

MATH1052

**Multivariate Calculus
and
Ordinary Differential Equations**

WORKBOOK
Semester 2, 2010

How to use this workbook

This workbook is designed for use during lectures. It contains copies of all of the slides that will be shown in lectures, giving you time to listen and think in class, rather than spending the entire time writing.

There are exercises, definitions and examples in the workbook for you to fill in. These will be covered in lectures, and you should write down all the information given. The lecturer will make it clear when they are covering something that you need to complete. Any highlighted formulae or results such as

$$1 + 1 = 2$$

should be common knowledge to all humans on the planet, including you.

The completed workbook will act as a study guide to assist you in working through assignments and preparing for the mid-semester and final exams. For this reason, it is very important to attend lectures.

If after reading (and perhaps even studying) parts of the material in these notes your mathematical life remains troubled, do not despair. The *First Year Learning Centre* is open every day from 2–4pm in the Priestley building (building 67).

For any further information about the course, including copies of assignments and worksheets, please go to *Blackboard* at <http://blackboard.elearning.uq.edu.au>

These notes have been carefully prepared, but there will inevitably be some errors. We are continually trying to improve the notes, and any suggestions and or comments are very welcome.

Overview

MATH1052 is a second course in calculus. It will enable you to use and visualise functions of two or more variables. Most importantly it extends your knowledge of calculus so you can differentiate multivariable functions and find their maxima and minima. These ideas are basic to the complex optimisation problems that occur when mathematics is used in economics, finance, and engineering.

MATH1052 also introduces ordinary differential equations (ODEs); one of the basic tools in mathematical modelling. ODEs are used to describe how things change with time. In engineering and the physical sciences ODEs are used to work out the motion of particles and satellites, the rates of chemical reactions, or to model electric fields. In biology they are used to model populations (to determine safe harvesting levels, for example) and to model epidemics.

The textbook for the course is *Calculus (6th Edition)* by James Stewart (Brooks/Cole). This is the only text you need for the course.

Contents

1	Functions of Several Variables	11
1.1	Introduction	11
1.1.1	Visualising functions	12
1.1.2	Main points	13
1.2	Conic Sections	14
1.2.1	Parabolas	14
1.2.2	Circles	17
1.2.3	Ellipses	18
1.2.4	Hyperbolas	20
1.2.5	Main points	21
1.3	Scalar Equation for a Plane	22
1.3.1	Horizontal planes	22
1.3.2	Vertical planes	23
1.3.3	Arbitrary planes	23
1.3.4	Main points	27
1.4	Vector Equation of a Plane I	28
1.4.1	Revision of vectors	28
1.4.2	Vector equation of a plane I	32
1.4.3	Main points	34
1.5	Equations for a Line	35
1.5.1	Parallel, skew and orthogonal Lines.	41
1.5.2	Vector equation of a plane II	44
1.5.3	Distance from a Point to a Plane	46
1.5.4	Main points	48
1.6	Contour Diagrams	49
1.6.1	Main points	58
1.7	Cross-sections of a Surface	59
1.7.1	Main points	64

2	Partial Derivatives and Tangent Planes	65
2.1	Limits and Continuity	65
2.1.1	Review of the 1-variable case	65
2.1.2	Multivariable limits	67
2.1.3	Multivariable continuity	71
2.1.4	Main points	71
2.2	Partial Derivatives	72
2.2.1	Slope in the x -direction	72
2.2.2	Slope in the y -direction	73
2.2.3	Partial derivatives for $f(x, y, z)$	75
2.2.4	Higher order derivatives	76
2.2.5	Main points	76
2.3	The Tangent Plane	77
2.3.1	Review for $f(x)$	77
2.3.2	Equation for a tangent plane	78
2.3.3	Smoothness	81
2.3.4	Main points	81
2.4	Linear Approximations	82
2.4.1	Review for $f(x)$	82
2.4.2	Linear approximations for $f(x, y)$	83
2.4.3	Estimating small changes	85
2.4.4	Estimating error	86
2.4.5	Extra reading: differentials	87
2.4.6	Main points	87
2.5	Gradients and Directional Derivatives	88
2.5.1	Directional derivatives	88
2.5.2	The gradient vector ∇f	90
2.5.3	Properties of the gradient vector	92

2.5.4	Main points	97
2.6	The Chain Rule and Applications	98
2.6.1	Review for $f(x)$	98
2.6.2	The chain rule for $f(x, y)$	99
2.6.3	Larger Chain Rules	100
2.6.4	Implicit Differentiation	102
2.6.5	Main points	104
3	Max and Min Problems on Surfaces	105
3.1	Quadratic Approximation	105
3.1.1	Review for $f(x)$	105
3.1.2	Quadratic approximations of $f(x, y)$	108
3.1.3	Main points	113
3.2	Critical Points and Optimization	114
3.2.1	Review for $f(x)$	114
3.2.2	Critical points for $f(x, y)$	116
3.2.3	Global maxima and minima	127
3.2.4	The Matlab command fminsearch	129
3.2.5	Main points	129
3.3	Constrained Optimisation and Lagrange Multipliers	130
3.3.1	Lagrange multipliers	130
3.3.2	Main points	134
4	Ordinary Differential Equations	135
4.1	Introduction	135
4.1.1	Main points	135
4.1.2	Examples of ODEs	135
4.1.3	How do ODEs arise?	136
4.1.4	Solution to an ODE	137

4.1.5	Initial value problem: fluid flow	138
4.1.6	The order of an ODE	138
4.1.7	Dr. VO's bike at constant speed	139
4.2	Motion of Projectiles	140
4.2.1	Main points	140
4.2.2	Gravity in one dimension	140
4.2.3	Apples in two dimensions	141
4.2.4	Example: hitting a target	142
4.2.5	Realistic models	144
4.2.6	A model with air resistance	144
4.3	Slope Fields and Equilibrium Solutions	146
4.3.1	Main points	146
4.3.2	Example: the slope field of $y' = 2y$	147
4.3.3	Equilibrium solutions	147
4.3.4	Stability of equilibrium solutions	151
4.3.5	Existence and uniqueness of solutions	152
4.3.6	Long-term behaviour	153
4.4	Euler's Method for Solving ODEs Numerically	154
4.4.1	Main points	154
4.4.2	Revision: equation of a straight line	154
4.4.3	Euler's method using Matlab; an example	157
4.4.4	Demonstrating errors in Euler's method	158
4.5	Separable ODEs	159
4.5.1	Definition	159
4.5.2	Solving separable ODEs	160
4.5.3	Singular solutions	164
4.6	Applications, Heat Equation, Population Growth	167
4.6.1	Newton's law of cooling	167

4.6.2	Malthus: “...doomed to misery and vice”	170
4.6.3	Verhulst to the rescue; down with the Malthusian law	172
4.6.4	Extra reading: numerically solving ODEs	176
4.7	Solving Linear First Order Equations	177
4.7.1	Definition	177
4.7.2	Partially worked-out example: RL circuit	182
4.8	Linear Second-Order Differential Equations with Constant Coefficients	184
4.8.1	Definition	184
4.8.2	The superposition principle	185
4.8.3	Reduction of order	186
4.8.4	Homogeneous linear second-order ODEs with constant coefficients	189
4.8.5	Extra reading: numerical solutions	196
4.9	Inhomogeneous Linear Second-Order ODEs	197
4.9.1	Inhomogeneous linear second-order ODEs with constant coefficients	197
4.10	Simple Harmonic Motion and Damped Oscillations	207
4.10.1	Main points	207
4.10.2	The undamped spring	207
4.10.3	Example	210
4.10.4	The damped spring	211
4.10.5	Underdamping	212
4.10.6	Critical damping	215
4.10.7	Overdamping	216
4.10.8	Summary of motion of a spring:	218
4.10.9	The pendulum	219

5	Parametrisation of Curves and Line Integrals	221
5.1	Parametrisation of Curves	221
5.1.1	Curves in \mathbb{R}^3	228
5.1.2	Piecewise linear paths	229
5.1.3	Extra reading: Matlab plotting parametrised equations	229
5.2	Position Vectors, Velocity and Acceleration	230
5.2.1	Position vector	230
5.2.2	Parametrising straight lines	230
5.2.3	Velocity vector	233
5.2.4	Acceleration	233
5.2.5	Who wants to be a millionaire?	235
5.2.6	Extra reading: stone on a string	236
5.3	Line Integrals and Work Done	237
5.3.1	Length of a curve	237
5.3.2	Work done by a constant force	241
5.3.3	Work done over a curve	242
5.3.4	Extra reading: length of a spiral	246
5.4	Gradient Fields	247
5.4.1	Main points	247
5.4.2	Conservative fields	247
5.5	Work Done or Path Integrals for Non-conservative Fields	252
6	MATLAB Guide	255
6.1	Introduction	255
6.1.1	Getting started	255
6.1.2	Getting help	255
6.2	The Basics	256
6.3	M-files	257

6.3.1	If statements	258
6.3.2	For and while loops	259
6.4	Visualising Functions	261
6.4.1	Ezplot	261
6.4.2	Ezcontour	262
6.4.3	Ezsurf and ezmesh	264
6.4.4	Functions of three variables	264
6.5	Visualising Data	264
6.5.1	Plot and plot3	265
6.5.2	Contour, surf and mesh	265
6.6	Animations	266
6.7	Optimisation	268
6.7.1	Using fminsearch	268
6.7.2	Lines of best fit	269
6.8	Differential Equations	271
6.8.1	Slope fields	271
6.8.2	Euler's method	273
6.8.3	Runge–Kutta method	273
6.8.4	Systems of differential equations	275
6.9	Parametric Functions	277
6.10	Common commands and operations	278
6.10.1	Common numerical operations and constants	278
6.10.2	Vector and matrix operations	278
6.10.3	Special matrices	279

1 Functions of Several Variables

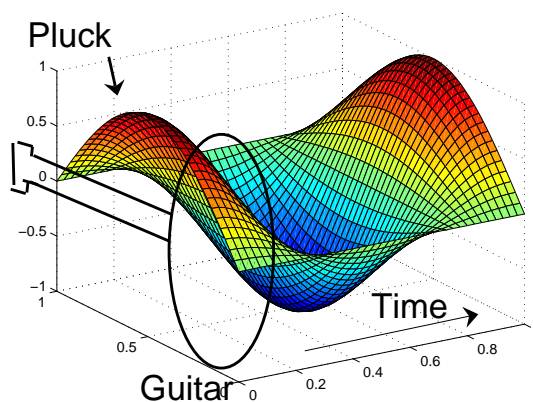
1.1 Introduction

From MATH1051 or earlier mathematics courses, you should be familiar with functions f of one variable. The tools of calculus were useful because you could:

- Sketch the graph $y = f(x)$ of f .
- Find the minima and maxima of f .
- Analyse the slope of f by calculating f' .
- Find approximations of f using Taylor series.
- Find solutions to $f(x) = 0$.

In this first module of MATH1052, we shall undertake these same basic ideas, but in the framework of functions of several variables (multivariate functions). Most of the theory for this section is covered in Stewart, Chapter 15.

Many familiar formulae are essentially just functions of more than one variable. For example, the volume V of a box is a function of its width, height and depth: $V(w, h, d) = whd$, and the profile of a guitar string is a function of time and position along the string: $f(x, t) = A \sin x \cos t$.



1.1.1 Visualising functions

Consider the volume of a cylinder as a function of two variables:

$$V(r, \ell) = \pi r^2 \ell.$$

A first method of visualising this function is to plot V for fixed ℓ and varying r .



Using Matlab, the `ezplot` function is the simplest graphing aid:

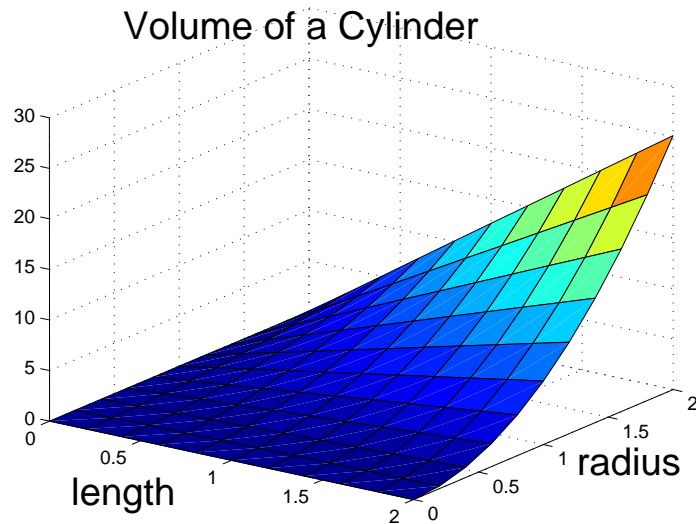
```
ezplot('pi*r^2*1', [0,5])  
hold on; ezplot('pi*r^2*2', [0,5])  
hold on; ezplot('pi*r^2*0.5', [0,5])
```

Alternatively, we could keep r fixed and plot V as a function of ℓ :

```
ezplot('pi*1^2*1', [0,5])  
hold on;  
ezplot('pi*0.5^2*1', [0,5])  
ezplot('pi*2^2*1', [0,5])
```

Better still is visualising the surface by sketching the graph of $z = f(x, y)$. In Matlab, the `ezsurf` function will prove to be our quickest tool:

```
ezsurf('pi*r^2*l', [0,2,0,2])
```



Note how you can change the point of view of the plot with the 'rotate' button.

In Matlab we can plot contour lines for the volume function using the `ezcontour` function:

```
ezcontour('pi*r^2*l', [0,2,0,2])
```

Using these tools, you should be able to replicate the plots in Stewart, Section 15.1.

We shall start the course by looking at **quadratic functions of two variables**. These have contours which are circles, ellipses, parabolas or hyperbolas.

1.1.2 Main points

- What is a function of several variables and why are they useful?
- Methods to visualise a function of two variables.
- Matlab commands `ezplot`, `ezsurf` and `ezcontour`.

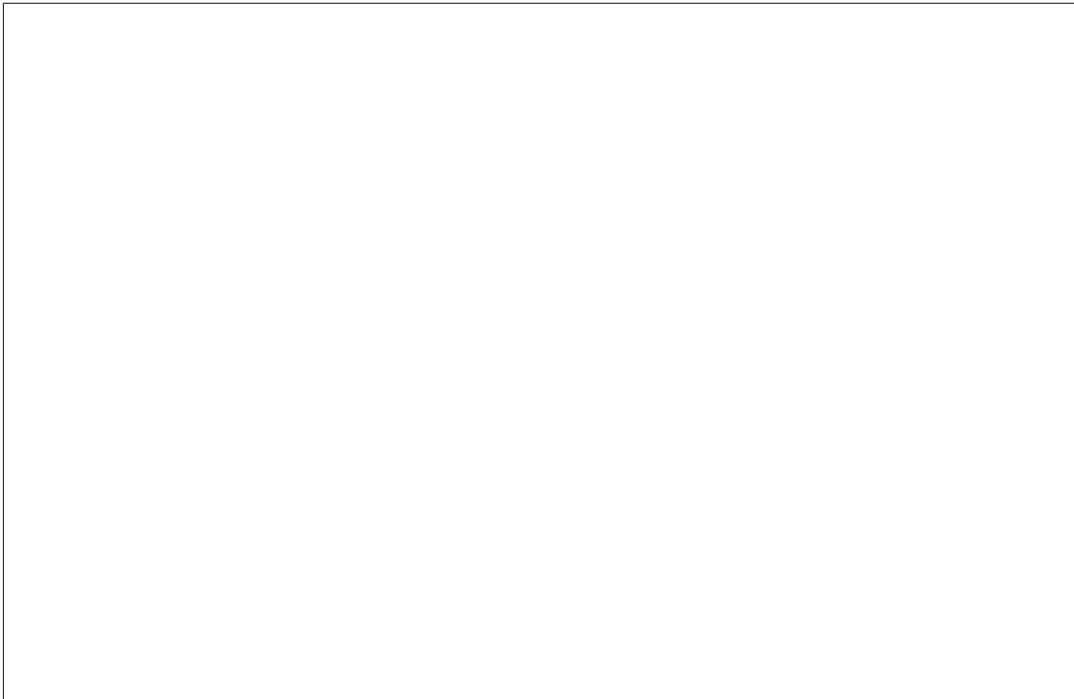
1.2 Conic Sections

The goal of this section is for you to become an expert in graphing parabolas, circles, ellipses and hyperbolas (without use of a graphics calculator, of course). The key-trick to this is “completing the square”. See also Stewart, Section 11.5.

1.2.1 Parabolas

Example: Sketch the following parabolas:

$$y = x^2, \quad y = x^2 + 3, \quad y = (x + 2)^2, \quad y = 2x^2, \quad y = 2(x + 2)^2 + 3.$$



Putting a quadratic $y = ax^2 + bx + c$ in the form

$$y = a(x - h)^2 + k,$$

is called **completing the square**.

Question: What is the interpretation of the point (h, k) ?

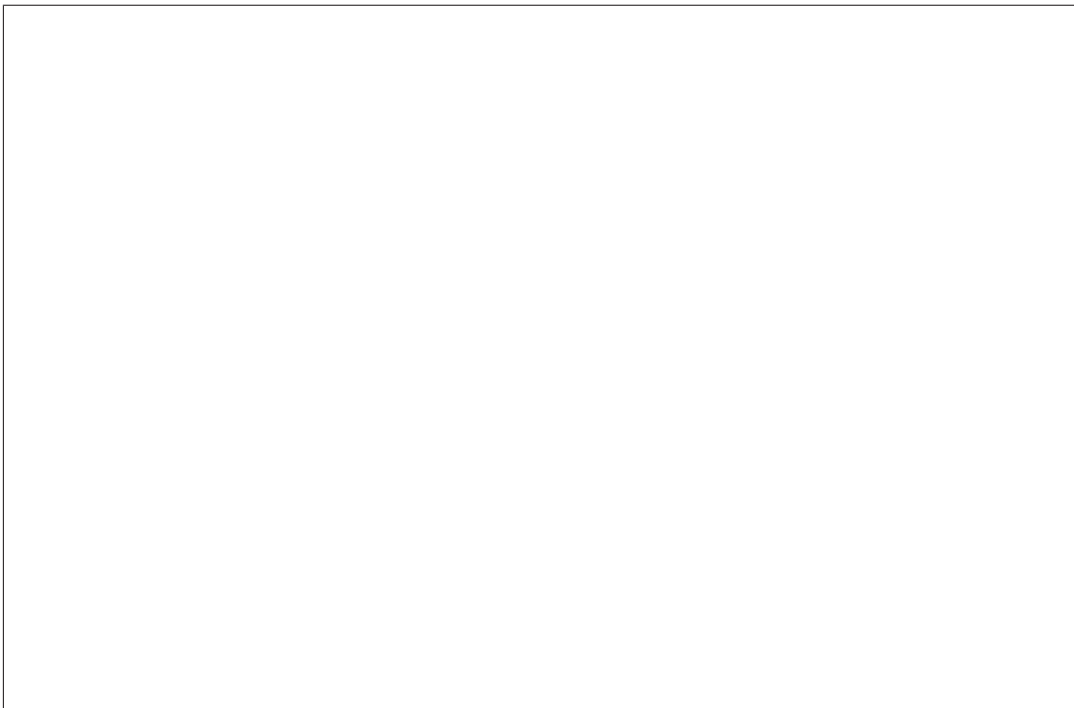
Completing the square makes use of the identity

$$x^2 + 2xh + h^2 = (x + h)^2.$$

Example: Sketch the parabola $y = x^2 + 6x + 8$.

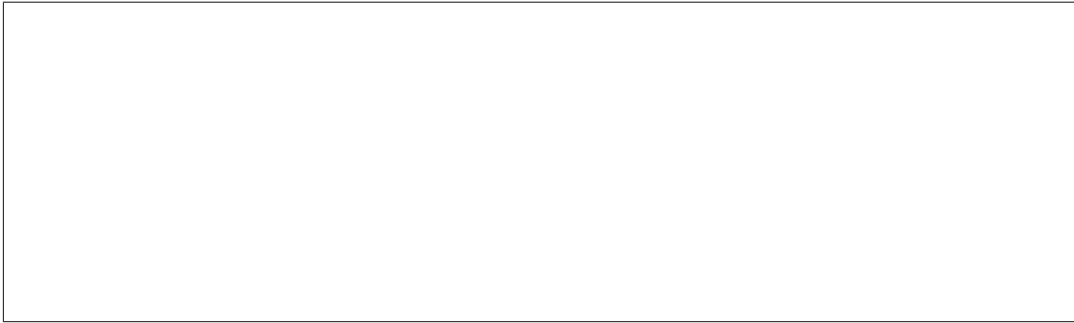


Example: Sketch the parabola $y = -3x^2 - 6x + 1$.



