 Newton's law of cooling

From everyday experience, it is reasonable to assume that the rate at which a hot object cools down, is proportional to how much it is hotter than its surrounding. The mathematical formulation of this idea was first investigated by Sir Isaac Newton, and is given by

$$\frac{dT}{dt} = -k(T - 20)$$

where T is the temperature of the object, the temperature of the surroundings was taken to be a room temperature of 20 °C, and t is the time. The negative sign was inserted on the RHS so that with a positive proportionality constant k this would represent a temperature ***decrease*** *if the object is hotter than room temperature (i.e., it cools down) and an* ***increase*** *if the object is colder.*

The mathematical form of this model is the same as that for radioactive decay, even though the physical situation it describes is entirely different. So it is also solved by using the solution for an equation with constant coefficients:

$$\frac{dT}{dt} + k\,T = 20k$$

which means $a = k$, $b = 20\ k$ and the solution is

$$T = \frac{20\,k}{k} + c\,e^{-k\,t} = 20 + c\,e^{-k\,t}$$

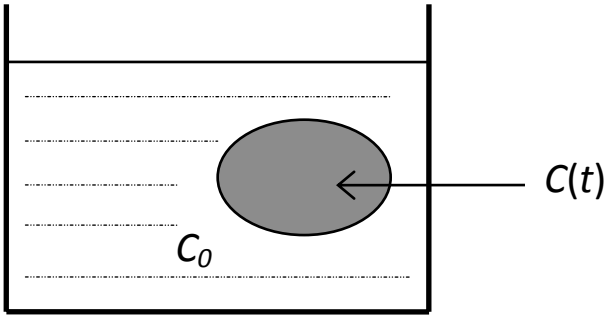
The solution still contains two unknown constants – k, as well as the integration constant c – and these could be determined if we know the temperature of the object at two different time values. However, note that the two constants play different roles in the model. The constant k determines the rate of cooling, and is a property of the object, e.g. would be different for a Thermos flask than for a metal or glass flask of the same size and shape. If we know enough about the properties of the materials the value of k might be calculated without actually heating up and measuring temperatures. On the other hand, c describes the particulars of the experiment, e.g. the initial temperature to which the object was heated.

So we see again that there is a difference between the mathematical perspective (where these are just constants, and essentially equivalent) and the perspective of the system for which we have set up a mathematical model.

Another significant difference is that while the mathematical solution just gives a number, for use in a mathematical model the quantities used are measured in units and for the example above the correct solution would be a temperature value T in degrees Celsius only.

6.6.4 Diffusion

In a cell, the volume is constant in an environment of solute concentration C_0. How does the concentration of solute inside the cell change with time, as it diffuses into the cell from the outside?



Define the variables:

$C(t)$ = intracellular concentration
$m(t)$ = mass of solute in cell
A = area of cell membrane
V = volume of cell

Relations between variables:

$$m(t) = V.C(t)$$

From **Ficks Law** of diffusion, the rate at which solute diffuses into the cell is proportional to the difference in concentration:

$$\frac{dm}{dt} = kA(C_0 - C)$$

where k = permeability of the cell membrane.

Now we combine the two equations to eliminate m so that we only have one unknown function of time:

$$\frac{dm}{dt} \qquad \frac{dC}{dt}$$

$$\therefore \frac{dC}{dt} = \frac{kA}{V}(C \ - C)$$

Separate the equation

$$\frac{dC}{(C_0 - C)} = \frac{kA}{V}dt$$

Integrate

$$-\ln(C_0 - C) = \frac{kA}{V}t + \tilde{c}$$

Exponentiate

$$C_0 - C = K^{'}.e^{-\frac{kA}{V}.t}$$
$$\therefore C = C_0 + K^{'}e^{-\frac{kA.t}{V}}$$

Here $K^{'}$ = the constant of integration.

dsolve(diff(C(t),t)=k*A/V*(Co-C(t)),C(t));

$$\mathrm{C}(t) = Co + \mathrm{e}^{\left(-\frac{k\,A\,t}{V}\right)}\_C1$$

Work out for yourself what this means for the following initial conditions, i.e. when *C(0)* is related to C_0 (the concentration in the environment) as follows:

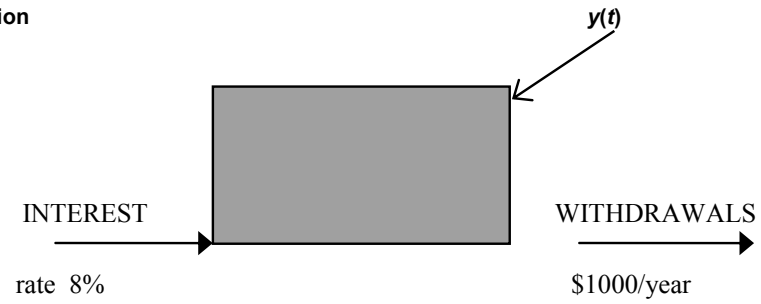
(a) $C(0) = C_0$ (b) $C(0) > C_0$ (c) $C(0) < C_0$

NB – *C(0)* is the cellular concentration at the time $t = 0$, which is something different from C_0!

6.6.5 Bank account balance

An account contains $10,000 and accrues interest at 8% compounded continuously. Frequent withdrawals are made from the account at the overall rate of $1000 per year.
When will there be only $1500 in the account?
How much will have been withdrawn from the account by that time?

Solution



Let y(t) represent the amount in the account; the initial balance is $y(0)$=10,000

Now	$y'(t)$ = Rate in - Rate out = Rate in - 1,000 Rate in = $0.08.y(t)$
Thus	$y'(t) = 0.08.y(t)-1000$
Rewriting this gives	$y'(t) - 0.08.y(t)=-1000$
This is of the form	$\frac{dy}{dt} + ay = b$
whose solution is	$y = \frac{b}{a} + ce^{-at}$
Thus	$y(t) = \frac{-1000}{-0.08} + ce^{+0.08t} = 12500 + ce^{0.08.t}$

dsolve(diff(y(t),t)=8/100*y(t)-1000,y(t));

$$y(t) = 12500 + e^{(2/25\,t)}\_C1$$

But $y(0) = 10{,}000$

Thus $10000 = 12500 + ce^{0.08.0}$

$= -2500$

$y(t) = 12500 - 2500e^{0.08}$

dsolve({diff(y(t),t)=8/100*y(t)-1000, y(0)=10000},y(t));

$$y(t) = 12500 - 2500\, e^{(2/25\,t)}$$

There will be \$1500 in the account when $1500 = 12500 - 2500\, e^{0.08t}$

$$\therefore e^{0.08t} = \frac{11000}{2500}$$

Taking logarithms gives $0.08t = \ln\left(\frac{110}{25}\right)$; $t = 18.52$ years

t=evalf(solve(1500=12500-2500* exp(2*t/25),t));

$$t = 18.52005676$$

By this time \$ 18,520 will have been withdrawn from the account.

6.6.6 Spread of Infection – the logistic equation.

How will the number of people that are infected increase when we introduce one infected person into a population of size n?

We need to take into account that only people who are not yet infected are susceptible to be infected; so introduce the variables

$x = x(t)$, number of susceptibles
y = y(t), number of infectives
n = the total population into which we introduce 1 infective

Relationships: $x + y = n + 1 \quad ; \quad for\ all\ t > 0$

The rate of increase of infectives is $\frac{dy}{dt} = \beta\, y\, x$

Now $\frac{dy}{dt} = \beta.y.(n + 1 - y)$

To simplify the notation we let n+1 = k, giving $\frac{dy}{dt} = \beta.y.(k - y)$

This equation is often called the logistic differential equation. It is not, in fact, a linear equation; but it is simple enough that it can be solved by **separating** the variables $\frac{dy}{y(k-y)} = \beta.dt$

In order to integrate this, we need to split the LHS of the equation into partial fractions

$$\frac{1}{y(k-y)} = \frac{A}{y} + \frac{B}{(k-y)} = \frac{A(k-y) + By}{y(k-y)}$$

Equate the coefficients of y^0 and y^1 in the numerators: $1 \equiv Ak + y(B - A)$; $0 \equiv B - A$

Thus $A = \frac{1}{k}$ and $B = \frac{1}{k}$

convert(1/(y*(k-y)),parfrac,y);

$$\frac{1}{k\,y} - \frac{1}{k\,(-k + y)}$$

Thus

$$LHS = \int \frac{dy}{ky} + \int \frac{dy}{k(k - y)}$$

$$\therefore \frac{1}{k}\left[\ln(y) - \ln(y - k)\right] = \frac{1}{k}\ln\left(\frac{y}{y - k}\right)$$

But the RHS = β *dt*; integrating gives β *t*+*c*, and so

$$\ln\left(\frac{y}{y - k}\right) = \beta k t + c k$$

Exponentiate

$$\frac{y}{(y - k)} = Ce^{\beta k t} \quad \text{where} \quad C = e^{ck}$$

$$y = y\,C\,e^{\beta k t} - k\,C\,e^{\beta k t}$$

$$\therefore y = \frac{k\,C\,e^{\beta k t}}{C\,e^{\beta k t} - 1}$$

Dividing this equation through by $e\beta^{kt}$ and introducing a new arbitrary constant D = 1/C this becomes

$$y = \frac{k}{1 - De^{-\beta k t}}$$

dsolve(diff(y(t),t)=beta*y*(k-y),y(t));

$$\frac{1}{y(t)} = \frac{1 + e^{(-\beta\,k\,t)}\_C1\,k}{k}$$

The Maple result looks a little different, but is really the same since both C1 and D are just arbitrary constants, including the sign!