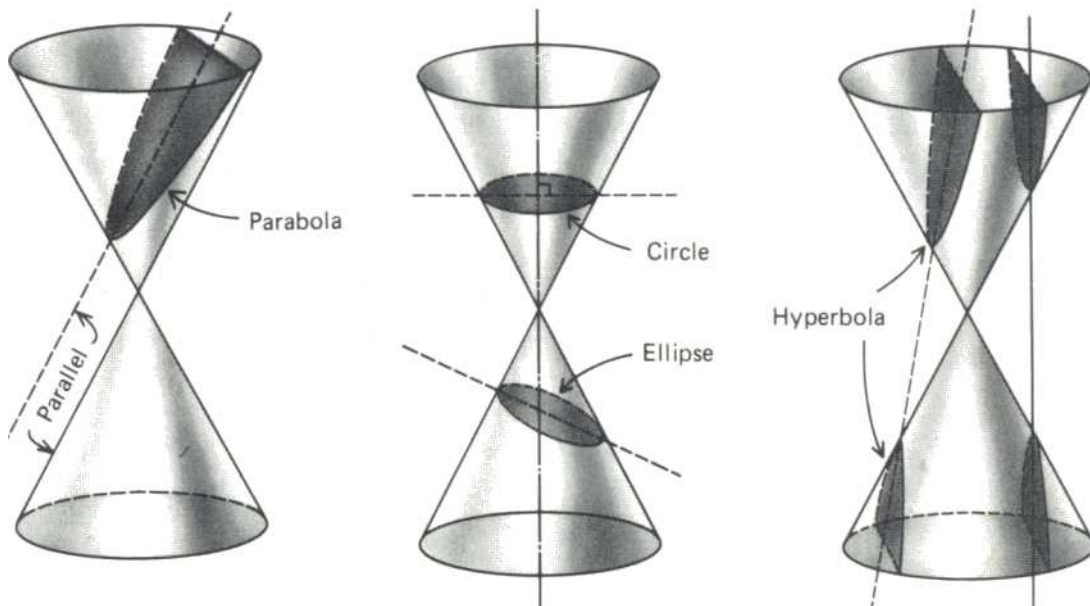
Example: Prove the quadratic formula. In other words, given $ax^2 + bx + c = 0$, by completing the square, show that

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$



Parabolas, circles, ellipses and hyperbolas are all examples of **conic sections**; they all arise by intersecting a double cone such as $z^2 = x^2 + y^2$, with a plane.

See the diagrams in Stewart p. 690. Slice an upright cone horizontally and you get a circle. Tip your knife a bit and you will get an ellipse. Keep tipping and you will get a parabola. Continue tipping and you will cut out a hyperbola.



1.2.2 Circles

- A circle centered at the origin $(0, 0)$ with radius r is given by the equation

$$x^2 + y^2 = r^2.$$

- A circle with radius r and center (h, k) corresponds to

$$(x - h)^2 + (y - k)^2 = r^2.$$

Example: Sketch the circle $x^2 + 10x + y^2 - 4y + 20 = 0$.



1.2.3 Ellipses

In standard form, an ellipse centered at the origin is given by the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$



An ellipse may also be defined as the set of points in a plane the sum of whose distances from two fixed focal points F_1 and F_2 is a constant, and you should definitely ask your lecturer to use two pins and a piece of string to demonstrate this. If $a \geq b$, then the foci are at $(\pm c, 0)$ where $c^2 = a^2 - b^2$. These lie on the major axis (x -axis here). The ellipse intersects the x axis at the vertices $(\pm a, 0)$, and the y -axis at the vertices $(0, \pm b)$.

Question: What is the equation of an ellipse centered at the point (h, k) ?

Example: Sketch the ellipse $x^2 - x + 9y^2 = 0$.



We could check this in Matlab with the command

```
ezplot('x^2-x+9*y^2=0', [-1,2], [-1,1])
```

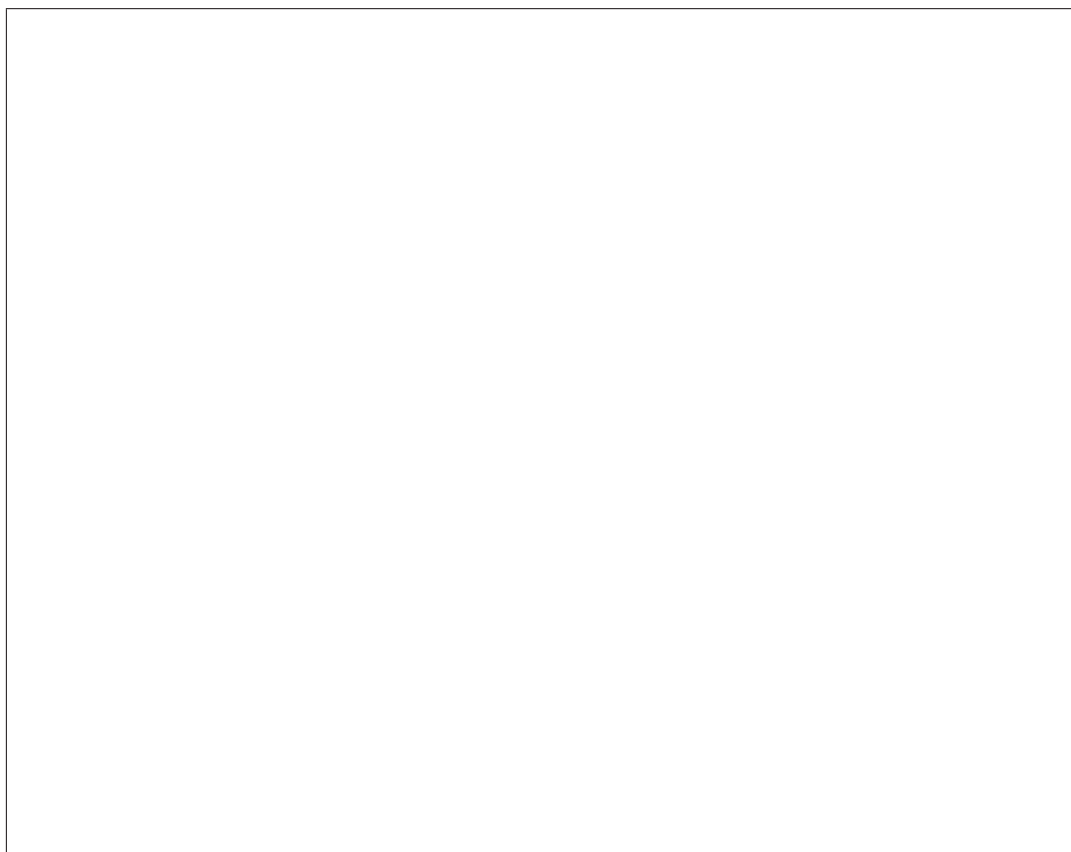
1.2.4 Hyperbolas

A hyperbola centered at the origin, with asymptotes $y = \pm bx/a$, is given by the equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

or by

$$-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$



A hyperbola may also be defined as the set of points in a plane, the difference of whose distances from two fixed focal points F_1 and F_2 is a constant. For such a hyperbola, the foci are at $(\pm c, 0)$ and $(0, \pm c)$ respectively, where $c^2 = a^2 + b^2$.

Question: What is the equation of a hyperbola centered at the point (h, k) ?

Exercise: Sketch the hyperbola $x^2 - 4y^2 = 9$.



Another kind of hyperbola is the very familiar $y = 1/x$.

1.2.5 Main points

- By completing the square, you should be able to
 - (i) sketch any curve of the form $ax^2 + by^2 + cx + dy + e = 0$,
 - (ii) identify this curve as a straight line (or lines), circle, parabola, ellipse or hyperbola, and
 - (iii) point out its main features such as turning points, asymptotes and intercepts with the x - and or y -axes.
- You should also be able to obtain the equation of a curve of the above form from a sketch with features given.

1.3 Scalar Equation for a Plane

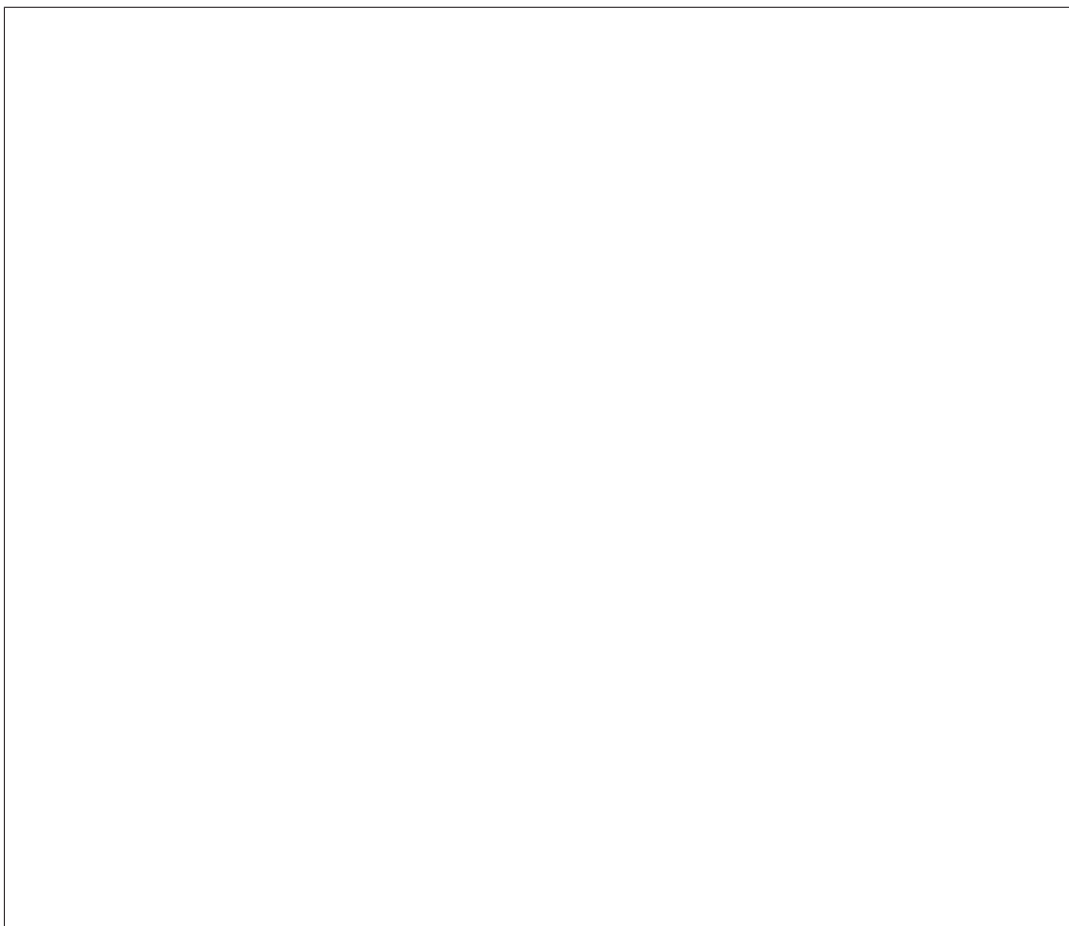
The equation of a plane can be expressed using scalars or vectors. In this section we will sketch planes in \mathbb{R}^3 and determine their scalar equations.

This material is covered in Stewart, Section 13.5.

1.3.1 Horizontal planes

The x - and y -axes lie in the horizontal plane $z = 0$. All other horizontal planes are parallel to $z = 0$ and are given by the equation $z = c$.

Example: Sketch the horizontal plane $z = 2$.



1.3.2 Vertical planes

A vertical plane has the form $ax + by = d$; it depends on x and y only and z does not appear. If you are not told this is an equation of a plane, or, equivalently, an equation in \mathbb{R}^3 , then you cannot distinguish it from the equation of a line in \mathbb{R}^2 .

Example: Sketch the plane $x + y = 1$.

First sketch the line $x + y = 1$ in the xy -plane; the plane contains this line and is parallel to the z axis.



1.3.3 Arbitrary planes

The general equation of a plane in \mathbb{R}^3 is given by

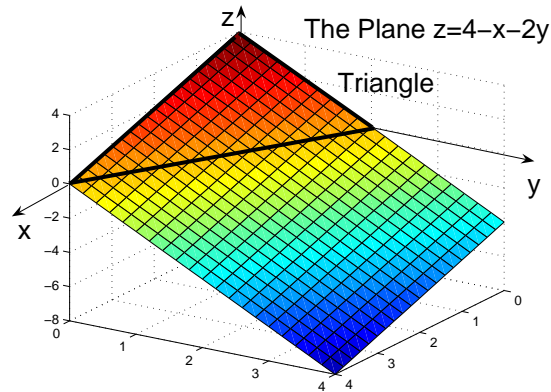
$$ax + by + cz = d$$

with a, b, c, d fixed real numbers. If the plane is not vertical, i.e., $c \neq 0$, this equation can be rearranged so that z is expressed as a function of x and y :

$$z = F(x, y) = -(a/c)x - (b/c)y + (d/c) = mx + ny + z_0.$$

To plot the plane in Matlab, simply plot $F(x, y)$ using `ezsurf`. The easiest way to sketch the plane by hand is to use the **triangle method**: If all of $a, b, c \neq 0$ the plane $ax + by + cz = d$ intercepts each axis at precisely one point. These three points make up a triangle which fixes the plane.

Example: The plane $x + 2y + z = 4$ intersects the x -axis at $x = 4$, the y -axis at $y = 2$ and the z -axis at $z = 4$.



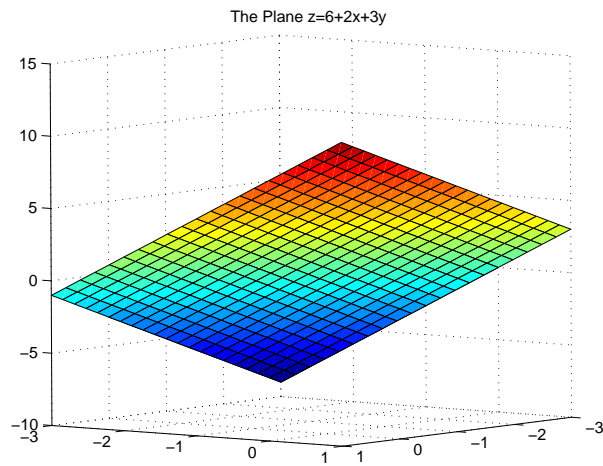
The triangle method is based on the simple fact that *any* three points that lie in a plane uniquely determine this plane *provided these three points do not lie on a single straight line*.

Example: Use the triangle method to sketch the plane $2x + 3y - z + 6 = 0$.

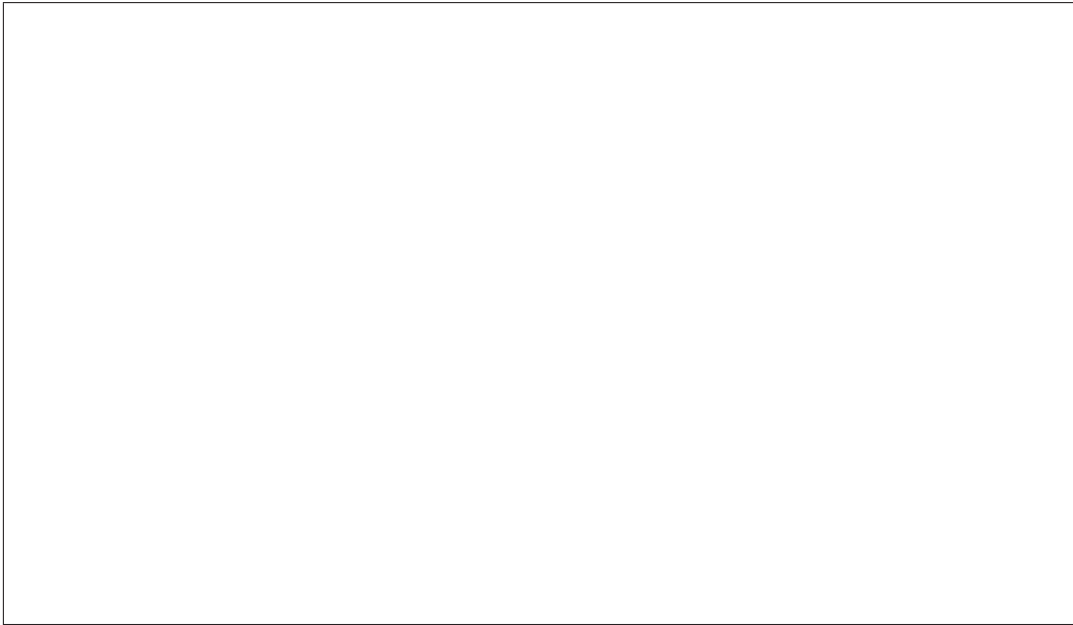


In Matlab, the graphing command is simply:

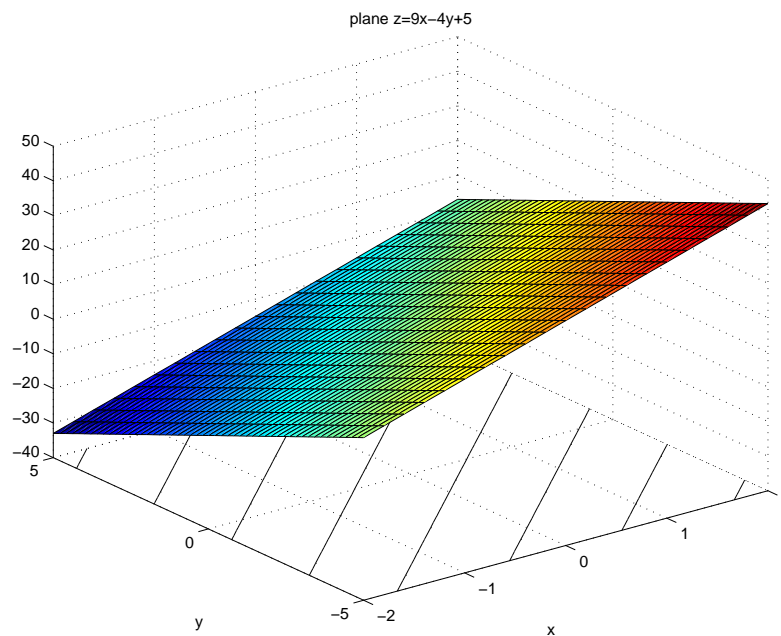
```
ezsurf('2*x+3*y+6', [-2,2,-5,5])
```



Example: Find the equation of the plane through $(0, 0, 5)$, $(1, 3, 2)$ and $(0, 1, 1)$.



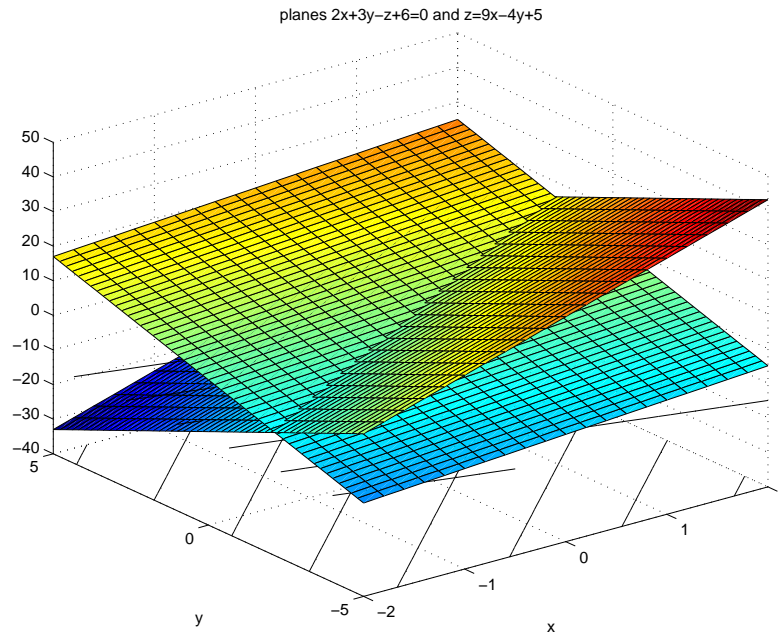
Important remark: It is customary to say *the* equation of a plane, even though it is not unique. Multiplying the equation of a plane by a nonzero constant gives another equation for the same plane. For example, $x - 2y + 3z = 4$ and $-2x + 4y - 6z = -8$ are equations of the same plane.



The Matlab command to produce a plot of this plane would be:

```
ezsurf('9*x-4*y+5', [-2,2,-5,5])
```

The following picture shows the intersection of the planes $2x + 3y - z + 6 = 0$ and $z = 9x - 4y + 5$. Note that the intersection is a line.



To produce two or more plots on the same figure, in Matlab use the hold command:

```
hold on;
```

1.3.4 Main points

- You should be able to draw horizontal, vertical and general planes by hand.
- You should be able to plot non-vertical planes in Matlab.
- You should be able to find the equation of a plane through three given points that are not on a straight line.
- You should know the scalar equation of a plane.

1.4 Vector Equation of a Plane I

For this section see also Stewart, Sections 13.1–13.5.

1.4.1 Revision of vectors

Pythagoras gives the distance d between the points (x_0, y_0) and (x_1, y_1) as

$$d = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}.$$

In three dimensions, the distance d between (x_0, y_0, z_0) and (x_1, y_1, z_1) is

$$d = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2}.$$

Exercise: Find the distance between the points $(1, 2, 3)$ and $(0, 0, 0)$.

The **norm** of a vector is its length. In Matlab, you can directly call the norm of a vector:

```
v=[1 2 3]
v =
     1     2     3
norm(v)
ans =
     3.7417
```

Recall that the **dot product** of two vectors \mathbf{a} and \mathbf{b} is given by

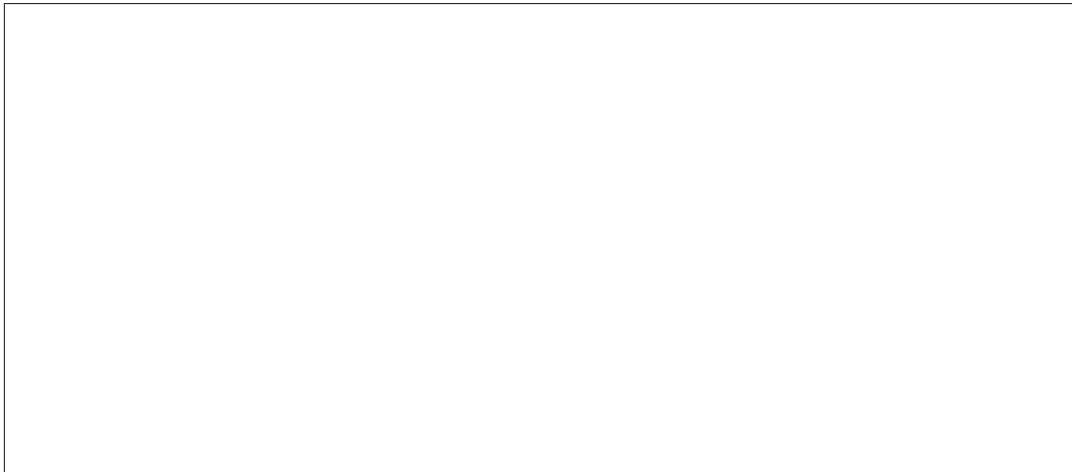
$$\mathbf{a} \cdot \mathbf{b} = (a_1, a_2, a_3) \cdot (b_1, b_2, b_3) = a_1b_1 + a_2b_2 + a_3b_3.$$

It is also true that

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta,$$

where θ is the angle between \mathbf{a} and \mathbf{b} , and $\|\mathbf{a}\|$ is the **norm** of \mathbf{a} :

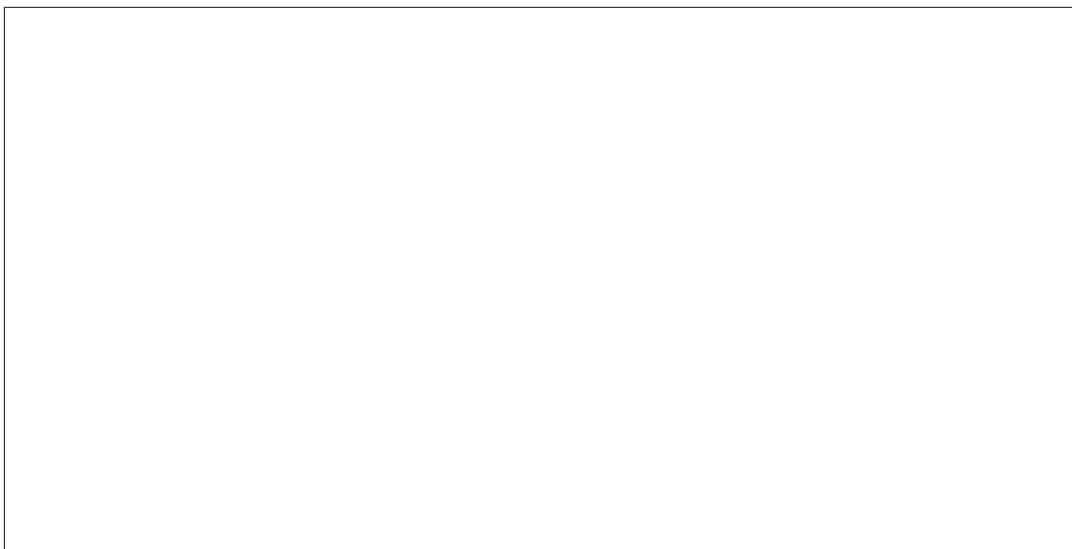
$$\|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + a_3^2}.$$



If $\mathbf{a} \cdot \mathbf{b} = 0$, then \mathbf{a} and \mathbf{b} are **orthogonal**.

The **projection** of \mathbf{b} onto \mathbf{a} is “the component of \mathbf{b} in the direction of \mathbf{a} ”:

$$\text{proj}_{\mathbf{a}} \mathbf{b} = \|\mathbf{b}\| \cos \theta \times \frac{\mathbf{a}}{\|\mathbf{a}\|} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|^2} \mathbf{a}.$$



Recall that the **cross product** $\mathbf{a} \times \mathbf{b}$ of two vectors in \mathbb{R}^3 is a vector orthogonal to both \mathbf{a} and \mathbf{b} given by

$$\mathbf{a} \times \mathbf{b} = (a_1, a_2, a_3) \times (b_1, b_2, b_3) = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

This is best memorised using a 3 by 3 determinant.



The cross product points in the direction given by the right-hand-rule.

It is also true that

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \theta.$$

This means that $\|\mathbf{a} \times \mathbf{b}\|$ gives the area of the parallelogram spanned by the vectors \mathbf{a} and \mathbf{b} .

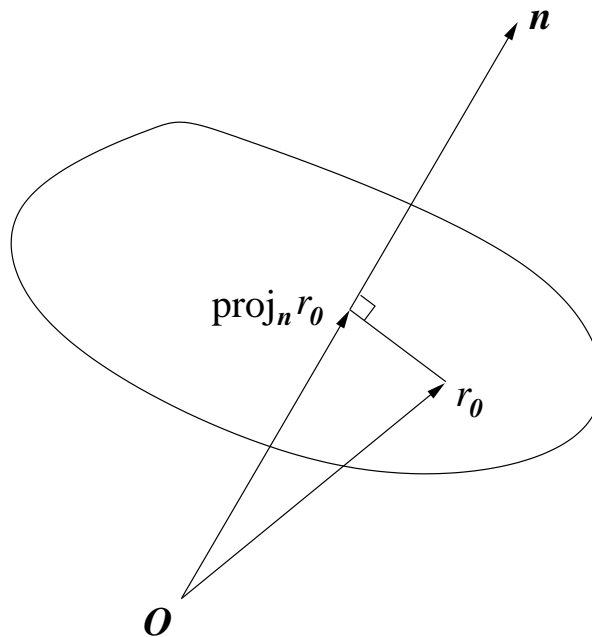
Example: Compute the area of the triangle with vertices $(1, 1)$, $(2, 3)$, $(4, 2)$.



1.4.2 Vector equation of a plane I

We have seen in Section 1.3 that any three points in \mathbb{R}^3 which do not lie on a straight line determine a plane. There is another geometric way to view a plane, which gives rise to the **vector equation of a plane**.

Let P be a plane and \mathbf{n} be a vector perpendicular to P . Such a vector is called a **normal** to P . Let \mathbf{r}_0 be a vector from the origin to a point r_0 in P and $\text{proj}_{\mathbf{n}}\mathbf{r}_0$ its projection onto \mathbf{n} . Then the length of this projection vector, i.e., $\|\text{proj}_{\mathbf{n}}\mathbf{r}_0\|$ is the distance of the plane to the origin.



If we had taken another vector, say \mathbf{r} , from the origin to a point r in P then

$$\text{proj}_{\mathbf{n}}\mathbf{r} = \text{proj}_{\mathbf{n}}\mathbf{r}_0.$$

In other words, the plane P is given by the collection of *all* points r whose corresponding vectors \mathbf{r} have the same orthogonal projection onto \mathbf{n} as \mathbf{r}_0 . Hence

$$\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0.$$

This is the **vector equation** for a plane.

The vector and scalar equations are in fact the same, as shown next.

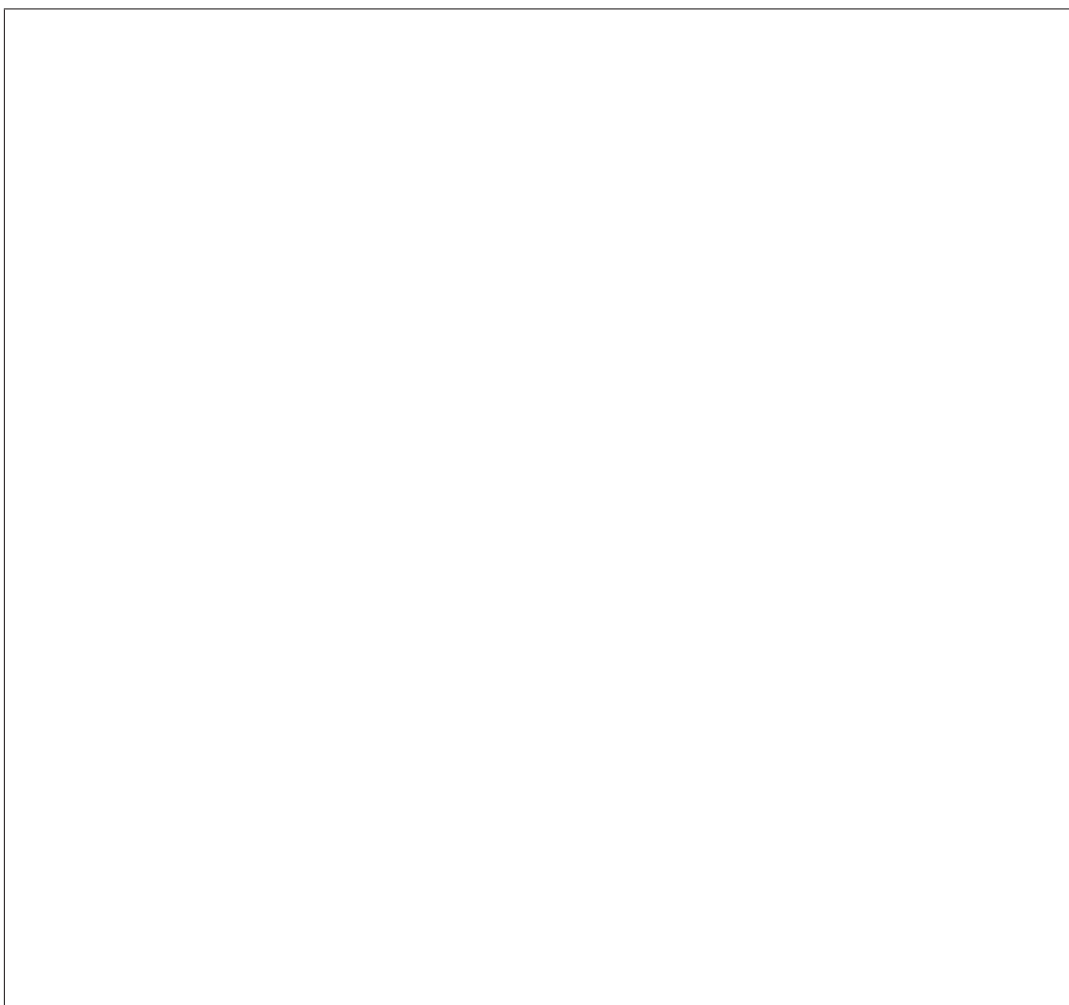
We start with the vector equation, where $\mathbf{n} = (a, b, c)$ is a normal, $\mathbf{r}_0 = (x_0, y_0, z_0)$ and $\mathbf{r} = (x, y, z)$. Then

$$\begin{aligned}\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) &= 0 \\ (a, b, c) \cdot (x - x_0, y - y_0, z - z_0) &= 0 \\ a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0 \\ ax + by + cz &= d\end{aligned}$$

where $d = \mathbf{n} \cdot \mathbf{r}_0 = ax_0 + by_0 + cz_0$.

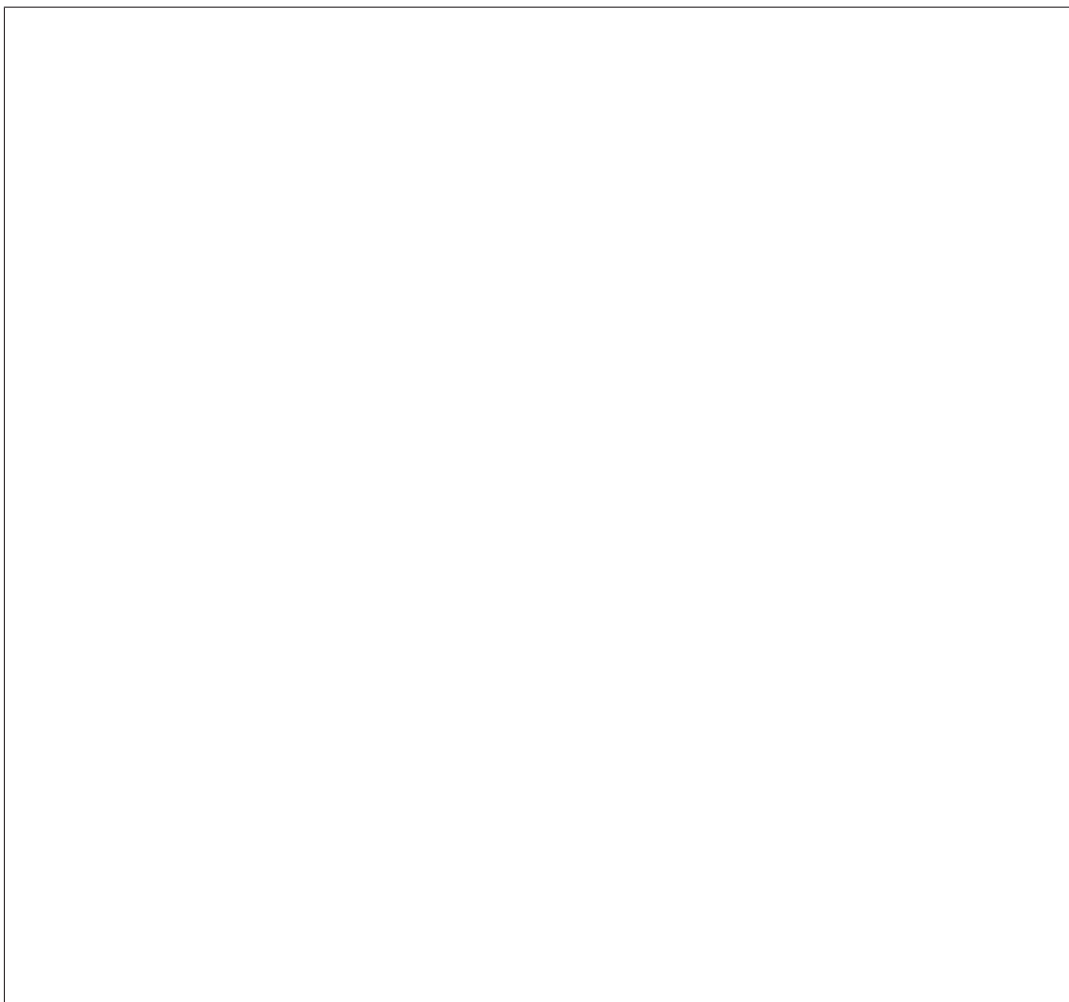
Important remark: Given a plane $ax + by + cz = d$ we now have a geometric interpretation of the vector (a, b, c) : it is a normal to the plane. Obviously the normal is not unique since any vector of the form $k(a, b, c)$, where k is a nonzero scalar, is also a normal.

Example: Find the vector and scalar equations for the plane passing through the point $(2, 4, -1)$ with normal $\mathbf{n} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k} = (2, 3, 4)$.



Our understanding of normal vectors allows us to compute angles between planes, see also Stewart Section 13.5.

Example: Find the angle between the planes $x+4y-3z = 1$ and $-3x+6y+7z = 0$.



Important remark: Contrary to popular belief, the angle between two planes cannot exceed $\pi/2$. It is a cardinal offense to ever hand in work that claims otherwise.

1.4.3 Main points

- You should be able to use the dot product, projections and the cross product.
- You should be able to derive the scalar equation from the vector equation of a plane.
- You should be able to find the equation of a plane given sufficient information.

1.5 Equations for a Line

There are three common ways to represent a line:

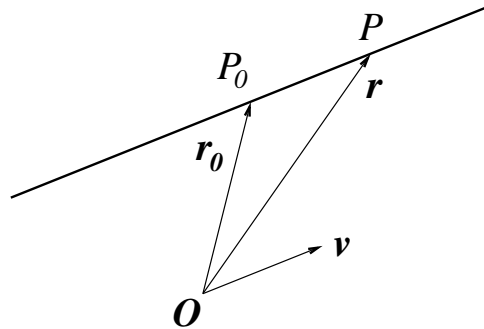
- (a) the vector representation;
- (b) the parametric representation;
- (c) symmetric equations, obtained by eliminating parameters in (b).

Which representation is best depends on the particular problem at hand.

(a) Vector representation

Let $P_0 = (x_0, y_0, z_0)$ be a point on the line L with corresponding position vector \mathbf{r}_0 and let \mathbf{v} be a vector parallel to L , known as a **direction vector**. For an arbitrary point $P = (x, y, z)$ lying on L we have

$$\overrightarrow{P_0P} = (x - x_0, y - y_0, z - z_0).$$



Since $\overrightarrow{P_0P}$ is parallel to \mathbf{v} , we have $\overrightarrow{P_0P} = \lambda\mathbf{v}$ for some scalar λ . Hence

$$\mathbf{r} = \mathbf{r}_0 + \overrightarrow{P_0P} = \mathbf{r}_0 + \lambda\mathbf{v}.$$

The equation

$$\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{v}, \quad \lambda \in \mathbb{R}$$

is known as the **vector equation** of the line L .

Important remark: As with *the* equation of a plane, *the* vector equation of a line is not unique. One can choose *any* point P_0 on the line as “starting point” and one can multiply the vector \mathbf{v} by any nonzero constant.

Example: Find (a) the vector representation of the line passing through $(5, 1, 3)$ and parallel to the vector $(1, 4, -2)$ and (b) the vector representation of the line passing through $(4, -3, 5)$ and parallel to the vector $(-2, -8, 4)$. What is the relation between the two lines?



(b) Parametric representation

The **parametric representation** of a line is a scalar representation of the vector equation $\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{v}$. Writing each vector in component form,

$$\begin{aligned}\mathbf{r} &= (x, y, z) \\ \mathbf{r}_0 &= (x_0, y_0, z_0) \\ \mathbf{v} &= (a, b, c)\end{aligned}$$

gives

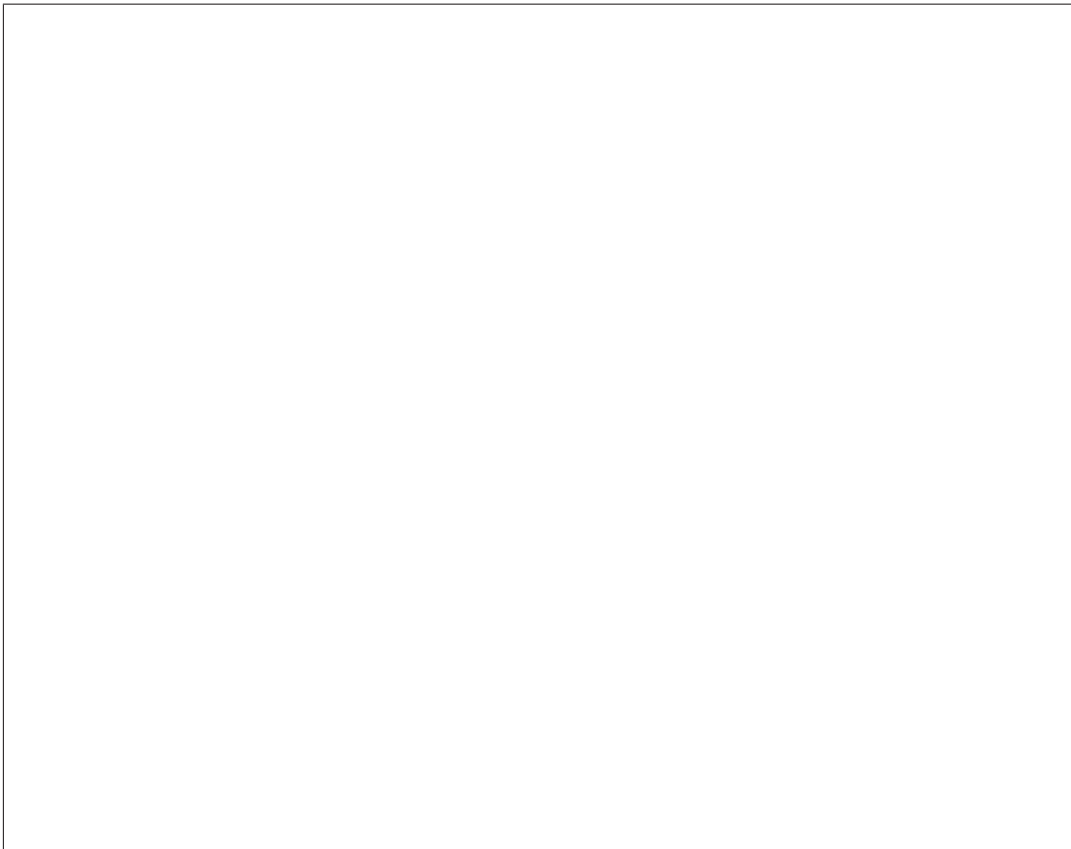
$$(x, y, z) = (x_0, y_0, z_0) + \lambda(a, b, c).$$

Matching components results in three scalar equations

$$\begin{cases} x = x_0 + a\lambda \\ y = y_0 + b\lambda \\ z = z_0 + c\lambda \end{cases} \quad (*)$$

known as the parametric equations of a line.