At $t = 0$, $y = 1$ and so

$$1 = \frac{k}{1 - D}$$

$$\therefore D = 1 - k = -n \quad ; \quad k = n + 1$$

Thus

$$y = \frac{n+1}{1 + n.e^{-\beta.(n+1).t}}$$

This is an important result known as the **Logistic Law**. Sketch a plot of its behaviour and check if it is a sensible model of the real world system!

```
simplify(dsolve({diff(y(t),t)=beta*y*(n+1-y),y(0)=1},y(t)));
```

$$y(t) = \frac{n + 1}{n\, e^{(-t\, \beta\, (n+1))} + 1}$$

6.6.7 Marketing strategy

A manufacturer expects to sell a product to 150,000 potential customers. Let y(t) be the number of customers buying the product by month t after its introduction. The manufacturer knows that sales will be determined, not only by advertising, but by how many customers are already using the product and how many are not. Marketing surveys suggest that the rate of customers attracted to the product is ruled by the differential equation

$$\frac{dy}{dt} = 0.004\, y(150 - y)$$

representing both the number of customers buying the product, y, and the number of potential customers yet to buy it, 150-y. The manufacturer initially gives away 250 copies of the product, so that y(0) =0.25 (thousand).

In how many months will the company have sold products to 90% of the anticipated market?

Solution

Since 90% of the 150 thousand potential customers is 135,000 we are thus asking at what value of t will y have the value of 135 (thousand). From the solution above we get using β = 0.004, k=150 that the coefficient in the exponential term is 0.6 = 0.004 * 150.

$$y = \frac{150}{1 + Ce^{-0.6t}}$$

Now we have to find C. To do this we use the initial condition that at time t=0, we had y=0.25 (thousand). So

$$0.25 = \frac{150}{1 + C}$$

which gives C=599. So to find when the market is 90% saturated we have

$$135 = \frac{150}{1 + 599e^{-0.6t}}$$

Rearranging gives

$$1 + 599e^{-0.6t} = \frac{150}{135}$$

which gives

$$e^{-0.6t} = 0.00018364$$

which taking logarithms gives

$$-0.6t = \ln(0.00018364)$$

and so finally $t = 14.33$ months.

So after this quite short time the market is essentially saturated for this product. The use of the logistic equation is quite common in product life-cycles.

6.7 Characterising Solutions Using a Phase Line

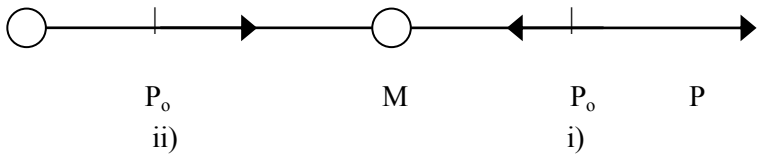
Graphical methods can often give information about the solution even if we have not yet found it in the form of a formula.

Take, for example, a logistic equation given by $\frac{dP}{dt} = kP(M - P)$

If we plot the solution to this that was found in section 6.6.6 as a function of time, it is seen that the fixed value M is asymptotically approached as time increases. The way that this is approached depends on the initial value P_0 of P:

1. $P_o > M$ then $P \to M$ from above
2. $P_o < M$ then $P \to M$ from below

This is represented graphically on a phase line as follows



This behaviour can be read directly from the differential equation, without actually solving it. Consider an initial value of type (i) above. At $P = P_0$, the RHS of the equation will be negative; so the LHS shows that P will decrease. However as long as P stays above M, the derivative remains negative so P decreases further, but at a slower rate. If we had chosen an initial value $P_0 = M$, the derivative is zero so in that case it does not change and stays at the value M forever. For the same reason, the P value that approaches M from above can never reach it, only apporoaches it asymptotically. A similar reasoning can be applied to the initial value case (ii); there the RHS and hence the derivative on the left is positive so the value M is approached from below. We describe M as a **stable point.**

Consider the equation $\frac{dP}{dt} = kP(M - P)(P - m)$

This equation is harder to solve than the logistic equation (try solving it yourself using Maple!). But a similar reasoning as above can easily be applied to the differential equation directly; only we now have 3 different cases for $P_0 < m$, $m < P_0 < M$ and $P_0 > M$ respectively.

The phase line representation in this case is

Now *m* is an **unstable point** while *M* is a **stable point.** Both of these are sometimes called **fixed points** of the DE.

6.8 Variation of Parameters method

We saw in this chapter that a linear first order differential equation or LFODE can be solved quite generally using the separation of variables method. However, in more complicated cases we will need to resort to other methods, and one of these is called **variation of parameters**. Although we did not need it for the equations treated so far, the case of 1st order linear DE's is a simple one to illustrate the method and show that it is in fact equivalent to separation of variables in this case.

Looking back at the introduction of separation of variables, we first solved the homogeneous equation and used its solution to help us find a solution for the non-homogeneous equation. This approach is also the basis of the variation of parameters method. Suppose we consider the general non-homogeneous equation:

$$\frac{dy}{dx} + y\,P(x) = Q(x)$$

The corresponding homogeneous equation is

$$\frac{dy}{dx} + y\,P(x) = 0$$

and the solution to this we have previously found to be

$$y = A\,e^{-f(x)}$$

where *f(x)* is found from the equation we were given as

$$f(x) = \int P(x)\,dx$$

In this solution, the constant A was just a constant of integration, that we would normally determine from the initial conditions. Now consider the solution that we found for the non-homogeneous equation. We can write that in the form

$$y = e^{-f(x)}g(x) = g(x)e^{-f(x)}$$

where we have once again introduced a new function, g(x) this time, that is calculated from the function $Q(x)$ that we were given in the non-homogenous equation:

$$g(x) \quad \int e^{f(x)} Q(x)\,dx$$

If we compare the two solutions, we see that the solution to the non-homogenous equation looks quite a lot like the one for the homogeneous equation, only the integration **constant** A in the homogenous solution was replaced by a **function** $g(x)$ in solving the non-homogeneous one.

That is the basic idea of the variation of parameters method: we solve the homogeneous equation, then replace all integration constants with new, unknown variables (hence **variation** of parameters) and put the resulting expression back into the non-homogenous equation and try to solve it for the new variables.

Let us do that last step for the linear 1st order equation. We pretend that we do not know yet what the expression for *g(x)* is, but find it by substituting *y* back into the non-homogenous equation:

$$\frac{d}{dx}\left[g(x)e^{-f(x)}\right]+\left[g(x)e^{-f(x)}\right]P(x)=Q(x)$$
$$\therefore g(x)\left[\frac{d}{dx}e^{-f(x)}+e^{-f(x)}P(x)\right]+e^{-f(x)}\frac{dg}{dx}=Q(x)$$

But the expression in the square brackets is zero because it is in fact just the left hand side of the homogeneous equation. It is easy to see that what remains of the equation above is very simple to integrate and one gets the solution for *g(x)* just as it was found before.

Notice that the fact that a large part of the equation just dropped away was no coincidence – when we changed the *A* to a function, we created a product; the product rule of differentiation ensures that when that is put back in the equation there will be two terms coming from the derivative, and since one of the factors in the product was just the solution of the homogeneous equation, one of those two terms are bound to exactly cancel all the other terms (apart from the derivative term) that were on the left of the homogeneous equation and that also occur in the non-homogeneous equation.

So applying variation of parameters is always guaranteed to remove some terms from the non-homogenous equation – if it does not happen, you must have made a mistake! What is not guaranteed, is whether you will be able to solve the equation that remains. In many cases, like the one above, the equation that remains is simpler than the one you started with and can be solved. However, in some cases you might still have an equation that is hard to solve, and then variation of parameters simply did not help.

6.9 The Main Points Again – A stepwise strategy for solving FODE's.

6.9.1 Step 1 – standard form

Get the differential equation into the standard form: $\frac{dy}{dx}+y\,P(x)=Q(x)$

It is not important which symbols are used for the variables, but remember that *y* represents the unknown function that you are trying to find, and *x* is the differentiation variable.
The important things are:

- *The first term on the LHS of the equation is a first order derivative, and it has no coefficient (remove by dividing if necessary).*
- *The only other term that contains y, is also on the LHS, and this only contains y to the 1st power.*
- *Any terms that do not contain y (but may contain x) are on the RHS.*

In addition to ordinary algebraic manipulation of the equation to get it into the standard form, it is sometimes useful to transform to another variable. In particular, notice that if we add a constant to y (i.e. put $u(x) = y(x) + k$), the derivative is not changed. Simple non-homogeneous equations can often be changed into homogeneous this way – for example, if it happens that $Q(x) = k\,P(x)$.

6.9.2 Step 2 – constant coefficients

Are $P(x)$ and $Q(x)$ constants, say $P(x) = a$ and $Q(x) = b$?

Then we can take a shortcut and write down the solution immediately: $y(x) = \frac{b}{a} + c\,e^{-ax}$

Here c is an integration constant determined from the initial conditions, and you are finished.

ELSE

6.9.3 Step 3 – homogeneous or non-homogeneous equation?

Is the RHS zero? Then it is a homogeneous equation, otherwise non-homogeneous. If it is a non-homogeneous equation, write down the homogeneous equation anyway by replacing $Q(x)$ by zero.

6.9.4 Step 4 – solve homogeneous equation.

There will only be two terms on the LHS. Move one of them to the RHS, and separate all the y's on one side and all the x's on the other, by cross-multiplying in such a way that the numerator on one side is proportional to dy and on the other to dx.

Integrate each side, remembering to add a integration constant.

At least one of the integrals will be a *ln* function. Get rid of it by taking exponents. This will change the additive integration constant into a multiplying constant, i.e. a coefficient c.

If the original equation was homogeneous, you are done – just find c from initial conditions; else go on to

6.9.5 Step 5 – integrating factor

Divide through so that you are only left with c on the RHS and the LHS is a product. This product is used for two separate purposes:

First, the part of it that multiplies y is the integrating factor. Identify this factor, and multiply the original non-homogeneous equation through by it.