Example: Find the parametric equations for the line passing through $(5, 1, 3)$ and parallel to $\mathbf{i} + 4\mathbf{j} - 2\mathbf{k}$.



(c) Symmetric equations

The parameter λ can be eliminated from the parametric equations of a line. For example, by eliminating λ from each of the three equations $x = 5 + \lambda$, $y = 1 + 4\lambda$, $z = 3 - 2\lambda$ for the line L , we obtain

$$\lambda = x - 5 = \frac{y - 1}{4} = \frac{z - 3}{-2}.$$

The equations

$$x - 5 = \frac{y - 1}{4} = \frac{z - 3}{-2}$$

are known as the **symmetric equations** of L .

What these equation really are is a set of two non-identical, non-parallel planes

$$x - 5 = \frac{y - 1}{4} \quad \text{and} \quad x - 5 = \frac{z - 3}{-2},$$

which, as we know, must intersect to give a line.

The general form of the symmetric equations of the line (*) from the previous page is given by

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

but some care is needed when one (or two) of a , b or c is equal to zero. For example, if $a = 0$ the above needs to be replaced by

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

and so on.

Example: Find the symmetric equations of the line that passes through the points $(2, 4, -3)$ and $(3, -1, 1)$.



Example: Find the symmetric equations of the line $\mathbf{r} = (1, 2, 3) + \lambda(0, 1, 0)$.



To plot a line in three-dimensional space, use the parametric form. The Matlab command is `ezplot3`. As input it takes three functions of a parameter, and the range of the parameter.

Example: Plot the line

$$\begin{cases} x = 100 + 200t \\ y = 200 + 300t \\ z = 300 - 10t \end{cases}$$

for t between -100 and 100 . The Matlab command would be

```
ezplot3('100+200*t', '200+300*t', '300-10*t', [-100,100])
```

You can grab and rotate the resulting axes.

1.5.1 Parallel, skew and orthogonal Lines.

Two lines are **parallel** if, when written as $\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{v}$, and $\mathbf{r} = \hat{\mathbf{r}}_0 + \mu\mathbf{u}$, the direction vectors \mathbf{u} and \mathbf{v} are linear multiples of each other, i.e., $\mathbf{u} = k\mathbf{v}$ for some nonzero scalar k .

Two straight lines in three dimensional space rarely intersect. Non parallel, non-intersecting lines are called **skew lines**.

Example: Show that the lines L_1 and L_2 given by

$$L_1 : \begin{cases} x = 1 + \lambda \\ y = -2 + 3\lambda \\ z = 4 - \lambda \end{cases} \quad L_2 : \begin{cases} x = 2\mu \\ y = 3 + \mu \\ z = -3 + 4\mu \end{cases}$$

are skew lines.



Two lines are **orthogonal** if their direction vectors are orthogonal. An easy test for orthogonality is to take the dot product of the direction vectors; if and only if this gives zero are the lines orthogonal.

Example: Show that the lines L_1 and L_2 given by

$$L_1 : \frac{x-1}{2} = y-3 = \frac{z+5}{4},$$

$$L_2 : 2-x = \frac{5-y}{2} = z-5.$$

are orthogonal.



Further theory

Unlike lines, planes can not be skew in 3 dimensions. Given two planes, the only possibilities are that they intersect, or that they are parallel. If two planes are parallel, then their normal vectors must be parallel. If two (distinct) planes intersect, then they intersect in a line.

Example: Find the parametric equations for the intersection of the planes

$$z = x + y \quad \text{and} \quad 2x - 5y - z = 1.$$



1.5.2 Vector equation of a plane II

Previously we discussed the scalar equation of a plane, which is an equation of the form $ax + by + cz = d$, and vector equation a plane, which is an equation of the form $\mathbf{n} \cdot (\mathbf{r} - \mathbf{r}_0) = 0$.

There is a second type of vector equation of a plane, which is similar in form to the vector equation of a line. One fixes a point P_0 on the plane with corresponding position vector \mathbf{r}_0 , but now not one, but two (non-parallel) vectors \mathbf{u} and \mathbf{v} parallel to the plane are required to fully determine the plane:

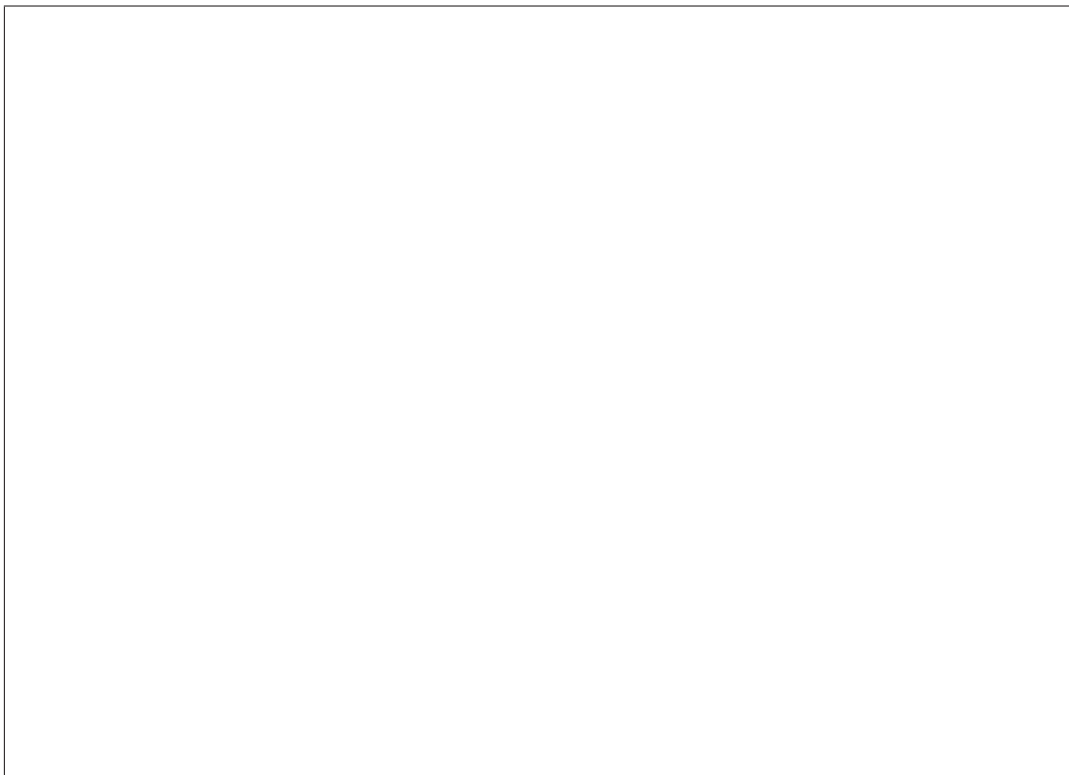
$$\mathbf{r} = \mathbf{r}_0 + \lambda\mathbf{u} + \mu\mathbf{v}, \quad \lambda, \mu \in \mathbb{R}.$$

Example: Sketch a figure explaining the above equation geometrically.

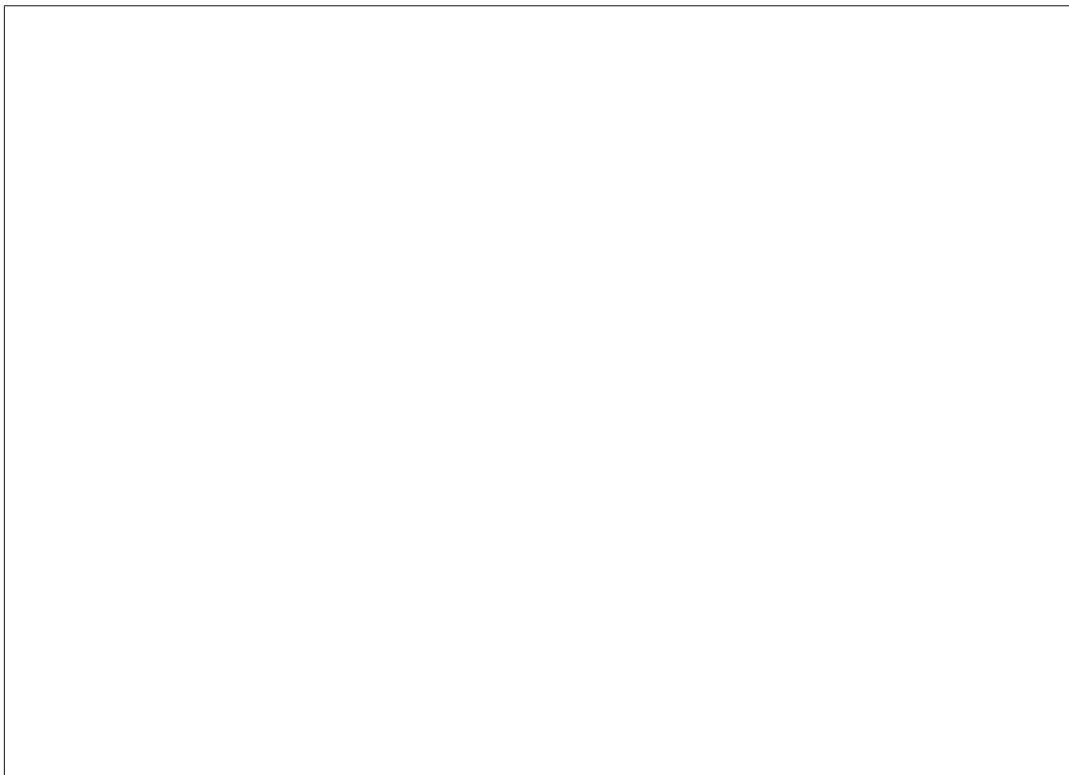


Important remark: In some sense the above vector equation of a plane is even less unique than the vector equation of a line. Again one can take \mathbf{r}_0 to be any point on the plane and one can multiply both \mathbf{u} and \mathbf{v} by nonzero constants. But unlike a line, we can also replace \mathbf{u} and \mathbf{v} by any other pair of (independent) vectors parallel to the plane. For example, $(x, y, z) = (1, 0, 1) + \lambda(0, 1, 2) + \mu(3, 1, -1)$ and $(x, y, z) = (4, 1, 0) + \lambda(-1, 0, 1) + \mu(-3, 1, 5)$ are different vector representations of the same plane! So do not panic if in your exam the person next to you writes down a very different-looking equation; they still might have the right answer...

Example: Find the second type of vector equation of the plane through $(0, 0, 5)$, $(1, 3, 2)$ and $(0, 1, 1)$.



Example: Find the second type of vector equation for the plane $2x + y - z = 4$.



1.5.3 Distance from a Point to a Plane

A common problem that arises in applications is to find the distance from a point to a plane, where, by **distance**, we always mean **minimum distance**.

If $P = (x_1, x_2, x_3)$ is a point, with corresponding position vector \mathbf{p} , and Π is a plane (with normal $\mathbf{n} = (a, b, c)$) given by $ax + by + cz = d$, then the formula for the distance between P and Π is

$$D = \frac{|\mathbf{n} \cdot \mathbf{p} - d|}{\|\mathbf{n}\|} = \frac{|ax_1 + by_1 + cz_1 - d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Important remark: It is best not to try to memorise such a formula because, oddly enough, under exam pressure small details often subtly change for the worse. Also, it depends on one's choice of representation of a plane. Stewart, for example writes $ax + by + cz + d = 0$ in which case the numerator contains $+d$ instead of $-d$. Best is to *understand* the actual derivation given below.

One derivation of the above distance formula using orthogonal projections is given in Stewart, Section 13.5. Below we present an alternative method, which is not quite as slick, but which is very easy to carry out for explicit examples, even under exam conditions.

First we construct the line L through the point P and orthogonal to the plane Π :

$$L : \quad \mathbf{r} = \mathbf{p} + \lambda \mathbf{n}, \quad \lambda \in \mathbb{R}.$$

Next we determine the point, say Q , where L intersects Π . That is, we substitute the equation for L into the equation $\mathbf{r} \cdot \mathbf{n} = d$ for Π :

$$(\mathbf{p} + \lambda \mathbf{n}) \cdot \mathbf{n} = d.$$

Solving for λ yields

$$\lambda = \frac{d - \mathbf{p} \cdot \mathbf{n}}{\|\mathbf{n}\|^2},$$

so that the point Q , with position vector \mathbf{q} , is given by

$$\mathbf{q} = \mathbf{p} + \frac{d - \mathbf{n} \cdot \mathbf{p}}{\|\mathbf{n}\|^2} \mathbf{n}.$$

The distance between P and Π is now the distance between P and Q . But

$$\mathbf{p} - \mathbf{q} = \frac{\mathbf{n} \cdot \mathbf{p} - d}{\|\mathbf{n}\|^2} \mathbf{n}$$

so that

$$D^2 = \|\mathbf{p} - \mathbf{q}\|^2 = \frac{(\mathbf{n} \cdot \mathbf{p} - d)^2}{\|\mathbf{n}\|^4} \mathbf{n} \cdot \mathbf{n} = \frac{(\mathbf{n} \cdot \mathbf{p} - d)^2}{\|\mathbf{n}\|^2}.$$

The formula now follows by taking the square root on both sides *and remembering that a distance can never be negative*, whereas $\mathbf{n} \cdot \mathbf{p} - d$ can.

Example: Find the distance from the point $(1, 2, 3)$ to the plane $z = 4$.

Important remark: In assignments or exams we do not provide boxes whose size might indicate how hard or how trivial an actual question is. Hence it is important to always think first before you start to calculate. Sometimes, just thinking will be enough.

Example: Find the distance between the point $(1, 2, 3)$ and the plane $x + y + z = 1$.

1.5.4 Main points

- You should know the various representations for a line.
- Given one representation of a line, you should be able to find the other representations.
- You should be able to find the equation for a line given sufficient information.
- You should know how to find the intersection of two planes or a plane and a line.
- You should be able to find the distance from a point to a plane, or between two planes.

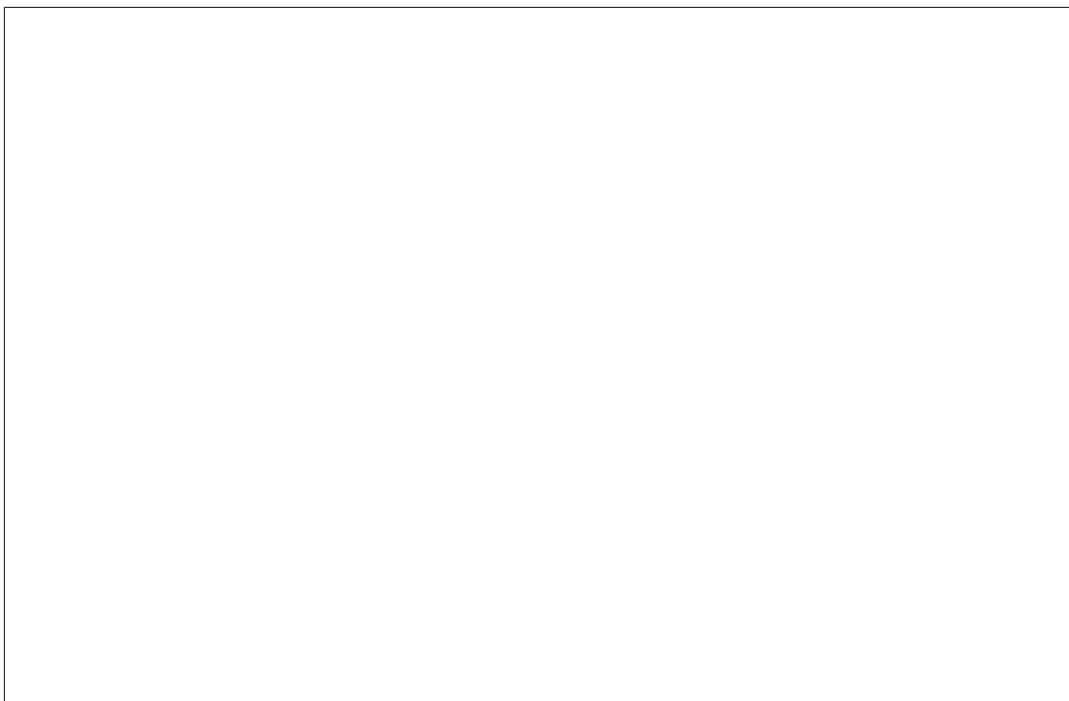
1.6 Contour Diagrams

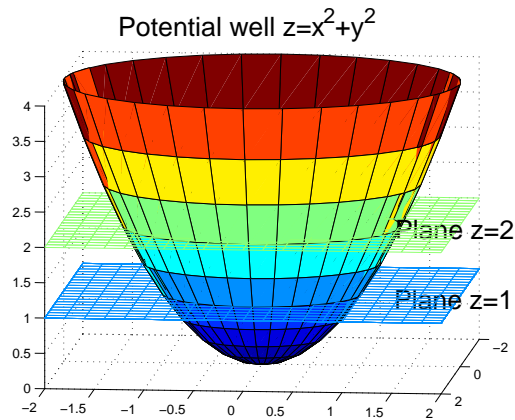
Geographical maps have curves of constant height above sea level, or curves of constant air pressure (isobars), or curves of constant temperature (isothermals). Drawing contours is an effective method of representing a 3-dimensional surface in two dimensions. We now look at functions f of two variables. A **contour** is a curve corresponding to the equation $z = f(x, y) = C$, see also Stewart, Section 15.1.

Consider the surface $z = f(x, y) = x^2 + y^2$ sliced by horizontal planes $z = 0, z = 1, z = 2, \dots$

Plane	Contour	Description
$z = 0$		
$z = 1$		
$z = 2$		
$z = 3$		
$z = 4$		

Note that as the radius increases, the contours are more closely spaced.





Example: Draw a contour diagram of f given by

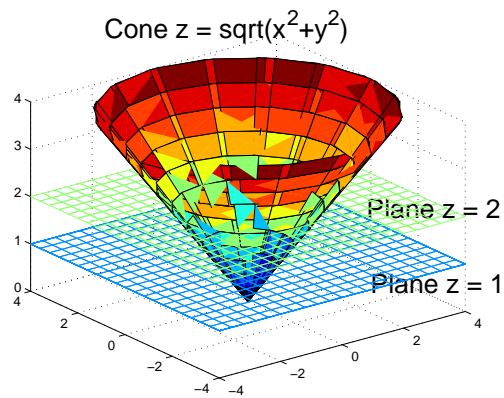
$$f(x, y) = \sqrt{x^2 + y^2}.$$

In Matlab, the command is

```
ezcontour('sqrt(x^2+y^2)', [-10,10,-10,10]);
```

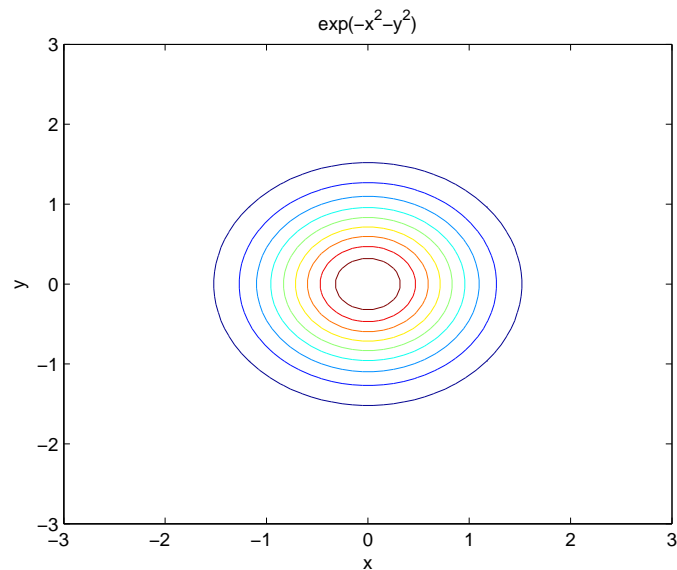
If horizontal planes are equally spaced, say $z = 0, c, 2c, 3c, \dots$, it is not hard to visualise the surface from its contour diagram. Spread-out contours mean the surface is quite flat and closely spaced ones imply a steep climb.