Doing this, the LHS of the equation will always just be the derivative (d/dx) of the product. Integrating the LHS will therefore immediately give you the product – the 2nd purpose.

All you still need to do, is to integrate the RHS, tidy up the equation and find the integration constant from initial conditions.

6.9.6 Non-linear equations

These are usually too hard to solve, but we came across one exception – the logistic equation, e.g. in the spread of infection example. Notice the extension of this idea to similar, but more complicated equations, that we used to illustrate the phase line representation.

In these cases, go straight to step 4 above – separating variables.

Apply a partial fraction expansion to the LHS to convert it to a form that can easily be integrated.

7 General Properties of Solutions to Differential Equations

7.1 Introduction

So far, we have seen how to solve 1st order differential equations. It would be a logical next step to go on to 2nd order equations i.e. containing a second derivative, as many textbooks do. From there, even higher order equations could be studied. However, as we will see in the next chapter, all of these can in fact be reduced to systems of 1st order equations and so in this course we will limit ourselves to that strategy.

However, before doing that it is good to stand back and get a brief overview of some properties that can be expected from such equations. We will only discuss these, without going into the mathematical proofs.

In solving a 1st order equation, we needed to integrate it once, and got one solution containing one integration constant that was determined from an initial condition. For a 2nd order equation, it has to be integrated twice, and this gives two distinct solutions, with two integration constants. Hence we also need two initial conditions to determine them, usually both the value of the unknown function and its derivative at $t = 0$.

A simple example is the following 2nd order homogenous equation with constant coefficients:

$$\frac{d^2y}{dt^2} + w^2 y = 0$$

This has two solutions: $y = cos\ (wt)$ or $y = sin\ (wt)$, as you can easily test by substitution. Then the most general solution is obtained by just superimposing these solutions, just as we did for vectors solving algebraic equations:

$$y(t) = A\cos(wt) + B\sin(wt)$$

Here the A and B are the integration constants that are found from two initial conditions.

If we want to solve the inhomogeneous equation

$$\frac{d^2y}{dt^2} + w^2 y = f(t)$$

that can be done by applying the same idea that we used for inhomogeneous LFODE's and algebraic equations. We just need any particular solution of the inhomogeneous equation, and add that to the homogeneous solution stated above.

So how can we find a particular solution? There are various ways, but the most straightforward is variation of parameters. In other words, we use the homogeneous solution, change the integration constants into functions of *t* and substitute this into the inhomogeneous equation.

With this brief preview of what to expect, we now proceed to list the general results. These are general in the sense that they apply to equations of any order, provided that the equation remains **linear**. By linear, we mean that *y* as well as any derivatives of *y* may only occur as first powers – no quadratic terms or functions, etc. These results are not proven here – refer to a textbook if you want to see that – but you have seen examples of these in the cases that we studied so far.

7.2 Homogenous Linear Equations

7.2.1 Principle of Superposition

- *If $y_1(x)$ and $y_2(x)$ are solutions of a linear homogenous DE,*
- *Then any linear combination $c_1 y_1(x) + c_2 y_1(x)$ is also a solution.*

7.2.2 Number of solutions

- *An N-th order linear homogenous DE has exactly N solutions that are linearly independent.*

By substitution we can find that the 2nd order equation

$$\frac{d^2y}{dt^2} + w^2 y = 0$$

is solved by any of the following expressions:

$$e^{i\omega t}, e^{-i\omega t}, \cos(\omega t), \sin(\omega t), \cos(\omega t + \phi)$$

So here are apparently 5 different solutions – but they are not linearly independent.

7.2.3 Linear independence of functions

- *The set of functions $y_1(x)$, $y_2(x)$, ... $y_n(x)$ are* ***linearly independent*** *if and only if* ***for all x***

$$c_1 y_1(x) + c_2 y_2(x) + \ldots + c_n y_n(x) = 0$$

only *occurs if $c_1=c_2=\ldots=c_n=0$.*

- *This is tested by checking that the Wronskian determinant $W \neq 0$*

This is the same idea as linear independence for vectors, if we consider a function to be a vector with an infinite number of components:

$$y(x) = \begin{bmatrix} y(x_1) \\ y(x_2) \\ \vdots \end{bmatrix}$$

where x_1, x_2, … are all the possible values that x can take.

The Wronskian

To test if the functions are independent, take any *x*; since the equation holds also for each neighbouring point, by subtraction it will hold for the derivative as well:

$$c_1\, y_1'(x) + c_2\, y_2'(x) + \ldots + c_n\, y_n'(x) = 0, \quad \text{for all x}$$

Repeat this *n* times to get a set of *n* equations for the n constants c_n. But this set can only have the desired trivial solution provided that the so-called "Wronskian" determinant *W* is not zero, for all *x*:

$$W = \begin{vmatrix} y_1(x) & y_2(x) & \cdots & y_n(x) \\ y_1'(x) & y_2'(x) & \cdots & y_n'(x) \\ \vdots & \vdots & \vdots & \vdots \\ y^n{}_1(x) & y^n{}_2(x) & \cdots & y^n{}_n(x) \end{vmatrix}$$

Example

Try this rule on the solutions of the 2nd order equation that were mentioned above:

$$W = \begin{vmatrix} \cos(\omega t) & \sin(\omega t) \\ -\omega \sin(\omega t) & \omega \cos(\omega t) \end{vmatrix} = \omega \left[\cos^2(\omega t) + \sin^2(\omega t) \right] = \omega$$

So this has the same value *w* which is not zero for any *t* value, and we can conclude that the solutions *cos(wt)* and *sin(wt)* are linearly independent. In the same way you can prove for yourself that for the pair of solutions *exp(iwt)* and *exp(-iwt)* the Wronskian is (-2*w*) which is not zero and so these two solutions are also linearly independent.

By contrast, however, see what happens when we take *exp(iwt), sin(wt)* and *cos(wt)*:

$$W = \begin{vmatrix} e^{i\omega t} & \sin(\omega t) & \cos(\omega t) \\ i\,\omega\, e^{i\omega t} & \omega \cos(\omega t) & -\omega \sin(\omega t) \\ -\omega^2\, e^{i\omega t} & -\omega^2 \sin(\omega t) & -\omega^2 \cos(\omega t) \end{vmatrix} = 0$$

This expression is 0 because the first and last rows of the determinant are identical, so these three solutions are **not** linearly independent.

7.2.4 General solution

- *All solutions of a N-th order homogenous DE can be written as a linear superposition of any set of N linearly independent solutions.*
  This is the general solution.

- *N initial conditions are necessary and sufficient to give a unique solution.*

7.2.5 Non-homogenous linear equations

Suppose that

- *$y_p(x)$ is a solution of the non-homogenous linear DE.*
- *$y_1(x)$, $y_2(x)$, ... $y_n(x)$ are solutions of the homogenous equation*
- *Then*

$$y(x) = y_p(x) + c_1 y_1(x) + c_2 y_2(x) + \ldots + c_n y_n(x)$$

is the most general solution of the non-homogenous DE..

This means that to solve an *N*-th order non-homogenous equation, we need to

1. Find a set of *N* linearly independent solutions of the associated homogenous equation.
2. Find only one solution of the non-homogenous equation; $y_p(x)$, called the particular solution; $y_p(x)$ contains no integration constants.
3. Find the *N* integration constants $c_1, c_2, \ldots c_n$ from *N* initial conditions.

8 Systems of Linear Differential Equations

8.1 Introduction

There are two reasons why we want to be able to solve systems of 1st order linear DE's:

- Many physical systems are described by a system of equations, in particular **compartment models** where the system is broken up into separate "compartments", each of which is described by a linear DE.
- Every n^{th} order differential equation can be converted to a system of n first order differential equations.

Let us consider an example of each case.

Example 1: A higher order equation

$$\frac{d^2u}{dt^2} + 2\frac{du}{dt} + 3u = f(t)$$

Let $x_1 = u$ and $x_2 = \frac{du}{dt}$

Then

$$\frac{dx_1}{dt} = x_2$$
$$\frac{dx_2}{dt} = \frac{d^2u}{dt^2}$$
$$= f(t) - 3x_1 - 2x_2$$

We can write this in the following way

$$\frac{dx_1}{dt} = 0x_1 + x_2 + 0$$

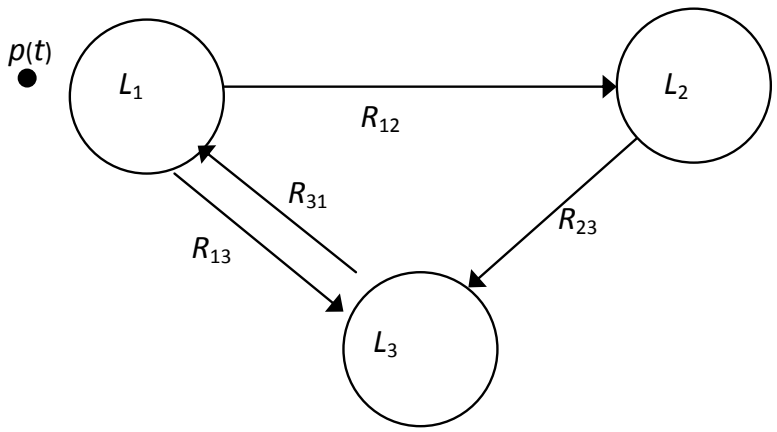
$$\frac{dx_2}{dt} = -3x_1 - 2x_2 + f(t)$$

and thus we can write this in matrix notation as

$$\frac{d\underline{x}}{dt} = \begin{bmatrix} 0 & 1 \\ -3 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ f(t) \end{bmatrix} = A\underline{x} + \underline{f}$$

If $\underline{f} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ then the system is called **homogeneous**.

Example 2 – A compartment model: A system of 3 lakes



Suppose we have a system of lakes, with a source *p(t)* feeding into lake 1, and with flows between the lakes as indicated in the diagram. Let the total water volume of lake *i* be denoted by the variable y_i, and we assume that the flow R_{ij} is proportional to the water volume in the lake that feeds it.

From this we get for the volume of Lake 1

$$\begin{aligned}\frac{dy_1}{dt} &= -R_{12} - R_{13} + R_{31} + p(t) \\ &= -k_{12}\, y_1 - k_{13} y_1 + k_{31} y_3 + p(t)\end{aligned}$$

and similar equations for the other lakes. Taken together, this forms a matrix system

$$\underline{y}' = A\underline{y} + \underline{f}$$

where

$$\underline{f} = \begin{bmatrix} p(t) \\ 0 \\ 0 \end{bmatrix}$$

and so the system is homogeneous if there is no external source (i.e., *p(t)* = 0).

As before, we start in either case by solving the homogenous equation system.

8.2 Homogenous Systems

We are now interested in solving a system where *A* is an (*n* x *n*) matrix:

$$\frac{d\underline{x}}{dt} = A\underline{x}$$

8.2.1 Method 1: Direct substitution

If the equation was a scalar equation, the solution would just have been $x = c\, e^{At}$. Let us try an analogous solution to the vector equation, but since we cannot put a matrix in the exponent we put an unknown scalar number λ that is to be determined, and change the integration constant into a constant vector to make sure that the equation conforms:

$$\underline{x}(t) = e^{\lambda t}\underline{z}$$

where *z* is a $n \times 1$ column matrix. Written out, the equations are

$$\begin{aligned}\frac{dx_1}{dt} &= a_{11}x_1 + a_{12}x_2 \\ \frac{dx_2}{dt} &= a_{21}x_1 + a_{22}x_2\end{aligned}$$

Note that these equations are **coupled**, i.e the derivative of x_1 is determined by both x_1 itself and the other variable(s), in this case x_2. Also note that this happens because the coefficient matrix A contains non-zero elements that are not on its diagonal, in this case a_{12} and a_{21}. The substitution we try, assumes that both variables share the same exponential time dependence factor, i.e.:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = e^{\lambda t} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

a constant vector

Therefore by differentiating and substituting we have

$$\begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \end{bmatrix} = \lambda e^{\lambda t} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

Thus by comparison

$$\lambda e^{\lambda t} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = A\underline{x}$$

$$= A e^{\lambda t} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

Thus, dividing by $e\lambda^{t}$ one gets

$$\lambda.\underline{z} = A.\underline{z}$$

and this is we recognize as the eigenvalue equation for A!

From the solution of the eigenvalue problem, we will in this case get two eigenvalues, each with its own eigenvector. Each of these represents a solution, as we expect for the 2nd order equation that we started with. Superposing these two solutions, with coefficients to be determined from the initial conditions, gives the final solution.

Example 1

Take the system representing the 2nd order equation in the introduction section. The coefficient matrix is:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -3 & -2 \end{bmatrix}$$

and to find its eigenvalues we construct the matrix

$$\begin{bmatrix} -\lambda & 1 \\ -3 & -2-\lambda \end{bmatrix}$$

Find the determinant

$$\begin{aligned} \det(A-\lambda.I) &= (-\lambda)(-2-\lambda)-(-3)(1) \\ &= 2\lambda + \lambda^{2} + 3 \\ &= \lambda^{2} + 2\lambda + 3 \end{aligned}$$

The eigenvalues are

$$\lambda = \frac{-2 \pm \sqrt{4-12}}{2}$$

$$= -1 \pm \sqrt{2}.i$$

The eigenvalues are complex – because A was not a symmetric matrix -, which means that also eigenvectors are complex, and we will not calculate them for this example. But check what Maple produces:

```
with(linalg): A:=matrix(2,2,[0,1,-3,-2]);
```

$$A := \begin{bmatrix} 0 & 1 \\ -3 & -2 \end{bmatrix}$$

```
lambda:=eigenvals(A);
```

$$\lambda := -1 + I\sqrt{2}, -1 - I\sqrt{2}$$

8.2.2 Method 2: Decoupling

Direct substitution into the coupled equations led in a natural way to an eigenvalue problem. Knowing that this is what we get anyway, we can make use of the eigenvector expansion to **decouple** the equations from each other first, just as we did with algebraic equations. We start with the eigenvector matrix P of the coefficient matrix A, and which by definition can be written as follows using the diagonal matrix D of eigenvalues:

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

Remember that the columns of P are the eigenvectors of A; so they satisfy the matrix equation

$$\mathbf{A\,P} = \mathbf{P\,D}$$
$$\mathbf{P^{-1}A\,P} = \mathbf{D}$$

If we multiply the matrix form of the DE system from the left by P^{-1} and insert the unit matrix $I=P\,P^{-1}$ at a convenient point into the result, we get

$$\mathbf{Y'} = \mathbf{A\,Y}$$
$$\mathbf{P^{-1}\,Y'} = \mathbf{P^{-1}\,A\,Y} = \mathbf{P^{-1}\,A\,P\,P^{-1}Y}$$

Now we recognize that the first part of the right hand side is just D, and if we define a new vector of unknowns $U=P^{-1}Y$ instead of Y, we can write this equation as

$$\mathbf{U'} = \mathbf{D\,U}$$

which is almost the same form as the one just before that Y satisfies, but with the big advantage that instead of just any old matrix A we have a diagonal matrix D on the right hand side. That means that the equations for the components of U are decoupled from each other, in other words we can forget that they form part of a system of equations, and solve each one separately using the methods we learned earlier. Once we have found the solutions for the U's, we can go back to the solutions for Y that we really want, just by performing a matrix multiplication, as follows from the way we defined U;

$$\mathbf{Y} = \mathbf{P\,U}$$

Because the elements of Y represent the solutions, and we will superimpose them with arbitrary coefficients, we see from the equation above that it would be a waste of time to normalise the eigenvectors – it would only give us a different set of coefficient values in the end.

To demonstrate this method, we first take an example that is simpler than the A we showed before, because the eigenvalues are real, not complex.

Example 2

$$y_1' = y_1 + y_2$$
$$y_2' = 4y_1 - 2y_2$$

Note that these equations for the y's are coupled. The initial conditions are

$$y_1(0) = 1$$
$$y_2(0) = 6$$

The coefficient matrix is

$$A = \begin{bmatrix} 1 & 1 \\ 4 & -2 \end{bmatrix}$$

This will be diagonalised by P whose columns are the linearly independent eigenvectors of A.

$$D = P^{-1}.A.P$$

The elements of D are the eigenvalues of A. So, we are back to solving the eigenvalue problem for A.

1. Find the eigenvalues of A $\det(A-\lambda.I) = \begin{vmatrix} 1-\lambda & 1 \\ 4 & -2-\lambda \end{vmatrix}$

$$= (1-\lambda\ )(-2-\lambda) - (4)(1)$$
$$= -2 + 2\lambda\ - \lambda\ + \lambda^{\,2} - 4$$
$$= \lambda^2 + \lambda\ - 6$$
$$= (\lambda\ + 3)(\lambda\ - 2)$$

A:=matrix(2,2,[1,1,4,-2]); lambda:=eigenvals(A);

$$A := \begin{bmatrix} 1 & 1 \\ 4 & -2 \end{bmatrix}$$
$$\lambda := 2, -3$$

The eigenvalues are 2 and -3.

2. Find the eigenvectors of A $A.\underline{x} = \lambda.\underline{x}$

When λ = 2

$$\begin{bmatrix} 1-2 & 1 \\ 4 & -2-2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 4 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

Changing to REF eliminates the last row and we get $x_1 = x_2 = 1$.

So the eigenvector associated with λ = 2 is $P_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Note that we do not bother to normalize this.

When λ = -3

$$\begin{bmatrix} 1+3 & 1 \\ 4 & -2+3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 4 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

So $x_2 = 1$, $x_1 = -1/4$ and

$$P_2 = \begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$$

eigenvects(A);

$$[-3, 1, \{[1 \quad -4]\}], [2, 1, \{[1 \quad 1]\}]$$

For method 1, we would at this point say that there are two solutions:

$$Y_1 = e^{2t}\begin{bmatrix}1\\1\end{bmatrix} \qquad ; \qquad Y_2 = e^{-3t}\begin{bmatrix}-\frac{1}{4}\\1\end{bmatrix}$$

and that the general solution is a superposition of these two solutions, with coefficients to be determined from the initial conditions.

In method 2, on the other hand, the next step is to make a matrix from the eigenvectors

$$\mathbf{P} = \begin{bmatrix}1 & -\frac{1}{4}\\1 & 1\end{bmatrix}$$

The diagonal matrix D could be found by using P to apply a similarity transformation to A and multiplying the matrix products out, but since we know the eigenvalues we can take a shortcut and write

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

Now we construct the matrix equation for the new vector U:

$$U' = \mathbf{D}\,U = \begin{bmatrix}2 & 0\\0 & -3\end{bmatrix}\begin{bmatrix}U_1\\U_2\end{bmatrix}$$

This is a set of (two) differential equations

$$U_1' = 2.U_1$$
$$U_2' = -3.U_2$$

These equations are for the U's are not coupled any more. They are each of the form $y' = k\,y$

which has the solution $y = c.e^{k.t}$

Thus
$$U_1 = c_1\,e^{2.t}$$
$$U_2 = c_2\,e^{-3.t}$$

But we want to know $\underline{Y}$. To get $\underline{Y}$ we multiply *PU* together

$$\underline{Y} = \begin{bmatrix} 1 & -\frac{1}{4} \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_1.e^{2.t} \\ c_2.e^{-3.t} \end{bmatrix}$$

and multiplied out this is

$$y_1 = c_1.e^{2.t} - \frac{1}{4}c_2.e^{-3.t} \qquad \text{A)}$$

$$y_2 = c_1.e^{2.t} + c_2.e^{-3.t} \qquad \text{(B)}$$

and these equations are the same as we got from method 1.