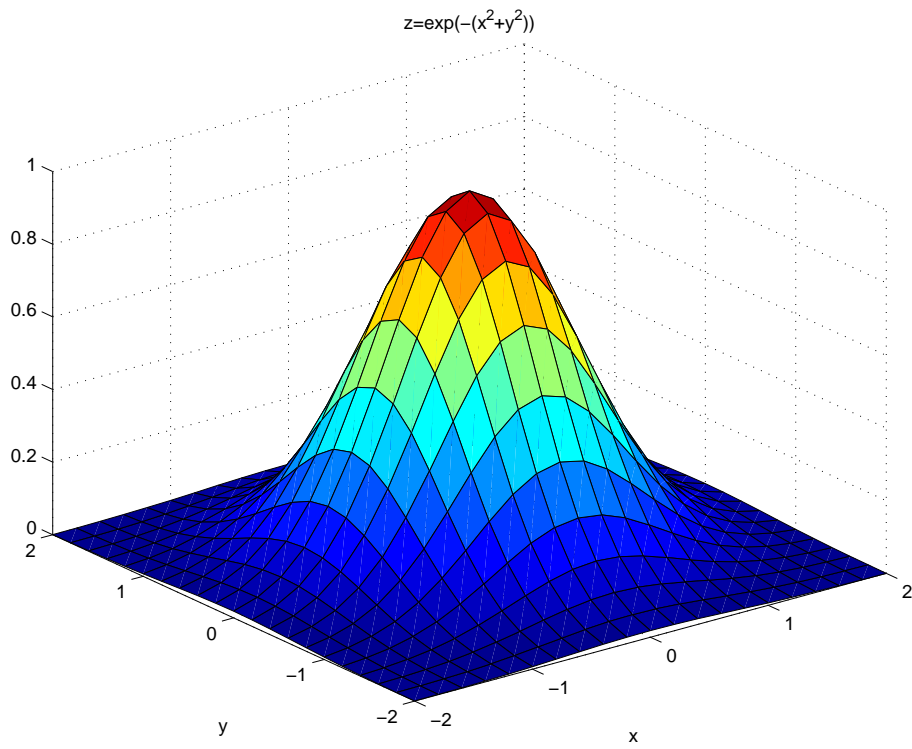




Note that the contours of the last two functions were all circles. Such surfaces have **circular symmetry**. When x and y only appear as $x^2 + y^2$ in the definition of f , then the graph of f has circular symmetry about the z axis. The height z depends only on the radial distance $r = \sqrt{x^2 + y^2}$.

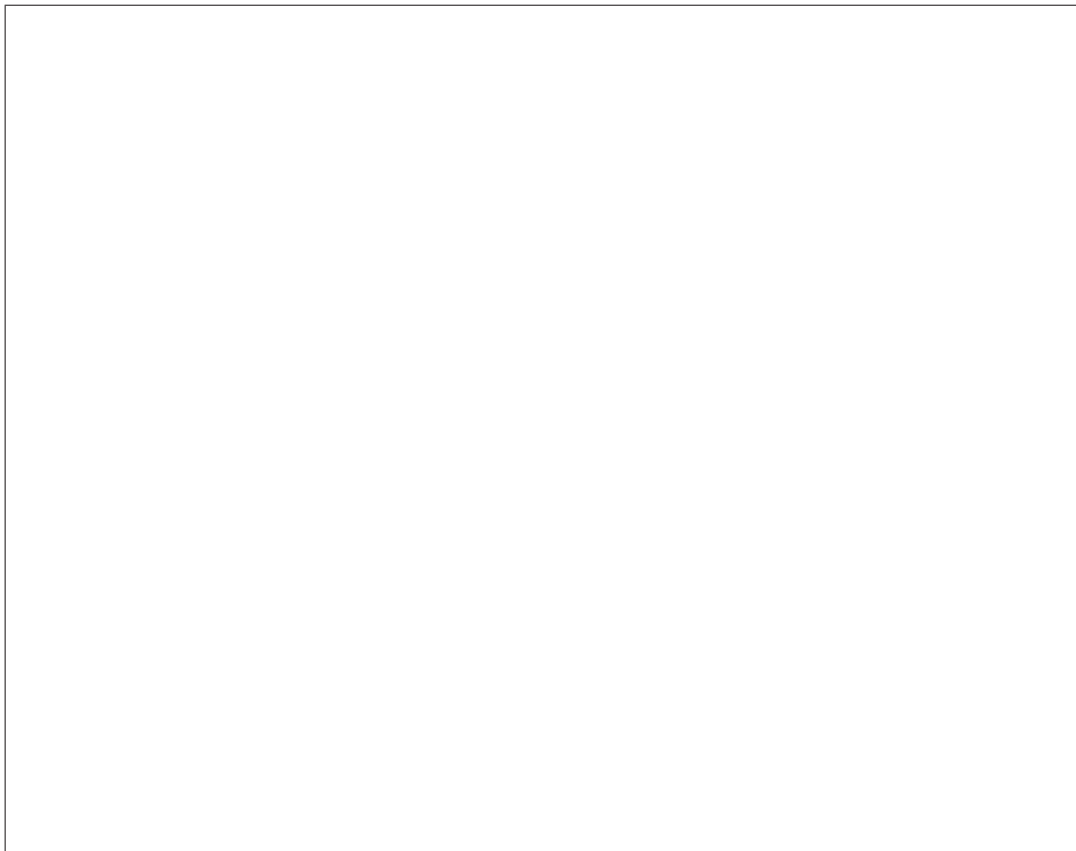
Example: $z = e^{-x^2 - y^2}$





Example: Draw a contour diagram for $z = f(x, y) = x^2 + 4y^2 - 2x + 1$.

Plane	Contour	Description
$z = 1$		
$z = 2$		
$z = 3$		



Example: A **saddle** $z = x^2 - y^2$ has hyperbolic contours.

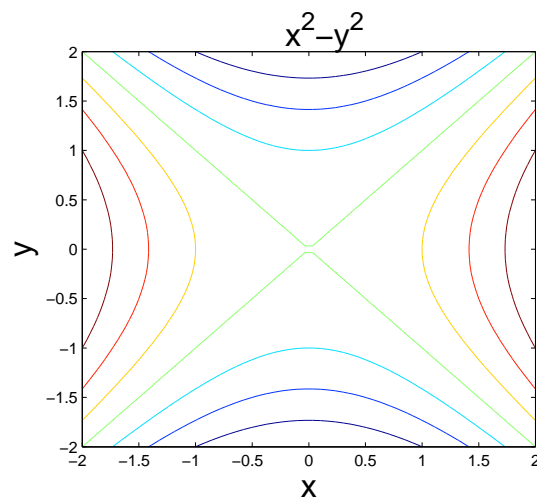
$$z = 0 :$$

$$z = 1 :$$

$$z = 2 :$$

$$z = -1 :$$

$$z = -2 :$$



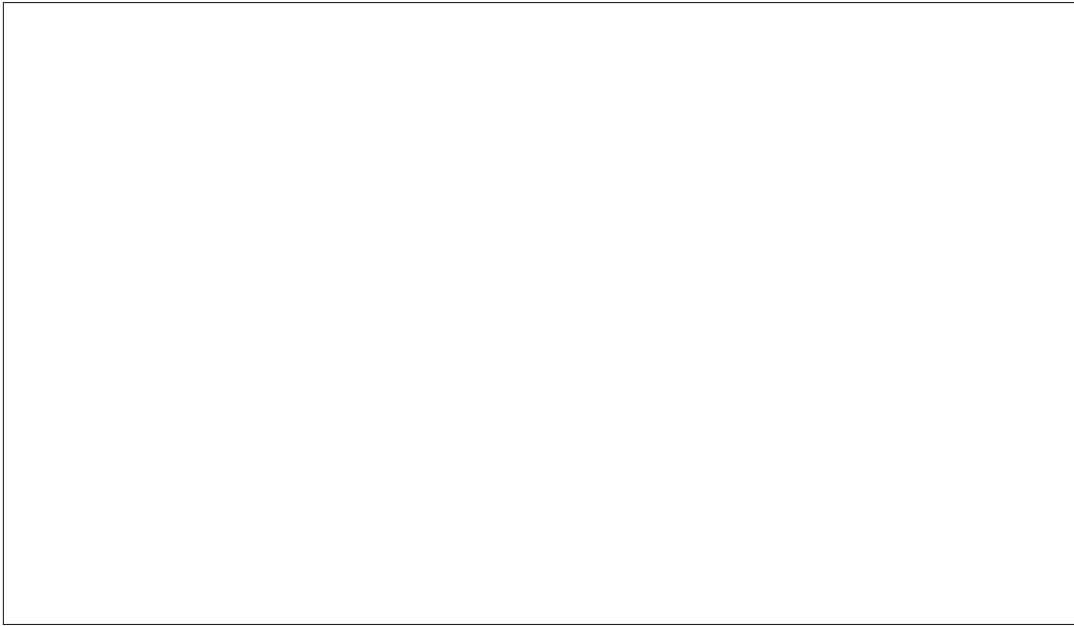
Example: Sketch the contour diagram for $z = f(x, y) = 9x^2 - 4y^2 + 2$ with contours at $z = -2, 2, 6$ and 10 .



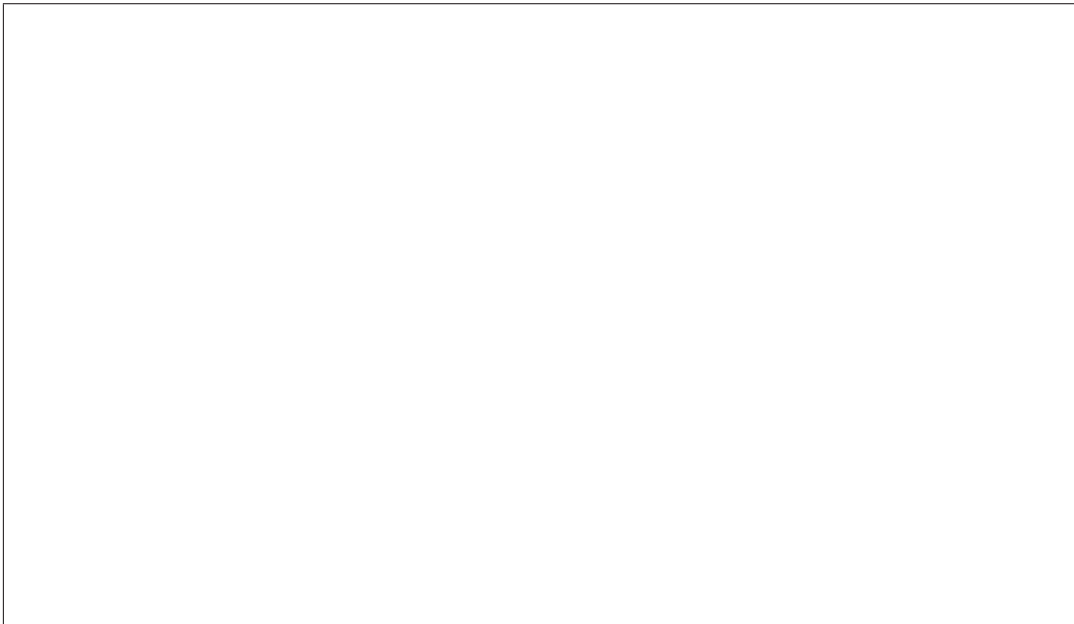
Example: Sketch a contour diagram for $z = f(x, y) = x^2$ and use this to sketch the graph of f .



Example: Sketch a contour diagram for $z = f(x, y) = x - y^2$.



Example: Sketch contour diagrams of $z = f(x, y) = x$ and $z = g(x, y) = x + y$.



Note: Contour diagrams of functions whose graphs are planes consist of **equidistant** parallel lines. (Equidistant lines in \mathbb{R}^2 are in fact always parallel, unlike \mathbb{R}^3 . Why?)

1.6.1 Main points

- You should be able to plot contour diagrams in Matlab using the `ezcontour` function (and its variants).
- You should be able to recognise circular symmetry in an equation.
- You should be able to match contour diagrams with functions.
- You should be able to sketch simple contour diagram.

1.7 Cross-sections of a Surface

A **cross-section** is the intersection of a surface with a vertical plane such as $y = C$, see also Stewart Section 13.6.

Example:

The height z of a vibrating guitar string can be expressed as a function of horizontal distance x , and time t

$$z = f(x, t) = A \sin x \cos t \quad \text{where} \quad 0 < x < \pi.$$

The snapshots where t is constant are cross-sections of the 'surface'.

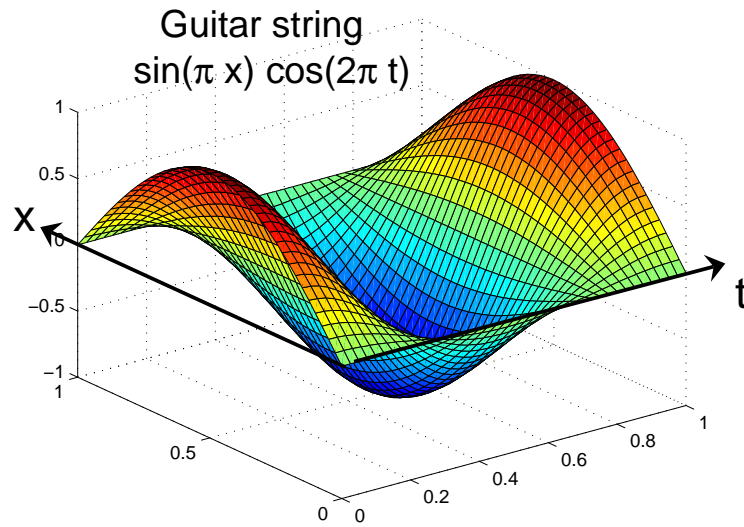


Varying time we get

$$\begin{aligned} t = 0 : & \quad z = A \sin x \\ t = \frac{1}{4}\pi : & \quad z = \frac{1}{2}\sqrt{2}A \sin x \\ t = \frac{1}{2}\pi : & \quad z = 0 \\ t = \frac{3}{4}\pi : & \quad z = -\frac{1}{2}\sqrt{2}A \sin x. \end{aligned}$$

These represent sine curves, with amplitudes between 0 and A .

We can also consider the cross-sections in x . For instance $x = \frac{1}{2}\pi$ (at the top of the sine wave), then $z = A \cos t$ which equals the amplitude of the sine wave.

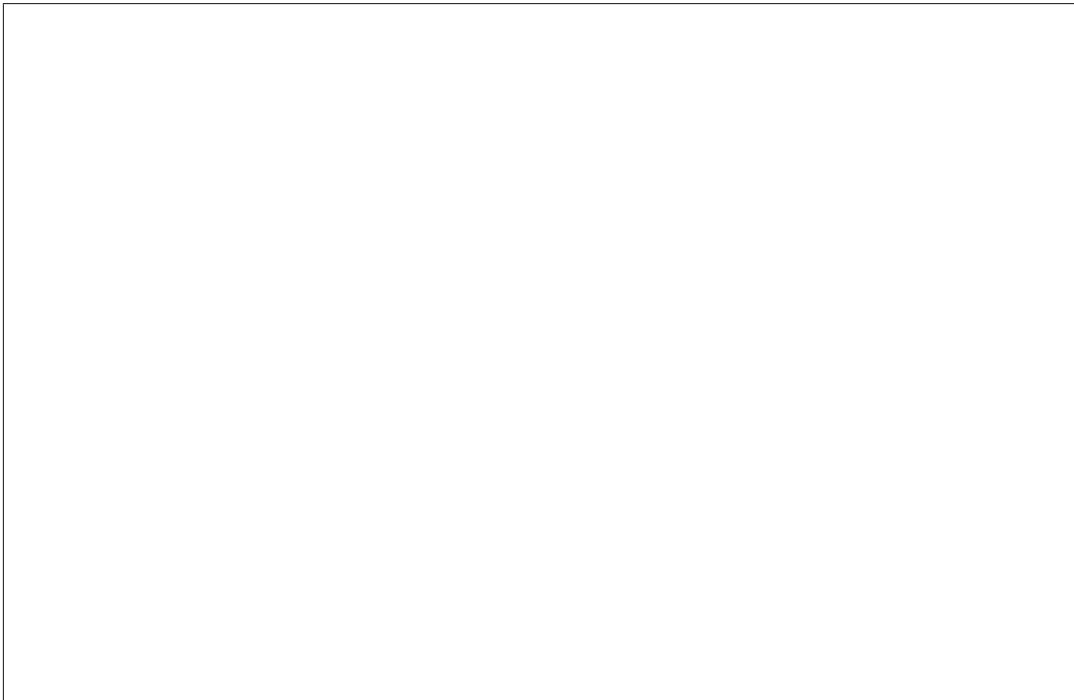


Matlab can be used to make a movie of the 2-dimensional surface by plotting cross-sections at different t values in sequence. The sequence of plots can be stored in a vector and played as a movie using the following code:

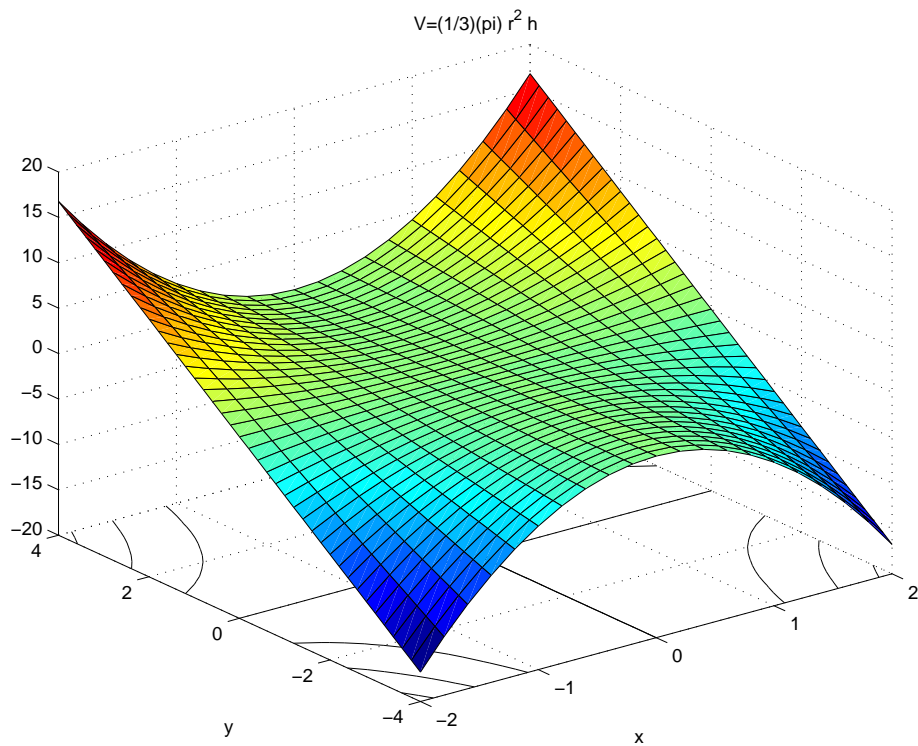
```
x=(0:0.25:1);  
for j=1:100  
    t=j/25;  
    z=sin(pi*x)*cos(2*pi*t);  
    plot(x,z);axis([0,1,-1,1]);  
    M(j)=getframe;  
end
```

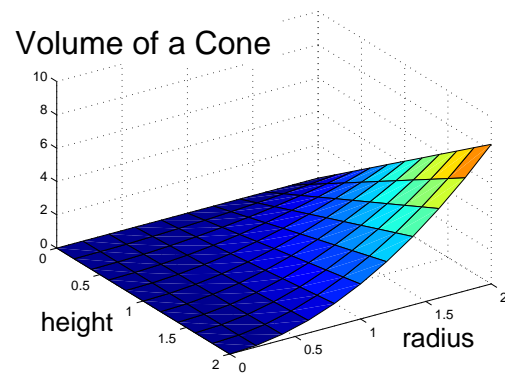
Note: **ezplot** cannot be used to do this because Matlab gets confused about which of t , x is a variable and which is a number.

Example: Use cross-sections to sketch the surface $V = \frac{1}{3}\pi r^2 h$ (volume of a cone).



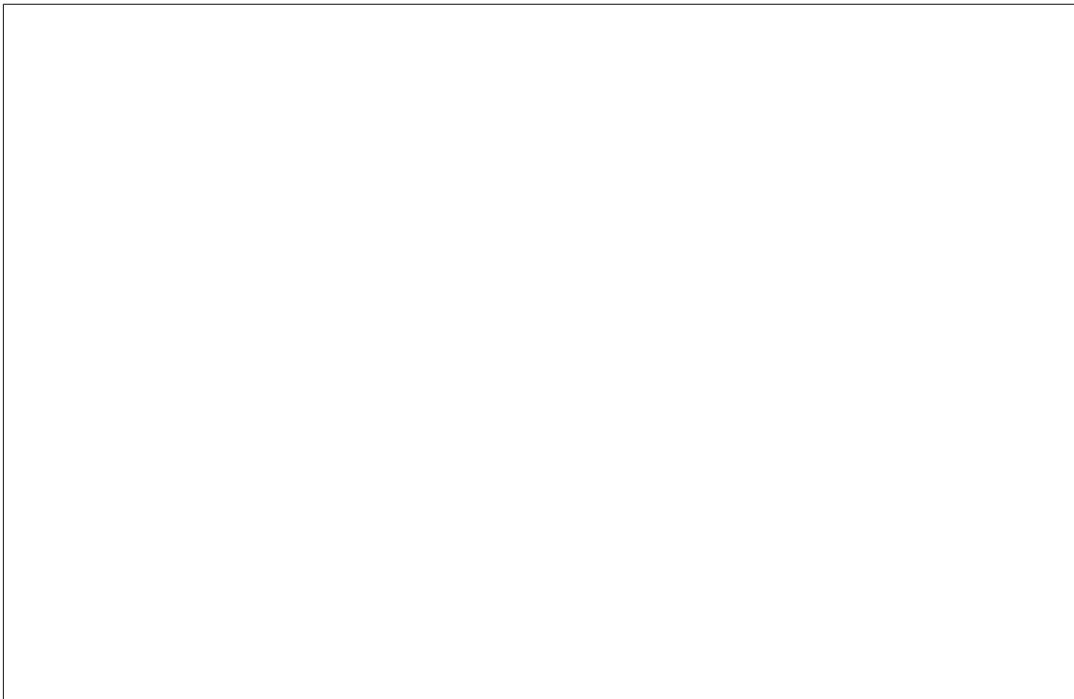
Here is the Matlab version (where $x = r$ and $y = h$).



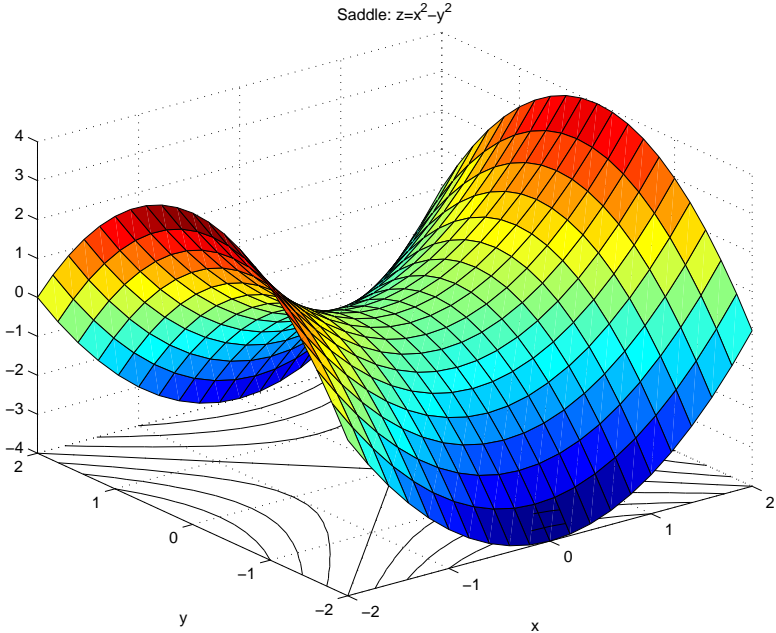


Example: The cross-sections of a saddle $z = x^2 - y^2$ are parabolas. For $y = y_0$ they point up: $z = x^2 - (y_0)^2$; and for $x = x_0$ they point down: $z = -y^2 + x_0^2$.

The surface is tricky to draw, unless you are an equestrian.



Here is the Matlab plot of the saddle $z = x^2 - y^2$.



Example: Use cross-sections to sketch the graph of $z = f(x) = x^2$.



1.7.1 Main points

- You should be able to construct cross-sections of multivariate functions.
- Cross sections are 2-dimensional graphs.
- Animation of cross-sections is another way to visualise multivariate functions.

2 Partial Derivatives and Tangent Planes

We will need to consider derivatives of functions of more than one variable. To do this, we first check how the familiar concepts of **limits** and **continuity** extend to functions of more than one variable. This material is covered in Section 15.2 of Stewart.

2.1 Limits and Continuity

2.1.1 Review of the 1-variable case

Let $f : D \rightarrow \mathbb{R}$ be a function with domain D an open subset of \mathbb{R} . For $a \in D$ we say that **the limit** $\lim_{x \rightarrow a} f(x)$ exists if and only if, (i) the limit from the left exists, (ii) the limit from the right exists, and (iii) these two limits coincide, i.e.,

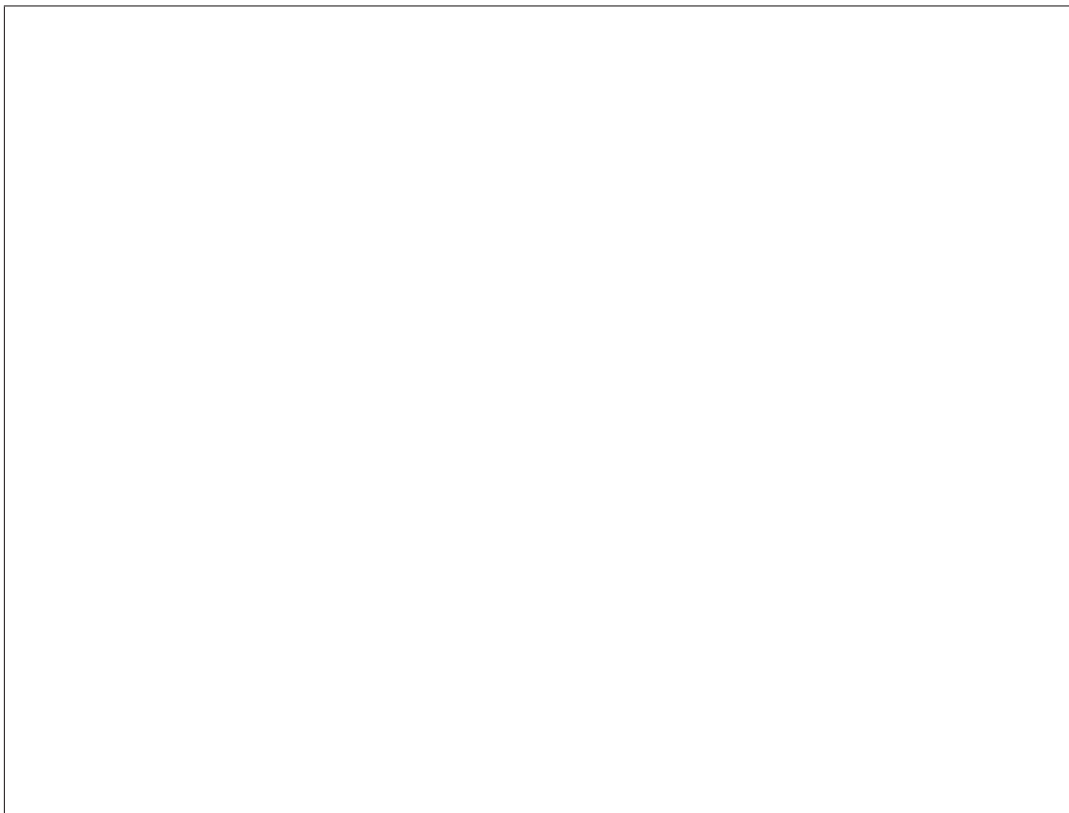
$$\lim_{x \rightarrow a^-} f(x) = \lim_{x \rightarrow a^+} f(x).$$

Furthermore, if the limit exists and is equal the actual value of f at a , i.e., if

$$\lim_{x \rightarrow a^-} f(x) = \lim_{x \rightarrow a^+} f(x) = f(a),$$

we say that f is **continuous** at $x = a$.

If f is continuous on all of D we say that f is a continuous function on D .



We can also consider the limit for points on the boundary of the domain D of a function. For example, if $f : (0, 2) \rightarrow \mathbb{R}$ is defined by $f(x) = 1/x$, then

$$\lim_{x \rightarrow 2^-} f(x) = \frac{1}{2}$$

but

$$\lim_{x \rightarrow 0^+} f(x)$$

does not exist.

Another instructive example is $f : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ given by $f(x) = x^2$. The domain is now the punctured real line, i.e., $D = (-\infty, 0) \cup (0, \infty)$, but

$$\lim_{x \rightarrow 0^-} f(x) = \lim_{x \rightarrow 0^+} f(x) = 0.$$

In this situation we also say that $\lim_{x \rightarrow 0} f(x)$ exists and in fact one can *fix the hole* by defining $f(0) = 0$, to extend f to a continuous function on all of \mathbb{R} .

Important remark: Never, ever compute $\lim_{x \rightarrow a} f(x)$ by blindly substituting $x = a$ in f . For example, if

$$f(x) = \begin{cases} x^2 & \text{for } x \neq 3 \\ \pi & \text{for } x = 3, \end{cases}$$

then $\lim_{x \rightarrow 3} f(x)$ exists and is given by 9 which is not equal to π : the function f is not continuous at $x = 3$.

As a second example, if $f : D \rightarrow \mathbb{R}$ with $D = \mathbb{R} \setminus \{1\}$ is given by

$$f(x) = \frac{x^2 - 1}{x - 1},$$

then $\lim_{x \rightarrow 1} f(x) = 2$. Those (and there will be some) who write “This limit gives 0/0 which does not exist” should hang their heads in shame.

2.1.2 Multivariable limits

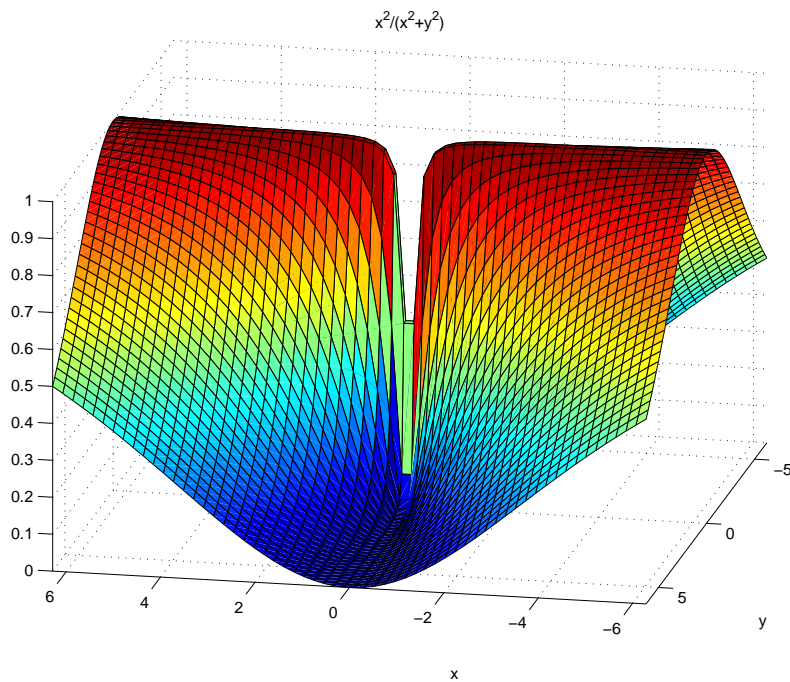
When f is a function of more than one variable, the situation is more interesting. There are more than two ways to approach a given point of interest. Consider the function

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

with domain given by $\mathbb{R}^2 \setminus \{(0, 0)\}$.

To see the graph of f in Matlab, type

```
ezsurf('x^2/(x^2+y^2)')
```



Next we consider the limit as $(x, y) \rightarrow (0, 0)$.

(i) Approaching the origin along $y = 0$:

(ii) Approaching the origin along $x = 0$:

Does $\lim_{(x,y) \rightarrow (0,0)} f(x,y)$ exist?

In general, for the limit $\lim_{(x,y) \rightarrow (a,b)} f(x,y)$ to exist, it is necessary that every path in D approaching (a,b) (the point (a,b) itself may or may not be in D) gives the same limiting value. This gives us the following method for finding if a limit does not exist.

Test for showing no limit exists

If $\begin{cases} f(x,y) \rightarrow L_1 & \text{as } (x,y) \rightarrow (a,b) \text{ along the path } C_1 \in D \\ f(x,y) \rightarrow L_2 & \text{as } (x,y) \rightarrow (a,b) \text{ along the path } C_2 \in D \end{cases}$
such that $L_1 \neq L_2$, then the limit $\lim_{(x,y) \rightarrow (a,b)} f(x,y)$ does not exist.

Important remark: The above notation is somewhat deficient and perhaps one should write

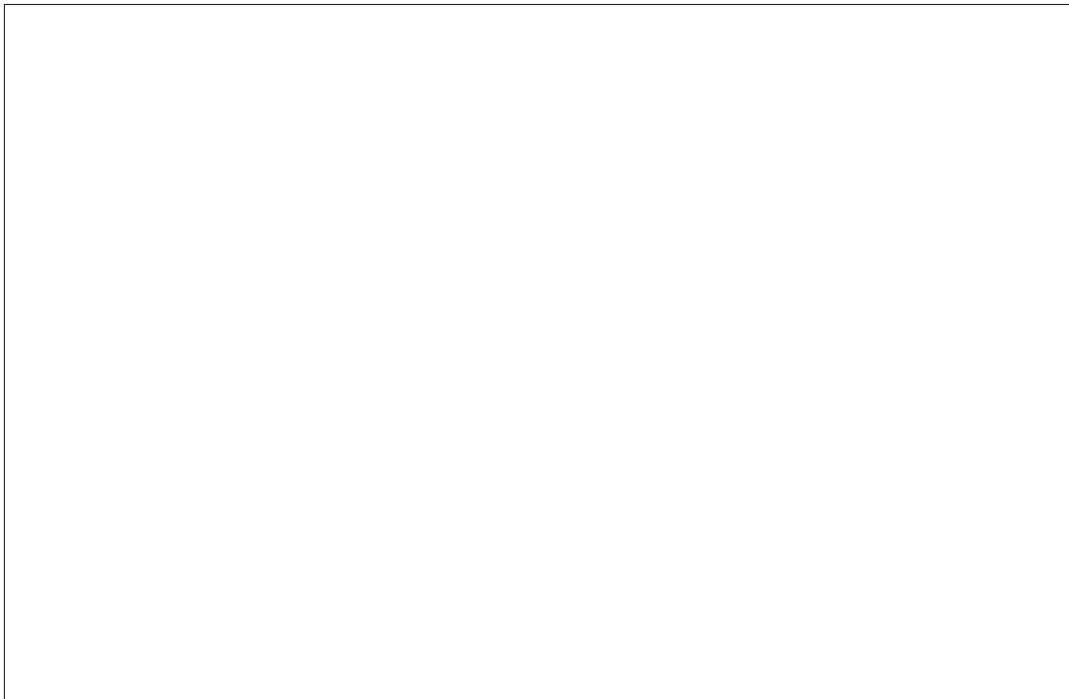
$$\lim_{(x,y) \rightarrow_D (a,b)} f(x,y)$$

to indicate that only paths in D terminating in (a,b) (which itself may or may not be in D) are considered. For example, if $f(x,y) = x^2 + y^2$ with $D = \{(x,y) : x^2 + y^2 < 1\}$ then $\lim_{(x,y) \rightarrow (1,0)} f(x,y)$ exists and is 1. However, if

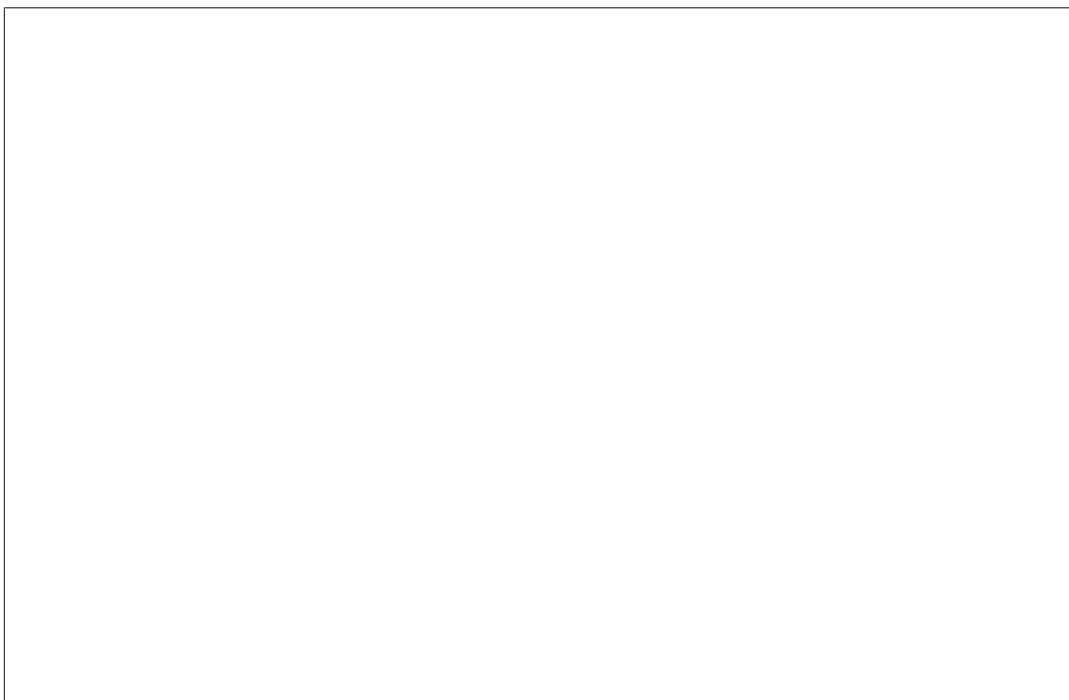
$$f(x,y) = \begin{cases} x^2 + y^2 & \text{for } D = \{(x,y) : x^2 + y^2 < 1\} \\ 0 & \text{for } D = \{(x,y) : x^2 + y^2 > 1\} \end{cases}$$

then $\lim_{(x,y) \rightarrow (1,0)} f(x,y)$ does not exist.

Example: Let $D = \mathbb{R}^2 \setminus \{(0, 0)\}$ and $f : D \rightarrow \mathbb{R}^2$ be given by $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$. Show that $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ does not exist.



Example: With the same D as above but now $f(x, y) = \frac{xy}{x^2 + y^2}$, show that $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ does not exist.



Important remark: There are infinitely many paths terminating in a given point, say (a, b) , in \mathbb{R}^2 , raising the question if one can ever prove that $\lim_{(x,y) \rightarrow (a,b)} f(x, y)$ does exist. The good news is that there are methods that can deal with infinitely many paths simultaneously. The bad news is that these methods (typically ϵ - δ proofs) are not part of this course. See Stewart Sec 15.2, Example 4 for a rigorous ϵ - δ proof that

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0,$$

where $f(x, y) = 3x^2y/(x^2 + y^2)$ and $D = \mathbb{R}^2 - \{(0, 0)\}$.

Exercise: Give a non-rigorous proof that the above limit is indeed correct by writing $x = r \cos \theta$ and $y = r \sin \theta$.



2.1.3 Multivariable continuity

Definition Given a function $f : D \rightarrow \mathbb{R}^2$, where D is an open subset of \mathbb{R}^2 . Let $(a, b) \in D$. Then $f(x, y)$ is **continuous** at (a, b) if

$$\lim_{(x,y) \rightarrow (a,b)} f(x, y) = f(a, b),$$

i.e., the limit $(x, y) \rightarrow (a, b)$ of $f(x, y)$ exists and is equal to $f(a, b)$.

If a function is continuous on all of D we say simply that it is **continuous** on D . Most of the functions we will consider are continuous. For example, polynomials in x and y are continuous on \mathbb{R}^2 . As a rule of thumb, if a function with domain D is defined by a single expression it will be continuous on D .

Example: Returning to our first example on page 69, $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$ is a continuous function despite the fact that $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ does not exist. The reason the non-existence of the limit does not affect continuity is that $(0, 0) \notin D$. However, if we had defined f on all of \mathbb{R}^2 by taking $f(0, 0) = 0$, then, all of a sudden f would no longer be a continuous function: f would be continuous on all of D except for the point $(0, 0)$.