Using the initial conditions $y_1(0) = 1$ and $y_2(0) = 6$, at t=0

$$1 = c_1 - \frac{1}{4}c_2 \qquad \text{from (A)}$$

$$6 = c_1 + c_2 \qquad \text{from (B)}$$

Subtracting (A) from (B) gives $c_2 = 4$ and then $c_1 = 2$. So the final solution is

$$y_1 = 2.e^{2.t} - e^{-3.t}$$
$$y_2 = 2.e^{2.t} + 4.e^{-3.t}$$

To do the solution in *Maple*, we first define a system of equations, then solve the system.

```
sys:=diff(y[1](t),t)=y[1](t)+y[2](t), diff(y[2](t),t)=4*y[1](t)-2*y[2](t);
```

$$sys := \frac{\partial}{\partial t} y_1(t) = y_1(t) + y_2(t),$$
$$\frac{\partial}{\partial t} y_2(t) = 4\, y_1(t) - 2\, y_2(t)$$

```
dsolve({sys,y[1](0)=1,y[2](0)=6},{y[1](t),y[2](t)});
```

8.2.3 Summary of the procedure for a homogenous system

1. *Find eigenvalues of A- the λ_i, and put them in the diagonal matrix D.*
2. *Find eigenvectors of A and make the eigenvector matrix P.*
3. *Solve u' = D u to get a set of solutions $u_i = c_i\, e^{\lambda_i t}$*
4. *Find y = Pu*
5. *Use the initial conditions to evaluate the c_i.*

8.3 The Fundamental Matrix

For our further work, it turns out that it is useful to rewrite the solutions in a slightly different way, which nicely separates the integration constants, the time dependent part and eigenvectors each in its own factor in a matrix product. First, we define a new matrix, called the **fundamental matrix** f, in the following way:

$$\boldsymbol{\varphi} = \mathbf{P}\begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & e^{\lambda_n t} \end{bmatrix}$$

The effect of the product is simply to multiply each column vector in *P* by its appropriate exponential time factor. Knowing P, we can usually write down f straight away in that form. Then, the solution of the equation system can be written as

$$\mathbf{Y} = \boldsymbol{\varphi}\mathbf{C} = \boldsymbol{\varphi}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

The expressions $P.\underline{u}$ and $\phi.\underline{C}$ are equivalent ways of writing the solution. You should check this for yourself, by writing out the various matrix products.

We illustrate the use of the fundamental matrix with the following example.

Example 3.

$$\frac{dx}{dt} = 2x + 3y$$
$$\frac{dy}{dt} = 2x + y$$

We use Maple to quickly solve the eigenvalue problem here. First define the coefficient matrix as Q, then find its eigenvalues and vectors:

```
Q:=matrix(2,2,[2,3,2,1]); eigenvects(Q);
```

$$[-1,\ 1,\ \{[-1 \quad 1]\}],\left[4,\ 1,\ \left\{\left[\frac{3}{2} \quad 1\right]\right\}\right]$$

This result needs some interpretation.

There are two answers here, each enclosed in square brackets. The first of the items inside the brackets is the values of the eigenvalues. In the first case this is -1, and in the second it is 4. The second value is unity in both cases indicating that this eigenvalue has multiplicity one, i.e. there is only one eigenvector associated with this eigenvalue. The next value is a set because it is enclosed in braces {..} and this is the set of eigenvectors associated with this eigenvalue. In the first instance this is the vector [-1 1] (for eigenvalue -1) and in the second instance it is the vector [3/2 1] for eigenvalue 4. Thus the fundamental matrix is defined as above to be

```
phi:=matrix(2,2,[3*exp(4*t),-exp(-t), 2*exp(4*t),exp(-t)]);
```

$$\phi := \begin{bmatrix} 3\,e^{(4\,t)} & -e^{(-t)} \\ 2\,e^{(4\,t)} & e^{(-t)} \end{bmatrix}$$

The solution to the DE system is constructed from this by multiplying it with the vector of constants:

```
evalm(phi &* matrix(2,1,[c1, c2]));
```

$$\begin{bmatrix} 3\,c1\,e^{(4t)} - c2\,e^{(-t)} \\ 2\,c1\,e^{(4t)} + c2\,e^{(-t)} \end{bmatrix}$$

Check this by getting the solution directly from Maple using its command **dsolve**:

```
sys:={diff(x(t),t)=2*x(t)+3*y(t), diff(y(t),t)=2*x(t)+y(t)};
dsolve(sys,{x(t),y(t)});
```

$$sys := \left\{ \frac{\partial}{\partial t} x(t) = 2\,x(t) + 3\,y(t), \frac{\partial}{\partial t} y(t) = 2\,x(t) + y(t) \right\}$$

$$\left\{ x(t) = \frac{3}{2}\,\_C2\,e^{(4t)} + \_C1\,e^{(-t)}, \; y(t) = \_C2\,e^{(4t)} - \_C1\,e^{(-t)} \right\}$$

The c's are swapped and the numerical coefficients found here, are slightly different from the ones we had using the fundamental matrix, but that is OK because the constants c_1 and c_2 are arbitrary anyway. What is important, though, is that the ratio of coefficients in front of each individual constant is the same in both forms of the solution, e.g. 3:2 for e^{4t}.

As you see, there is often a lot of calculation hidden behind a single Maple command. While it is tempting to just use Maple as a black box to solve the system of equations, you do need to understand what went on behind the scenes in order to be able to interpret the solution that comes out of it!

8.4 Repeated Eigenvalues

It can sometimes happen that there are repeated (also called degenerate) eigenvalues. That causes a problem, since all of those carry the same exponential factor so that we will then have only one independent solution, such as

$$\underline{Y_1} = \underline{P_1}\, e^{\lambda t}$$

Notice the difference between independent **eigenvectors** and **solutions** here. Depending on the matrix, there may or may not be a second independent eigenvector. But even if there is, it does not give us a second independent solution to the differential equation. For that reason a special procedure is needed to deal with this case. A full explanation of this is given in textbooks, and here we simply state the result, which is that the second independent solution is

$$\underline{Y_2} = \underline{P_1}\, t\, e^{\lambda t} + \underline{R}\, e^{\lambda t}$$

where $\underline{R}$ is calculated from

$$(A - \lambda I)\underline{R} = \underline{P_1}$$

This looks like an eigenvalue equation at first glance, but it is not – notice that the vector on the right is **not** the same as on the left, and is the already known vector that was found first. Note also that in this special case there is a term with the first power of t in the solution, not just the exponential.

For repeated eigenvalues, because we only use one eigenvector, the fundamental matrix cannot be constructed from the P matrix. Instead, we can directly use the solutions $Y_{1 \text{ and }} Y_2$:

$$\phi = \left[\underline{Y_1} \quad \underline{Y_2}\right] = \left[\underline{P_1}\, e^{\lambda t} \quad \underline{P_1}\, t\, e^{\lambda t} + \underline{R}\, e^{\lambda t}\right]$$

Remember that this is not a row vector as it may appear at first glance – since each element of the row vector is in fact itself a column vector, it is actually a matrix!

Example

Solve
$$\frac{dx}{dt} = x - y$$
$$\frac{dy}{dt} = x + 3y$$

First, we solve the eigenvalue problem:

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix}$$
$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & -1 \\ 1 & 3-\lambda \end{vmatrix}$$
$$= 3 - 4\lambda + \lambda^2 + 1$$
$$= (\lambda - 2)^2 = 0$$

There is, in this case, only the single eigenvalue λ=2, and its associated eigenvector is found to be $\underline{P_I}$=[1,-1] . Substituting that in the equation from which we have to solve for $\underline{R}$, we get

$$(A - \lambda I)\underline{R} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

The augmented matrix for this set of equations is easily put in REF:

$$\left(\begin{array}{cc|c} -1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right) \rightarrow \left(\begin{array}{cc|c} 1 & 1 & -1 \\ 0 & 0 & 0 \end{array}\right)$$

and from this it is not hard to find that the solution is $\underline{R}$ = [-1,0]. Using that to construct the second solution from the formula for Y_2 above, the resulting general solution of the system is

$$\begin{bmatrix} x \\ y \end{bmatrix} = c_1 \begin{bmatrix} e^{2t} \\ -e^{2t} \end{bmatrix} + c_2 \begin{bmatrix} te^{2t} - e^{2t} \\ -te^{2t} \end{bmatrix}$$

For practice, you should work through this example and confirm that you get the same solution.

8.5 Non-homogenous systems

Our plan of attack on non-homogenous systems is a combination of the strategy we used for algebraic equations, and the one that worked for a single differential equation.

The first step is to solve the associated system of homogenous equations, as we have just learned to do. That gives us the **complementary solution** as a vector Y_c (i.e., actually a set of solutions).

What remains, is to find the **particular solution**. And the way to do that, is to apply **variation of parameters** to the matrix form of the homogenous solution – which is why it was important to separate the integration constants (that are now going to become variables) in a separate factor by the introduction of the fundamental matrix.

We saw previously that variation of parameters works well because it leads to a cancellation of terms from the DE that leaves us with something simpler to solve. The same thing happens here, and we can do that once and for all by working through a little matrix algebra for the general case. Start with the complementary solution written in terms of the fundamental matrix:

$$\underline{y}_c = \phi \ \underline{c} \tag{1}$$

Varying the parameters means that we replace the constant vector $\underline{c}$ by a vector of unknown functions of time, $\underline{\boldsymbol{U}}$:

$$\begin{aligned}\underline{y}_p &= \phi\,\underline{U} \\ &= \begin{bmatrix}\phi_{11} & \phi_{12}\\ \phi_{21} & \phi_{22}\end{bmatrix}\begin{bmatrix}U_1(t)\\ U_2(t)\end{bmatrix}\end{aligned} \tag{2}$$

If we differentiate each of this pair of equations using the product rule, and rearrange the terms, this can be written as a matrix equation:

$$\underline{y}'_p = \phi'\underline{U} + \phi\underline{U}'$$

Notice that this has exactly the same form as if we just applied the product rule of differentiation directly to the matrix product.

The non-homogenous DE that the particular solution satisfies is written in matrix form as

$$\begin{aligned}\underline{y}'_p &= A\,\underline{y}_p + \underline{f} \\ &= A\phi\underline{U} + \underline{f}\end{aligned}$$

where (2) was used in the second step. Now inserting the derivative from above

$$\phi'\underline{U} + \phi\underline{U}' = A\phi\underline{U} + \underline{f} \tag{3}$$

The cancellation of terms that happens here, will remove the first term involving a derivative of the fundamental matrix. To see how that happens, we need to go back to the homogenous differential equation that the complementary solution satisfies:

$$\underline{y}'_c = A\,\underline{y}_c = \phi'\,\underline{c}$$

The second equals sign above follows by differentiating equation (1) and bearing in mind that $\underline{c}$ is constant. If we now substitute equation (1) also in the middle term of the last equation, we get

$$A\phi\,\underline{c} = \phi'\,\underline{c}$$

Now $\underline{c}$ is an arbitrary vector – it will have different values depending on the initial conditions. So the only way that this equation can hold for all possible different boundary conditions, is if

$$A\phi = \phi'$$

We see from this that the first term on each side of equation (3) above are the same, and so we are left with

$$\underline{U}' = \phi^{-1}\underline{f}$$

Because all the functions on the RHS are known functions of t, this is a very simple set of differential equations, which is easy to solve provided we can find the set of integrals:

$$\underline{U} = \int (\phi^{-1}\underline{f})\,dt$$

Once done, it is a simple matter to construct the particular solution because we just have to carry out the matrix product as required by equation (2)

$$\underline{y}_p = \phi\,\underline{U}$$

We saw previously that the particular solution should not contain any integration constants, because the complementary solution already contains all of them. To make that happen, we need to take the integration limits in the integral as t_0 and t respectively, where t_0 is the time at which the initial conditions are given (very often, $t_0 = 0$). By doing so, U and therefore also y_p will become zero at the initial time, and then at that time only y_c comes into play to satisfy the initial conditions.

The most general solution to a non-homogenous system of equations is found by superimposing the particular solution and the complementary solution, as before. To show that this still works for the matrix solution, we merely need to substitute it into the equation.

We wish to show that $\underline{y} = \underline{y}_c + \underline{y}_p$ is a solution to the non-homogeneous system.

Differentiate it: $\underline{y}' = \underline{y}'_c + \underline{y}'_p = A\underline{y}_c + (A\underline{y}_p + \underline{f}) = A(\underline{y}_c + \underline{y}_p) + \underline{f}$

where we used the knowledge that $\underline{y}_p$ satisfies the non-homogeneous equation in the second step

I.e., we have shown that $\underline{y}' = A\underline{y} + \underline{f}$

which is what we set out to prove.

We will now demonstrate the procedure step by step using an example problem.

8.5.1 Example

Solve the initial value problem

$$\frac{dy_1}{dt} = 3.y_1 - 4.y_2 + 1$$
$$\frac{dy_2}{dt} = 2.y_1 - 3.y_2 + t$$

subject to the initial conditions

$$y_1(0) = +1$$
$$y_2(0) = -1$$

This becomes

$$\underline{y}' = \begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix} \underline{y} + \underline{f} \quad ; \quad \underline{f} = \begin{bmatrix} 1 \\ t \end{bmatrix}$$

Steps

Ignore *f* and solve the homogeneous system

Find the eigenvalues of A:

$$\begin{vmatrix} 3-\lambda & -4 \\ 2 & -3-\lambda \end{vmatrix} = 0$$

These occur when

$$(3-\lambda\ )(-3-\lambda\ ) + 8 = 0$$

$$\lambda^{2} - 1 = 0$$
$$\lambda = \pm 1$$

Find the associated eigenvectors

- when $\lambda = -1$, the eigenvector is $\begin{bmatrix}1\\1\end{bmatrix}$
- when $\lambda = +1$, the eigenvector is $\begin{bmatrix}2\\1\end{bmatrix}$

A:=matrix(2,2,[3,-4,2,-3]);

$$A := \begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix}$$

eigenvects(A);

$$[1, 1, \{[2 \quad 1]\}], [-1, 1, \{[1 \quad 1]\}]$$

Form the matrix *P*

$$P = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$$

We can write down D directly by using the eigenvalues. However, to check if the numerical calculations were done correctly, we can also calculate it more laboriously from P and A:

$$D = P^{-1}.A.P = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Form the fundamental matrix $$\phi = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} e^{-t} & 0 \\ 0 & e^{t} \end{bmatrix} = \begin{bmatrix} e^{-t} & 2.e^{t} \\ e^{-t} & e^{t} \end{bmatrix}$$

Notice that the eigenvalues (that appear in the exponents) have to be put in the diagonal positions that correspond to the order in which the column vectors were taken. We could take the columns in another order if we like, but then the values on the diagonal of the second factor also have to be changed accordingly!

Calculate the complementary solution.

$$y_c = \phi C = \begin{bmatrix} e^{-t} & 2.e^{t} \\ e^{-t} & e^{t} \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 e^{-t} + 2c_2 e^{t} \\ c_1 e^{-t} + c_2 e^{t} \end{bmatrix}$$

Set up the DE for U $\underline{U}' = \phi^{-1}.\underline{f}$

$$\det(\phi) = 1 - 2 = -1 \neq 0$$

Because the determinant is not zero, the matrix can be inverted. Using the rule we encountered in the matrix part of the course to write down the inverse of a 2 x 2 matrix:

$$\phi^{-1} = -1.\begin{pmatrix} e^t & -2.e^t \\ -e^{-t} & e^{-t} \end{pmatrix}$$

$$\phi^{-1}.\underline{f} = -1.\begin{bmatrix} e^t & -2.e^t \\ -e^{-t} & e^{-t} \end{bmatrix}\begin{bmatrix} 1 \\ t \end{bmatrix}$$

$$\underline{U}' = \begin{bmatrix} 2.t.e^t - e^t \\ e^{-t} - t.e^{-t} \end{bmatrix}$$

Calculate $\underline{U}$ by integrating.

This has to be carried out element by element and we must evaluate this integral over the range $\int_{t_o}^{t}$ so that we get the correct constants of integration.

Integrate the first element by parts with

$$f = (2.t - 1) \quad f\,' = 2$$

$$g\,' = e^t \qquad g = e^t$$

Thus

$$\int u_1' \, dt = \int f\, g' = f\, g - \int f' g$$
$$= (2t - 1)e^t - \int 2e^t$$
$$= (2t - 3)e^t$$

Now

$$\int_0^t (2s - 1)e^s \, ds = (2s - 3)e^s \Big]_{s=0}^{s=t}$$
$$= (2t - 3)e^t - (2(0) - 3)1$$
$$= (2t - 3)e^t + 3$$

For the second vector component of U we follow the same procedure to integrate it:

$$u_2' = e^{-t}(1-t)$$

Again, integrate by parts with

$$f = (1-t) \qquad f\,' = -1$$

$$g\,' = e^{-t} \qquad g = -e^{-t}$$

Thus

$$u_2 = \int_0^t e^{-s}(1-s)\,ds$$

$$= (1-s).-e^{-s} - \int e^{-s}.ds$$

$$= (1-s).-e^{-s} + e^{-s}$$

$$= -e^{-s} + s.e^{-s} + e^{-s}$$

$$= s.e^{-s}$$

Thus

$$u_2 = s\,e^{-s}\Big]_{s=0}^{s=t} = t\,e^{-t}$$

So

$$\underline{U} = \begin{bmatrix} (2.t-3).e^{t}+3 \\ t.e^{-t} \end{bmatrix}$$

Now find the particular solution:

$$\begin{aligned} y_p &= \phi\,\underline{U} \\ &= \begin{bmatrix} e^{-t} & 2e^{t} \\ e^{-t} & e^{t} \end{bmatrix} \begin{bmatrix} (2t-3)e^{t}+3 \\ te^{-t} \end{bmatrix} \\ &= \begin{bmatrix} (2t-3)+3e^{-t}+2t \\ (2t-3)+3e^{-t}+t \end{bmatrix} \\ &= \begin{bmatrix} 3e^{-t}+4t-3 \\ 3e^{-t}+3t-3 \end{bmatrix} \end{aligned}$$

Find the integration constants from the initial conditions.

The general solution will be obtained by superimposing the particular and complementary solutions as follows

$$\begin{aligned} y &= y_p + y_c \\ &= \begin{bmatrix} e^{-t} & 2e^{t} \\ e^{-t} & e^{t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + y_p \end{aligned}$$

To get the integration constants c_1 and c_2 we can multiply this equation through by f^{-1} and get

$$\underline{C} = \phi^{-1}(t_0)\,\underline{y}(t_0)$$

because we have constructed y_p such that it falls away at $t = t_0 = 0$ in this case.