2.1.4 Main points

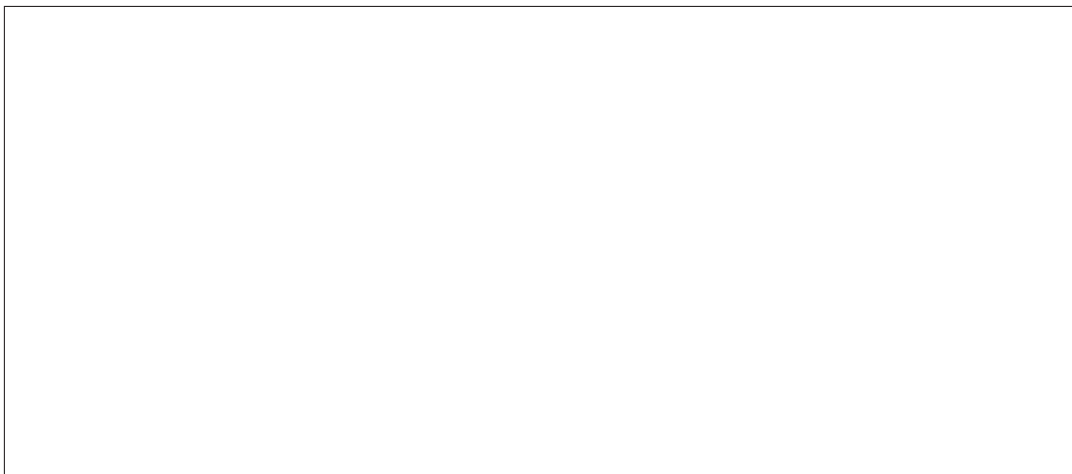
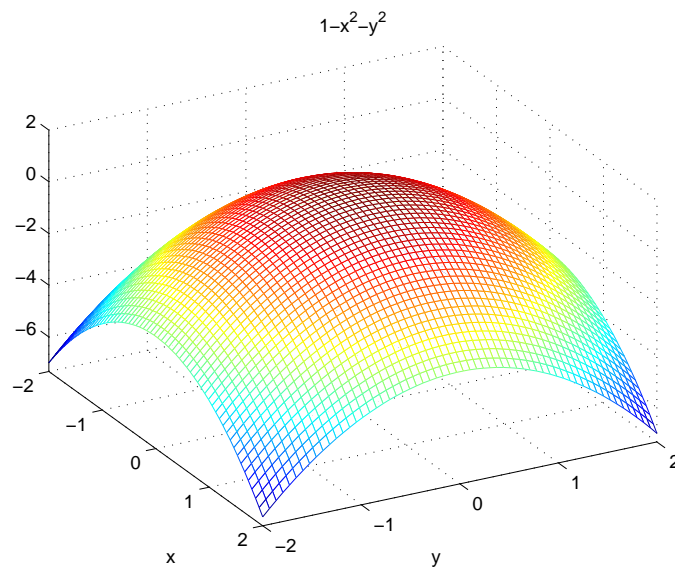
- You should be able to show when a limit does not exist.
- You should understand continuity of multivariate functions.

2.2 Partial Derivatives

This material is covered in Stewart, Section 15.3.

2.2.1 Slope in the x -direction

Consider the surface $z = f(x, y) = 1 - x^2 - y^2$ and the point $P = (1, -1, -1)$ on the surface. Draw the “ y -is-constant” cross-section through P and find the slope in the x -direction at P .

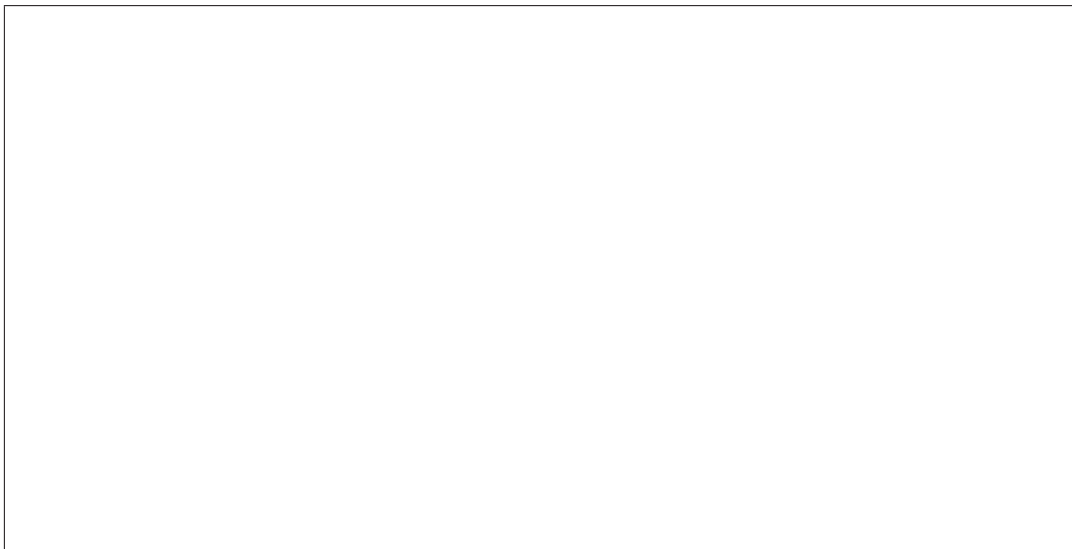
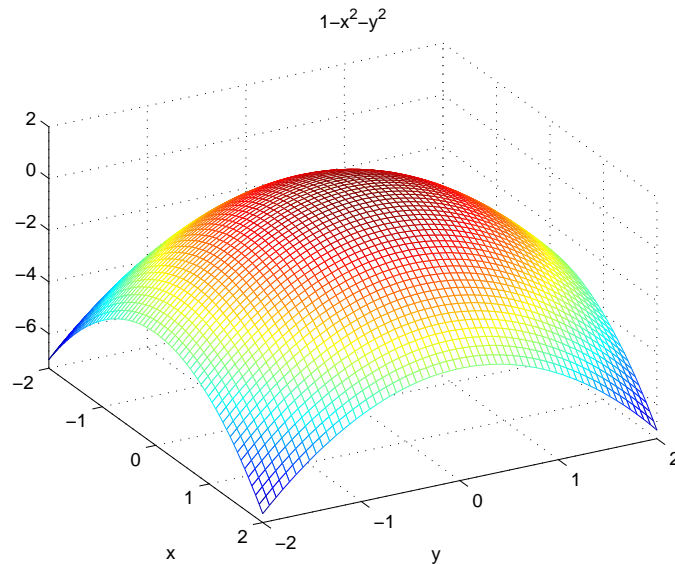


The slope in the x -direction, with y held fixed, is called the **partial derivative** of f with respect to x at the point (a, b)

$$\frac{\partial f}{\partial x}(a, b) = f_x(a, b) = \lim_{h \rightarrow 0} \frac{f(a + h, b) - f(a, b)}{h}.$$

2.2.2 Slope in the y -direction

Draw the “ x -is-constant” cross-section and find the slope at $P = (1, -1, -1)$ in the y direction, i.e., where $x = 1$.

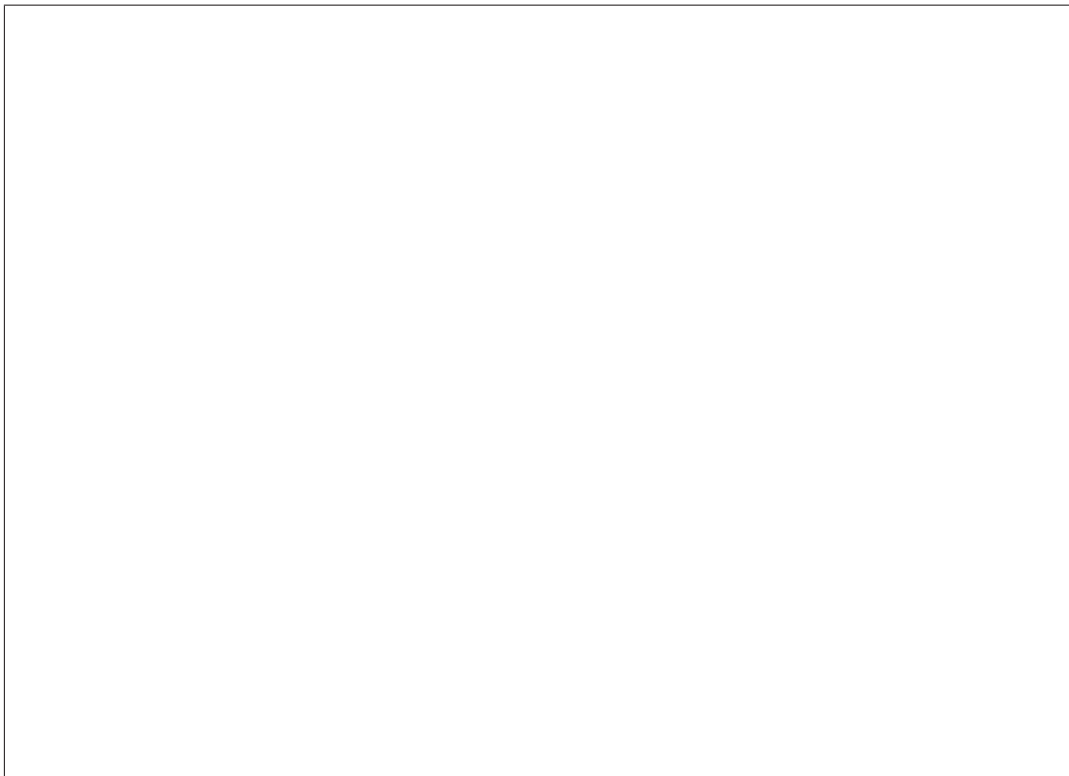


Similarly, the slope in the y -direction, with x held fixed, is called the **partial derivative** of f with respect to y at the point (a, b)

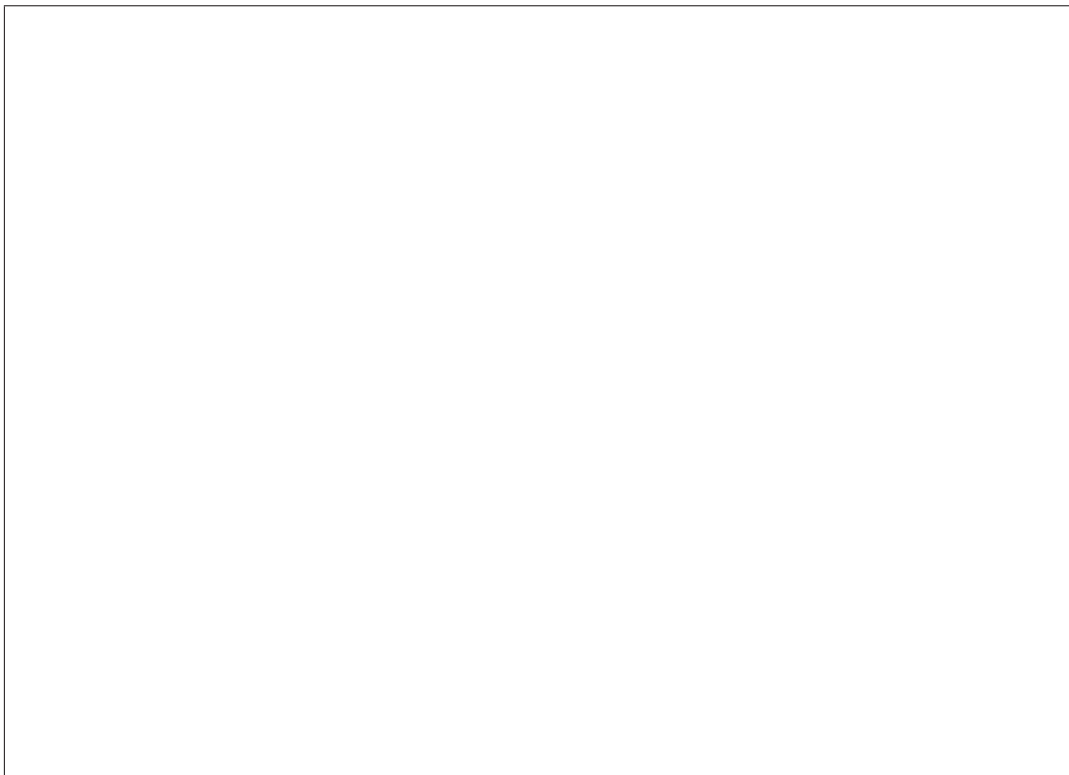
$$\frac{\partial f}{\partial y}(a, b) = f_y(a, b) = \lim_{h \rightarrow 0} \frac{f(a, b + h) - f(a, b)}{h}.$$

Normal rules of differentiation apply, we simply think of x as a constant.

Example: Find the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ of $f(x, y) = x \sin y + y \cos x$.



Example: Given $f(x, y) = xy^3 + x^2$, find $f_x(1, 2)$ and $f_y(1, 2)$.



2.2.3 Partial derivatives for $f(x, y, z)$

Example: The volume of a box $V(x, y, z) = xyz$.

If x changes by a small amount, say Δx , denote the corresponding change in V by ΔV . We can easily visualise that $\Delta V = yz\Delta x$. Therefore,

$$\frac{\Delta V}{\Delta x} = yz.$$

Letting $\Delta x \rightarrow 0$ we have $\frac{\partial V}{\partial x} = yz$.

For partial derivatives only one independent variable changes and all other independent variables remain fixed.

Example: Parallel resistance

In an electrical circuit, the combined resistance R , from three resistors R_1, R_2 and R_3 connected in parallel, is

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}.$$

What is the rate of change of the total resistance R with respect to R_1 ?

2.2.4 Higher order derivatives

The second partial derivatives of f , if they exist, are written as

$$\begin{aligned}f_{xx} &= \frac{\partial^2 f}{\partial x^2}, & f_{xy} &= \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right), \\f_{yy} &= \frac{\partial^2 f}{\partial y^2}, & f_{yx} &= \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right).\end{aligned}$$

If f_{xy} and f_{yx} are both continuous, then $f_{xy} = f_{yx}$.

2.2.5 Main points

- You should know the definition and meaning of partial derivatives
- You should be able to evaluate partial derivatives of functions.

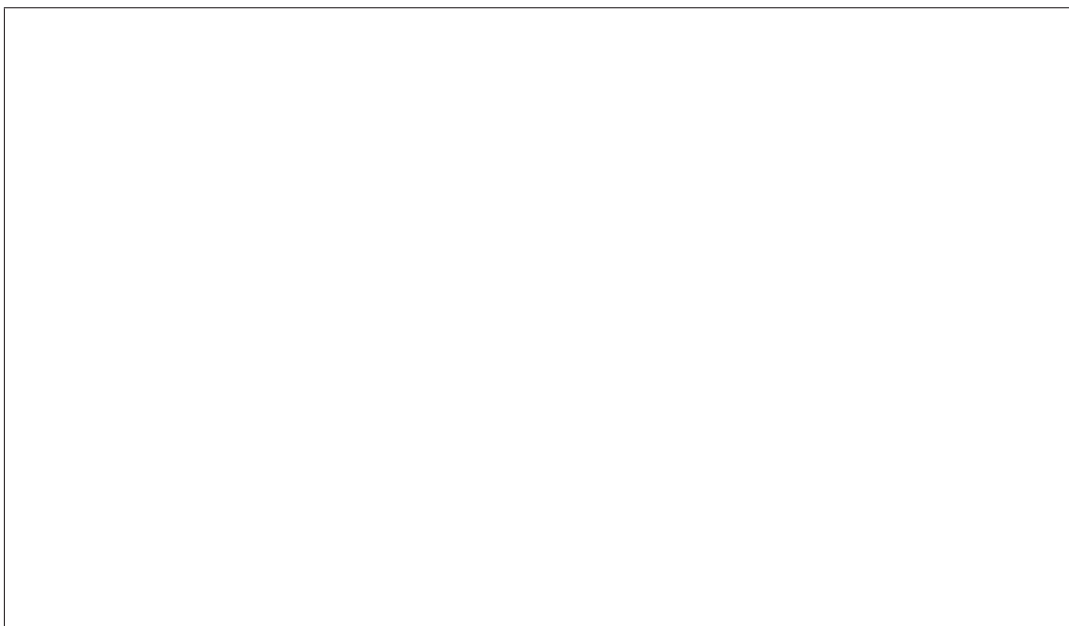
2.3 The Tangent Plane

This section is covered in Stewart, Section 15.4.

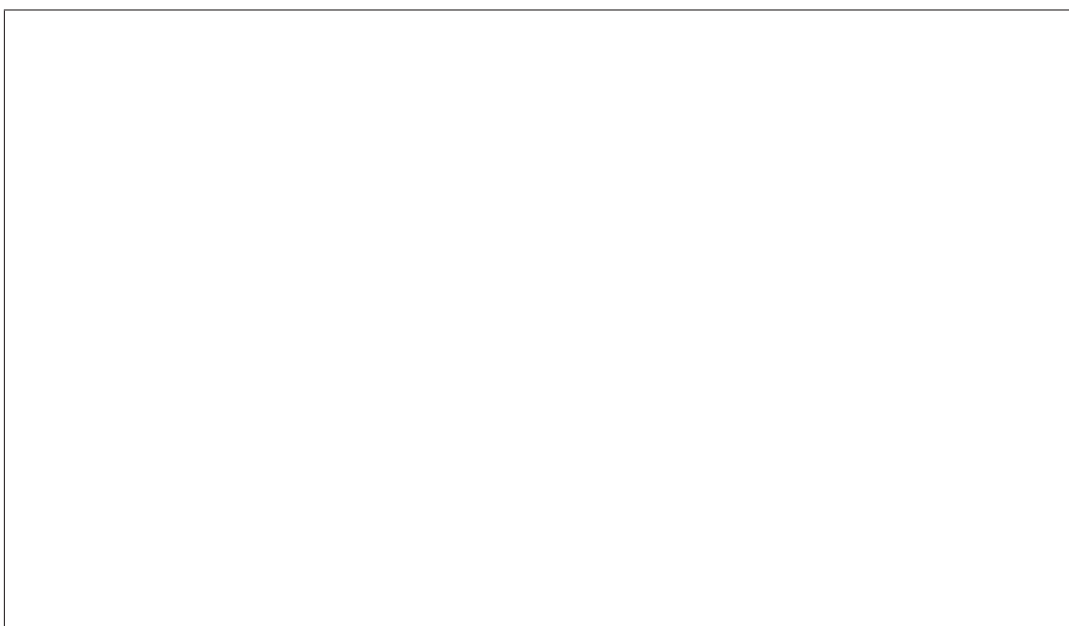
2.3.1 Review for $f(x)$

Recall that if $y = f(x)$ then the tangent line at the point $(a, f(a))$ is given by

$$y - f(a) = f'(a)(x - a) \quad \Leftrightarrow \quad y = f(a) + f'(a)(x - a).$$



Example: Find the tangent line to $y = f(x) = x^2$ at $x = 1$.



2.3.2 Equation for a tangent plane

In general, the equation of the tangent plane to a given surface $z = f(x, y)$ at $(a, b, f(a, b))$, is

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b),$$

or, equivalently,

$$(x, y, z) = (a, b, f(a, b)) + \lambda(1, 0, f_x(a, b)) + \mu(0, 1, f_y(a, b)), \quad \lambda, \mu \in \mathbb{R}.$$

Indeed, the first two components of this vector equation for the tangent plane imply $\lambda = x - a$ and $\mu = y - b$. Substituting this in the third component gives

$$z = f(a, b) + \lambda f_x(a, b) + \mu f_y(a, b) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

We also note that if we write $\Delta x = x - a$, $\Delta y = y - b$ and $\Delta z = z - f(a, b)$ then the vector equation for the tangent plane may be rewritten as

$$(\Delta x, \Delta y, \Delta z) = \lambda(1, 0, f_x(a, b)) + \mu(0, 1, f_y(a, b)), \quad \lambda, \mu \in \mathbb{R}.$$

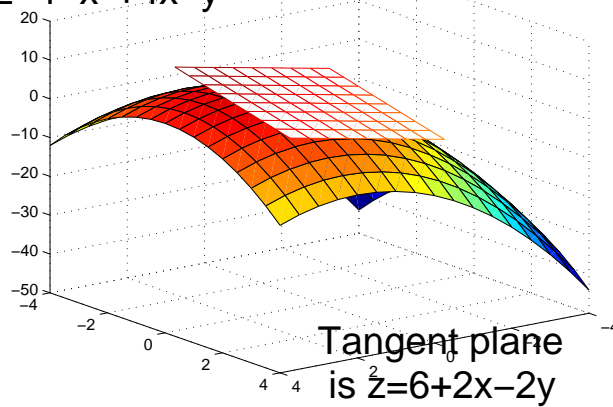
This shows that if $\Delta x = 1$ and $\Delta y = 0$ (i.e., $\lambda = 1$ and $\mu = 0$) then $\Delta z = f_x(a, b)$ and if $\Delta x = 0$ and $\Delta y = 1$ (i.e., $\lambda = 0$ and $\mu = 1$) then $\Delta z = f_y(a, b)$, in accordance with our interpretation of f_x and f_y as the respective slopes of f in x or y direction.

Example: What is the plane tangent to the surface $z = f(x, y) = 4 - x^2 + 4x - y^2$ at $(1, 1)$?