Remembering that $\underline{y}(0)$ is the vector of initial conditions that we were given

$$\begin{aligned} \underline{C} &= \phi^{-1}(0)\,\underline{y}(0) \\ &= \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix} \end{aligned}$$

Substitute the values for *C* into the general solution

Putting these constant values back into the two solutions individually, we have

$$y_1 = \overset{c_1}{-3.e^{-t}} + \overset{c_2}{4.e^{t}} + \underbrace{3.e^{-t} + 4.t - 3}_{y_{p\,1}} \quad = 4e^{t} + 4t - 3$$

$$y_2 = \overset{c_1}{-3.e^{-t}} + \overset{c_2}{2.e^{t}} + \underbrace{3.e^{-t} + 3.t - 3}_{y_{p\,2}} \quad = 2e^{t} + 3t - 3$$

So finally, the fully determined solution of the initial value problem is

$$y = \begin{bmatrix} 4.e^{t} + 4.t - 3 \\ 2.e^{t} + 3.t - 3 \end{bmatrix}$$

sys:=diff(y[1](t),t)=3*y[1](t)-4*y[2](t)+1, diff(y[2](t),t)=2*y[1](t)-3*y[2](t)+t;

$$sys := \frac{\partial}{\partial t} y_1(t) = 3\, y_1(t) - 4\, y_2(t) + 1,$$
$$\frac{\partial}{\partial t} y_2(t) = 2\, y_1(t) - 3\, y_2(t) + t$$

dsolve({sys,y[1](0)=1,y[2](0)=-1}, {y[1](t),y[2](t)});

$$\left\{ y_1(t) = -3 + 4\,t + 4\,\mathrm{e}^{t},\ y_2(t) = 3\,t - 3 + 2\,\mathrm{e}^{t} \right\}$$

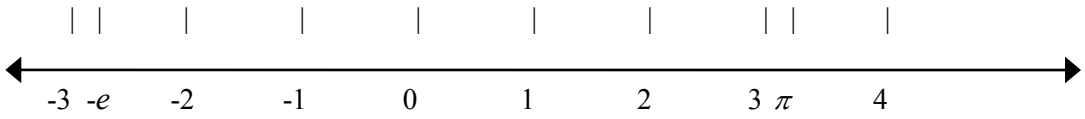
8.5.2 Summary of the procedure for a non-homogenous system

1. *Solve the homogeneous equation to get eigenvalues and the eigenvector matrix P.*
2. *Form the fundamental matrix ϕ from the eigenvctors.*
3. *Calculate the complementary solution $y_c = \phi \underline{c}$ (or equivalently = $P \underline{u}$)*
4. *Calculate* $\underline{U}' = \phi^{-1} f$
5. *Integrate this to calculate* $\underline{U} = \int_{t_0}^{t} \phi^{-1} f$
6. *Calculate the particular solution $\underline{y_p} = \phi \underline{U}$*
7. *Solve for $\underline{c}$ from the initial conditions :*
   $$\underline{c} = \phi^{-1}(t_0)\, \underline{y}(t_0)$$
8. *Then the final solution is $\underline{y} = \underline{y_c} + \underline{y_p}$.*

9 Appendix: Complex Numbers

9.1 Representing complex numbers

If x is any real number (represented here by a point on the real line)



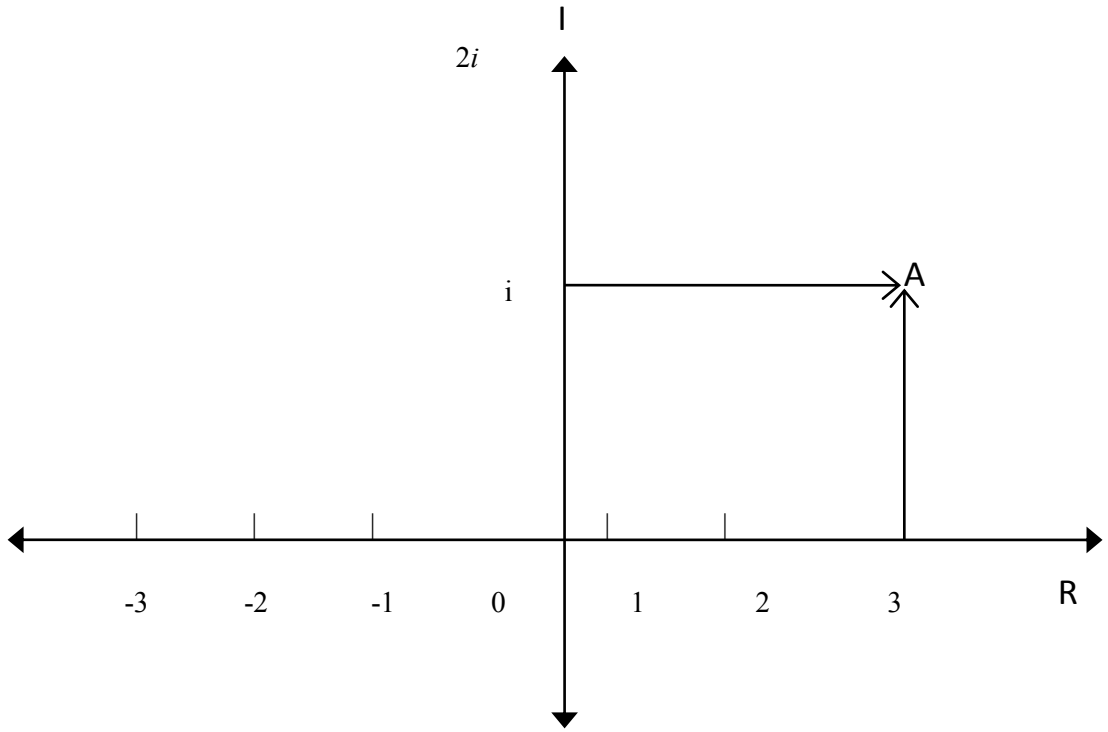
then $x^2 \geq 0$. But what about $x^2 = -1$?

We use the special number i, often called the "imaginary number", where $i^2 = -1$. We can see how Maple denotes this by doing

```
sqrt(-1);
```

I

Complex numbers arise when we have a + *i* b. To provide "reality" map these numbers into points on a plane, called the "complex plane" where the vertical axis represents the imaginary part:



A in the figure represents the complex number 3 + *i*.

Consider $x^2 + 2x + 2 = 0$; using the quadratic formula this is solved by

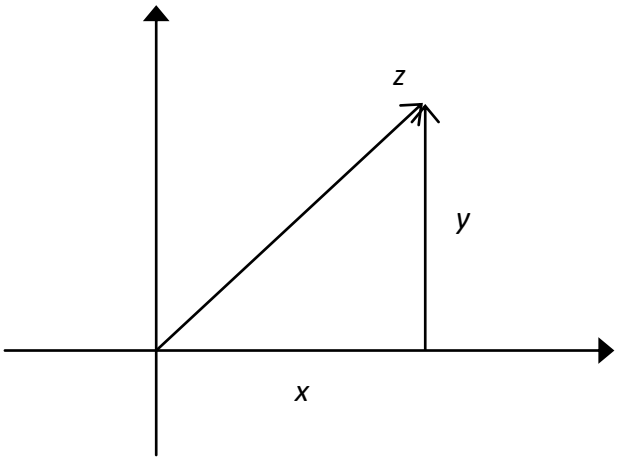
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$= \frac{-2 \pm \sqrt{4 - 4.1.2}}{2}$$
$$= \frac{-2 \pm \sqrt{-4}}{2} \quad = -1 \pm i$$

solve(x^2+2*x+2=0,x);

-1 + *I*, -1 − *I*

We often write complex numbers $z = x + i\,y$ where the point z is represented by the co-ordinates (x, y) in the complex plane.

Complex numbers have a real part (x) and an imaginary part (y), i.e. Re(z) = x and Im(z) = y.



The distance of z from the origin is called the **modulus,** written

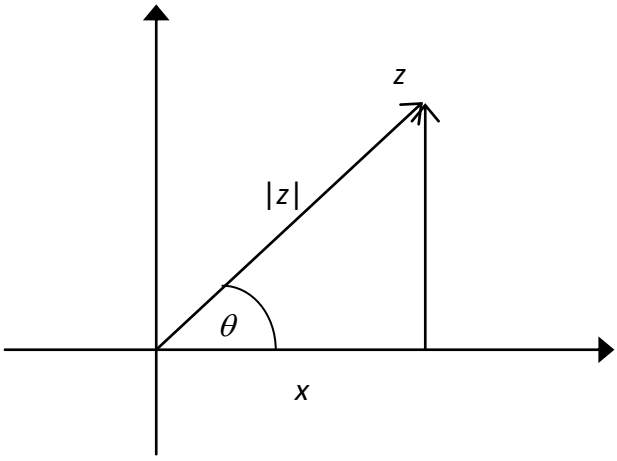
$$|z| = \sqrt{x^2 + y^2}$$

Example

How far is -12 + 5i from the origin?

$$|z| = \sqrt{x^2 + y^2} = \sqrt{(-12)^2 + 5^2} = \sqrt{169} = 13$$

If we continue this we can think of z as a vector in the complex plane, making an angle θ with the real axis.



We know $x = |z| \cos \theta$

$y = |z| \sin \theta$

Write $|z| = r$ then

$$z = r.\cos\theta + i.r.\sin\theta$$

$$= r(\cos\theta + i\sin\theta)$$

The angle can be calculated from x and y by the formula

$$\tan\theta = \frac{y}{x} \quad ; \quad \theta = \arctan\frac{y}{x}$$

When using this formula one would usually select a value of θ in the interval $0 < \theta < 2\pi$, avoiding negative θ values.

9.2 Algebraic operations

9.2.1 Addition

Let $z_1 = x_1 + i.y_1$ and $z_2 = x_2 + i.y_2$

$z_1 + z_2 = (x_1 + x_2) + i.(y_1 + y_2)$

$z_1 - z_2 = (x_1 - x_2) + i.(y_1 - y_2)$

9.2.2 Multiplication

$$(3 + 2\,i)\,(4 - 5\,i) = 3.4 - 3.5\,i + 2\,i.4 - 10\,i^2$$
$$= 12 - 15\,i + 8\,i - 10\,i^2$$
$$= 12 - 7\,i + 10$$
$$= 22 - 7\,i$$

(3+2*I)*(4-5*I);

$$22 - 7\,I$$

9.2.3 Complex conjugates and division

Remember $x^2 + 2x + 2 = 0$ gave

$$x_1 = -1 + i \qquad x_2 = -1 - i$$

x_1 and x_2 are called **complex conjugates** because their real parts are the same and their imaginary parts just differ in sign.

So $z = x + i\,y$ has conjugate $\bar{z} = x - i\,y$

Consider $z\,.\,\bar{z} = (x + iy)\,(x - iy)$

$$= x^2 - iyx + iyx - i^2y^2 = x^2 + y^2$$

So $z\,.\,\bar{z} = |z|^2$

To get $\frac{z_1}{z_2}$, multiply the top and bottom by $\bar{z}_2$

$$\frac{z_1}{z_2} = \frac{z_1}{z_2}\frac{\bar{z}_2}{\bar{z}_2} = \frac{1}{|z_2|^2} z_1 z_2$$

That is how we do division.

Example 1

Calculate $\dfrac{1+2i}{1+i}$

$z_1 = 1+ 2i$ $z_2 = 1 + i$

So $\bar{z}_2 = 1 - i$, thus $z_2 . \bar{z}_2 = (1+ i) (1 - i) = 1+ i - i - i^2 = 2$

$$\frac{(1+2i)}{(1+i)} = \frac{(1+2i)}{(1+i)}\frac{1-i}{1-i} = \frac{1}{2}\left[(1+2i)(1-i)\right]$$
$$= \frac{1}{2}\left[1+2i-i-2i^2\right] = \frac{1}{2}\left[3+i\right]$$

(1+2*I)/(1+I);

$$\frac{3}{2}+\frac{1}{2}I$$

Example 2

What is the reciprocal of i? $\dfrac{1+0.i}{0+i.1} = \dfrac{1}{i} = \dfrac{1(-i)}{i(-i)} = \dfrac{-i}{-i^2} = -i$

9.3 Euler's formula

Next, we investigate the meaning of exponential numbers that contain an i in the exponential. We will see that trigonometric functions give us a way to interpret this.

Start with the complex number $q = (\cos\theta + \mathrm{i}\sin\theta)$. Considering this as a function of θ we can differentiate it:

$$\frac{dq}{d\theta} = -\sin\theta + i\cos\theta = i(i\sin\theta + \cos\theta) = iq$$

Now recall the differentiation formula $\dfrac{d}{dx}e^{kx} = k\,e^{kx}$. We see that differentiating the function $f(x) = e^{kx}$, we get the same function f(x) but multiplied by the constant factor k from the exponent. In fact, the special number e is defined by this equation; it is the only number for which no other constant appears when its derivative is taken. So to satisfy the DE above for q, we have to take $q = e^{i\theta}$, i.e.

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

This famous equation is known as Euler's formula. It can also be proven in a more formal way by using series expansions (Taylor series) for the functions sin x, cos x and exp(x).

Applying Euler's formula the expression used above for any complex number z can now be rewritten

$$z = x + i\,y = r\,e^{i\theta}$$

Geometrically, we see that Euler's formula is equivalent to describing the vector in the complex plane that represents z bye its polar coordinates rather than its Cartesian coordinates.

Some special cases of interest are that, with r = 1, we get

$$i = e^{\frac{1}{2}i\pi} \qquad -1 = e^{i\pi} \qquad -i = e^{-\frac{1}{2}i\pi}$$

As the example with ±i demonstrates, to take complex conjugates we can change the same of the i in the exponent just as well as when we change the sign of i in the real and imaginary part representation. The polar representation of complex numbers also simplifies multiplication and division. If we have $z_1 = r_1\ exp(i\theta_1)$ and $z_2 = r_2\ exp(i\theta_2)$ we can calculate

$$z_1 z_2 = r_1 r_2\, e^{i(\theta_1+\theta_2)} \qquad \frac{z_1}{z_2} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}$$

Check for yourself that applying these formulas to the examples worked out above, gives the same results.

Finally, we can also invert Euler's formula to express the trigonometric functions in terms of complex exponentials:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \quad ; \quad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

9.4 Log, Exponential and Hyperbolic functions

From the polar representation it is straightforward to take logarithms of complex numbers:

$$z = x + iy = r\,e^{i\theta}$$
$$\ln z = \ln r + i\,\theta = \ln|z| + i\arctan(y/x)$$

From this formula, we can now take logarithms of imaginary numbers and even negative numbers; for example

$$\ln i = \frac{i\,\pi}{2} \quad ; \quad \ln(-1) = i\,\pi$$

Check this with Maple:

```
convert(I,polar);
```

$$\mathrm{polar}\left(1, \frac{1}{2}\pi\right)$$

Getting the log of a negative number: take z = -5

```
convert(-5,polar);
```

$$\mathrm{polar}(5, \pi)$$

and from this we conclude that ln (-5) = ln 5 + $i\,\pi$

Also, exponential functions with complex arguments become possible, for example

$$\exp(x + i\,y) = \exp(x)\exp(i\,y) = e^{x}\left(\cos y + i\sin y\right)$$

Raising complex numbers to powers also gives some interesting results when combined with Euler's formula. For example,

$$e^{in\theta} = \cos(n\theta) + i\sin(n\theta)$$
$$e^{in\theta} = \left(e^{i\theta}\right)^n = \left(\cos\theta + i\sin\theta\right)^n$$
$$\therefore \cos(n\theta) + i\sin(n\theta) = \left(\cos\theta + i\sin\theta\right)^n$$

The first of these equations is usually known as de Moivre's theorem. From the last line, we can get formulas for trigonometric functions of 2θ, 3θ etc. simply by multiplying out the appropriate power, and extracting its real and imaginary parts – much easier than deriving these geometrically!

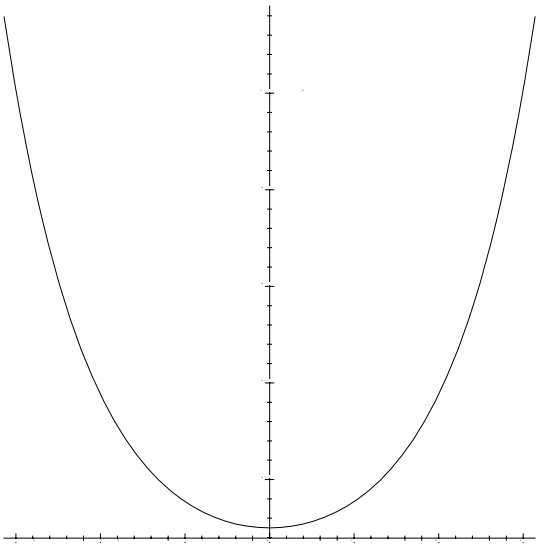
Finally, the sine and cosine functions for imaginary arguments are important enough that they are defined as special, so-called "hyperbolic" functions given by the formulas

$$\sinh(x) = \frac{e^x - e^{-x}}{2}$$

$$\cosh(x) = \frac{e^x + e^{-x}}{2}$$

A plot of cosh(θ):

```
plot(cosh(x),x=-Pi..Pi);
```



One practical use of this function is that a suspended cable, such as for a suspension bridge, takes the shape of the equation $y = \cosh(x)$ and is sometimes called a **catenary**.

Prove for yourself the relationships $\cos(iy) = \cosh(y)$; $\sin(iy) = i\sinh(y)$

Just as we have inverse trigonometric functions arcsin x and arcos x, we can define inverse functions usually written as sinh-1 and cosh-1. These are related to logarithms; for example, you can prove for yourself that, for real x and a,

$$\sinh^{-1}\frac{x}{a} = \ln\left(x \pm \sqrt{x^2 + a^2}\right) - \ln a$$

For this reason, the answer obtained for certain integrals from integration tables or even software like Maple and Mathematica, can sometimes be in terms of either logarithms or inverse hyperbolic functions. Such answers are equivalent except for an additive constant.

9.5 Differentiation Formulae

We have already seen that $\frac{d}{d\theta}\left(e^{i\theta}\right) = i.e^{i\theta}$. Generally, when differentiating

we can treat *i* like any other constant. Test this idea by differentiating the formula that was just derived for the sine function:

$$\frac{d}{d\theta}\left(\sin(\theta)\right) = \frac{d}{d\theta}\left(\frac{e^{i\theta} - e^{-i\theta}}{2i}\right)$$

$$= \frac{1}{2i}\left[i\,e^{i}\theta - (-i\,e^{-i}\theta)\right]$$

$$= \frac{1}{2i}\left[i\,e^{i}\theta + i\,e^{-i}\theta\right]$$

$$= \frac{e^{i\theta} + e^{-i\theta}}{2}$$

$$= \cos\,\theta$$

Now see what happens for the hyperbolic sine:

$$\frac{d}{dx}\left(\sinh(x)\right) = \frac{d}{dx}\left(\frac{e^{x} - e^{-x}}{2}\right)$$

$$= \frac{1}{2}\left[e^{x} - (-e^{-x})\right]$$

$$= \frac{1}{2}\left[e^{x} + e^{-x}\right] = \cosh\,(x)$$

Note: This is just like for the ordinary sine. The same happens with cosh – try it!

```
diff(cosh(x),x);
```

$$\sinh(x)$$