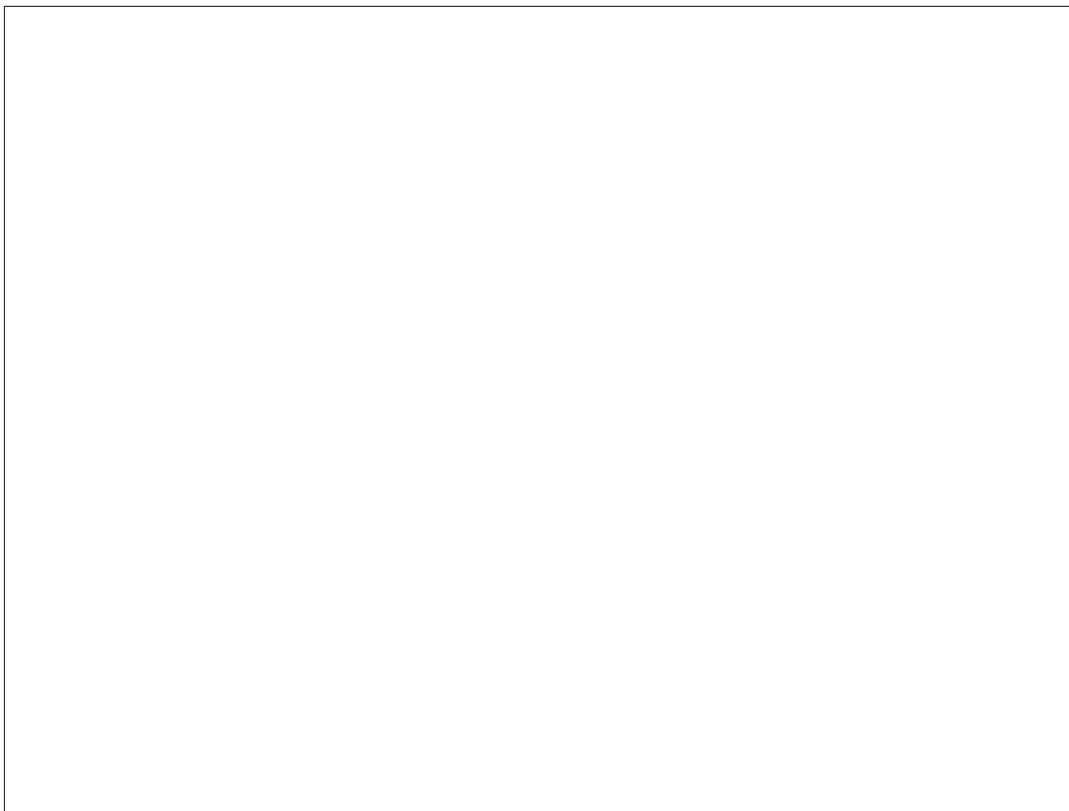
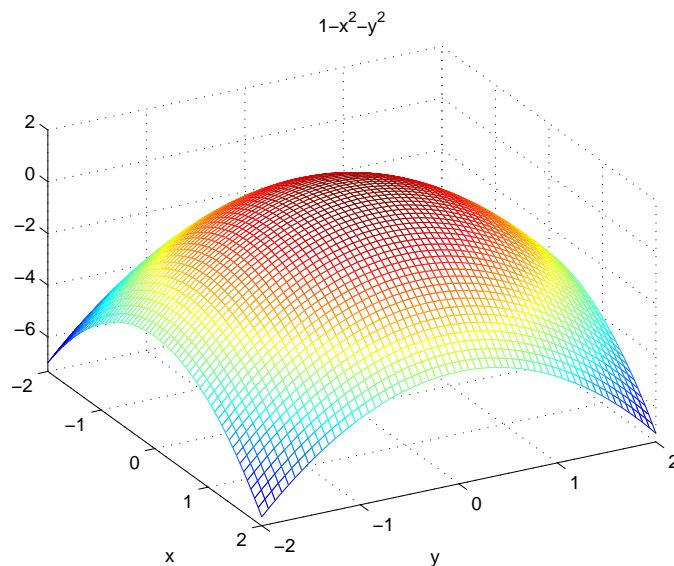
Tangent Plane at $(1, 1, 6)$ to

$$z = 4 - x^2 + 4x - y^2$$



Example: Find the tangent plane of $z = f(x, y) = e^{-x^2 - y^2}$ at $(x, y) = (1, 3)$.

Example: Find the equation for the tangent plane to the surface $z = 1 - x^2 - y^2$ at the point $P = (1, -1, -1)$.

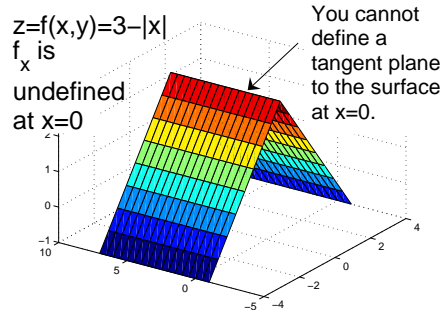


2.3.3 Smoothness

Can we always find partial derivatives and tangent planes?

Example.

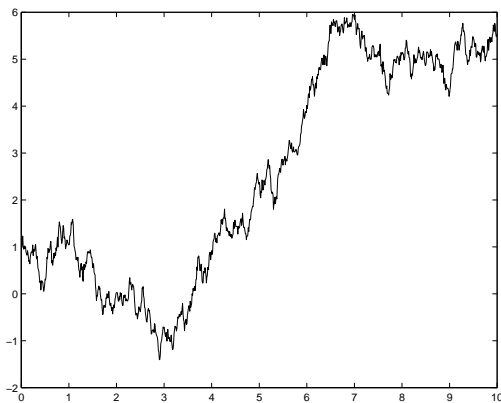
Simple cusp-like functions are not smooth.



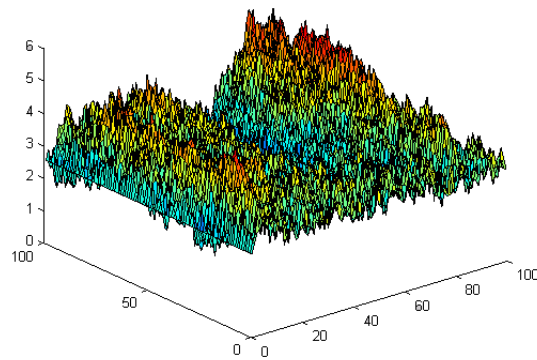
A surface $z = f(x, y)$ is **smooth** at (a, b) if f , f_x and f_y are all continuous at (a, b) . When you zoom in close enough to a smooth surface it looks like a plane.

One way to see this is to look at the contours. The contours of a plane are straight parallel lines, the same perpendicular distance apart. As you zoom into a smooth surface the contours straighten out. This means that close to (a, b) the surface is approximated by a plane; in fact it can be approximated by the tangent plane. Do straight contours imply smoothness?

Example: Brownian motion is not smooth. Look at the figure below. No matter how much you zoom in, it always looks rough—in fact, Brownian motion is a fractal. There are surface-analogues to Brownian motion, demonstrated with the fractal surface below.



Brownian motion



Fractal surface

2.3.4 Main points

- You should know how to find a tangent plane to a smooth surface, and recognise when a tangent plane or partial derivatives do not exist.

2.4 Linear Approximations

See Stewart, Section 15.4.

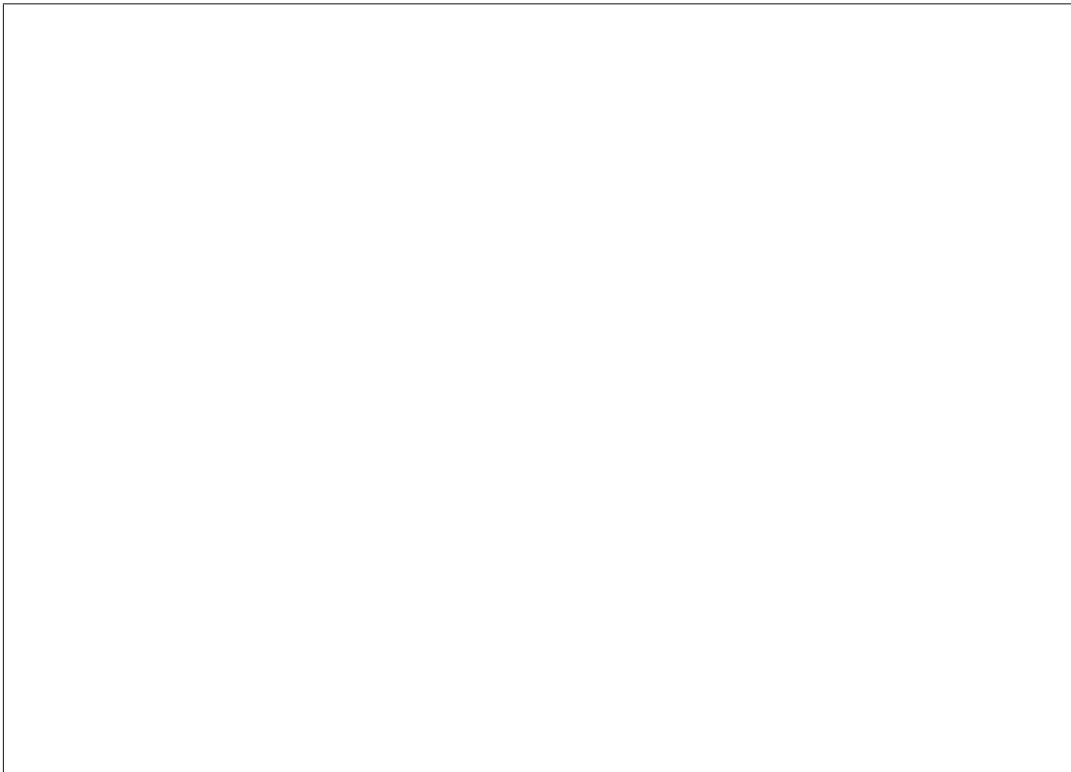
2.4.1 Review for $f(x)$

In science and engineering practice, a function $f(x)$, of one variable, is sometimes approximated by a straight line.

In particular, if we know an exact value $(a, f(a))$, then the tangent line through this point is usually an accurate approximation close to the point. It is called a **linear** or **first-order** approximation

$$f(x) \simeq f(a) + f'(a)(x - a).$$

Example: Estimate $e^{0.1}$.



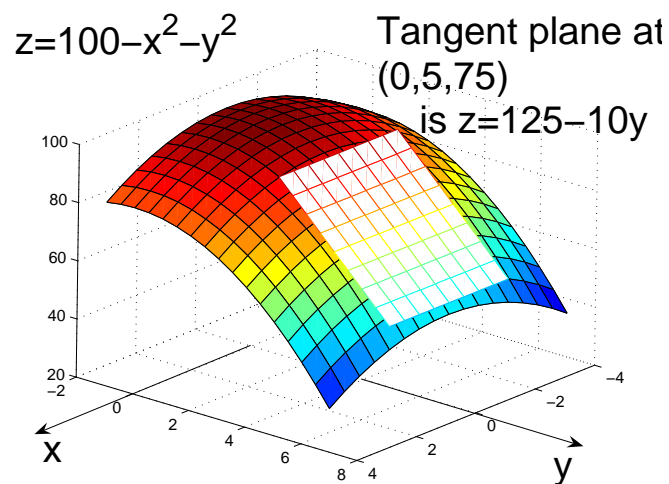
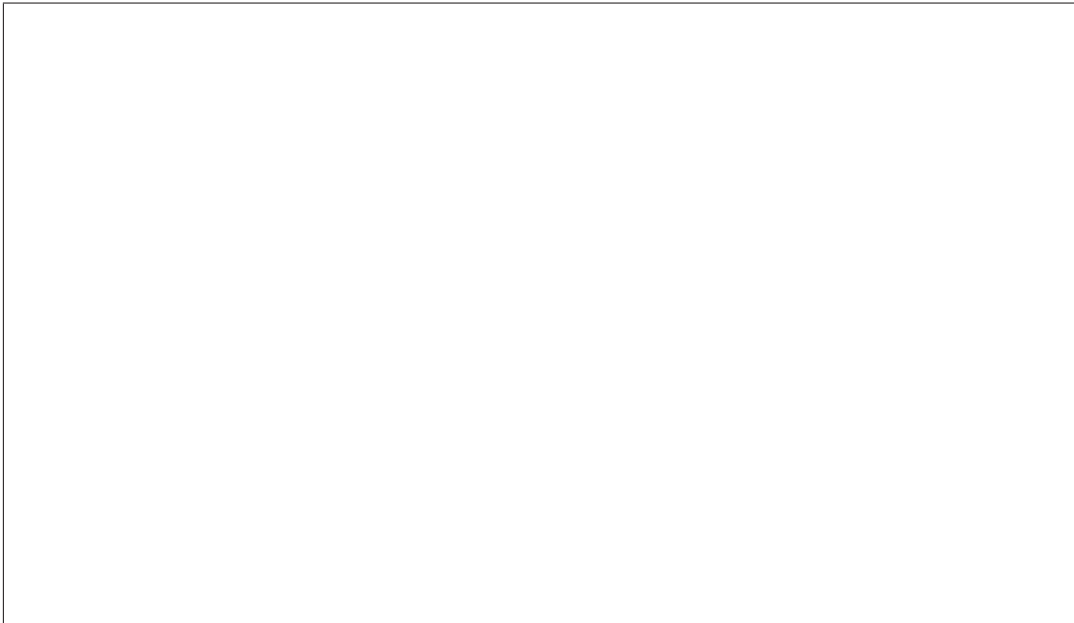
2.4.2 Linear approximations for $f(x, y)$

The corresponding **linear** or **first-order** approximation for a function f of two variables, near a known point (a, b) is the tangent plane. The linear approximation (i.e., tangent plane) is accurate close to the known point, provided f is smooth.

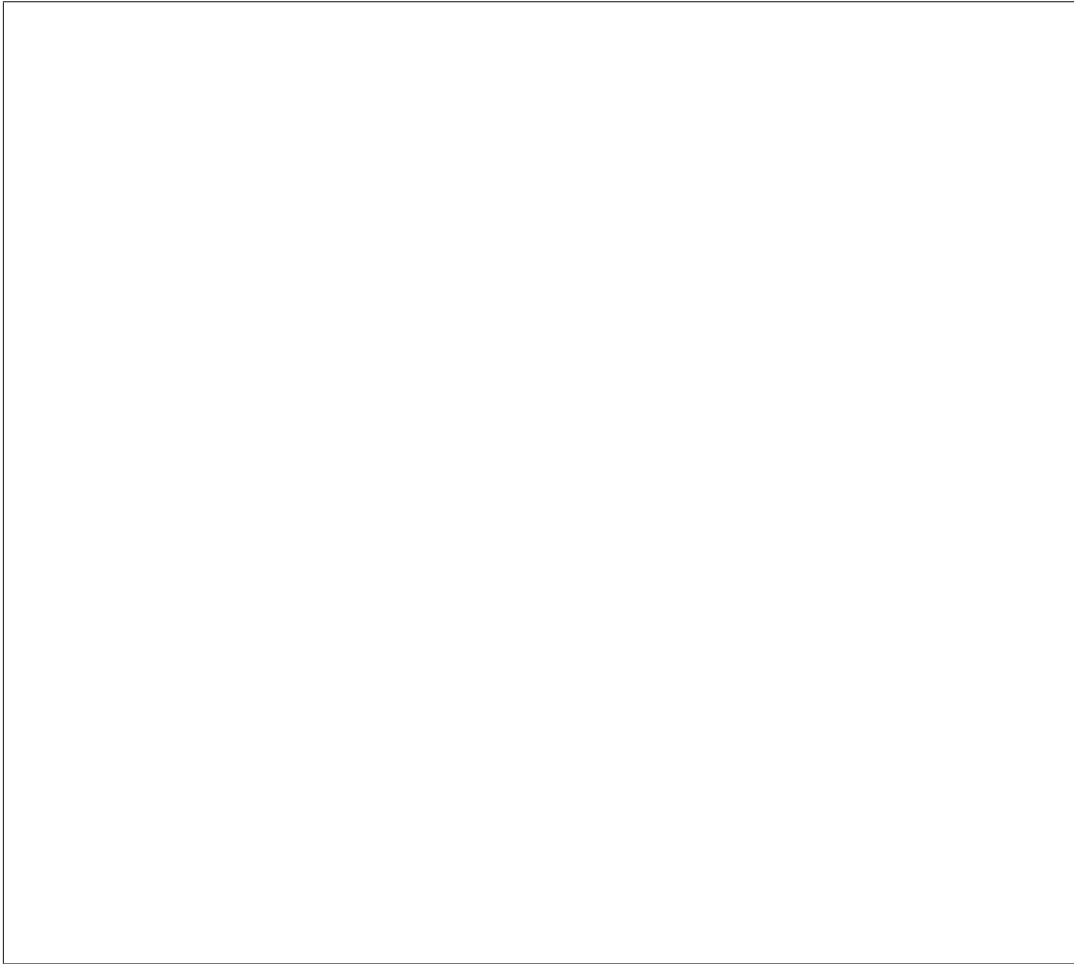
The linear approximation to f at (a, b) is

$$f(x, y) \simeq f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Example: The temperature in a region is given by $T(x, y) = 100 - x^2 - y^2$. Find the linear approximation to $T(x, y)$ near $(0, 5)$.

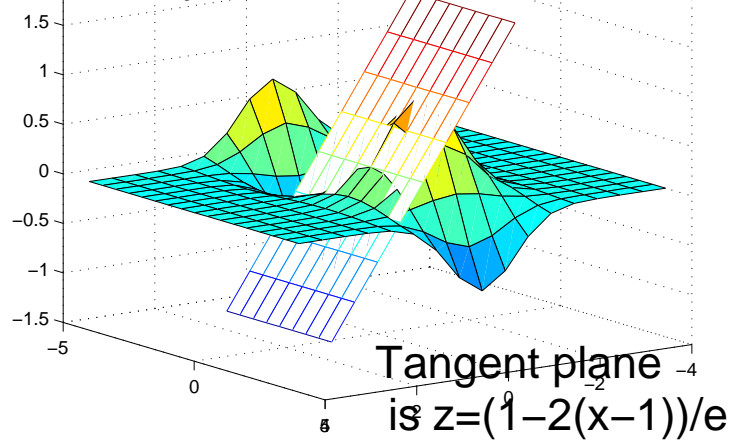


Example: Find the tangent plane to $z = e^{-x^2} \sin y$ at $(1, \frac{\pi}{2})$ and use it to find an approximate value for $e^{-(0.9)^2} \sin(1.5)$.



Tangent Plane at $(1, \pi/2, 1/e)$ to

$$z = e^{-x^2} \sin y$$



2.4.3 Estimating small changes

We may use the equation for the tangent plane to infer that

$$\Delta z \approx f_x(a, b)\Delta x + f_y(a, b)\Delta y,$$

where $\Delta x = x - a$, $\Delta y = y - b$ and $\Delta z = z - f(a, b)$ represent a small change in x , y and z respectively. This is useful for estimating small changes in z arising from small changes in x and y .

Example: Electric power is given by $P(E, R) = E^2/R$ where E is the voltage and R is the resistance. Find a linear approximation for $P(E, R)$ if $E \simeq 200$ (in Volts) and $R \simeq 400$ (in Ohms). Use this to find the effect that a change in E and R has on P .



2.4.4 Estimating error

If the error in x is at most E_1 and in y is at most E_2 , then a reasonable estimate of the worst-case error in the linear approximation of f in (a, b) is

$$|E| \approx |f_x(a, b)E_1| + |f_y(a, b)E_2|.$$

Exercise: Suppose when making up a metal barrel of base radius 1 (in metres) and height 2 (in metres), you allow for an error of 5% in radius and height. Estimate the worst-case error in volume.

2.4.5 Extra reading: differentials

If we reduce our small changes Δx and Δy down to **infinitesimal** changes dx and dy , we may rewrite the linear approximation (tangent plane) as

$$dz = df = f_x(a, b) dx + f_y(a, b) dy.$$

This infinitesimal change dz is called the **total differential**. Note, this equation is not just an approximation.

Example: Work done by a Gas

If the work done by an expanding gas, is given by $W = PV$, where P denotes pressure, and V volume, write an expression for the total differential dW .

2.4.6 Main points

- You should be able to work out total differentials for a given function of several variables.
- You should be able to calculate finite small deviations of the main function, given a small change in one of the variables. Note that this is a computationally more efficient method of approximation than simply evaluating the function at x and $x + \Delta x$, and then taking the difference.
- You should be able to estimate error using the linear approximation.

2.5 Gradients and Directional Derivatives

This material is covered in Stewart, Section 15.6.

2.5.1 Directional derivatives

The partial derivative f_x (or f_y) corresponds to the slope of $f(x, y)$ in the x -direction (or y -direction). We now turn our attention to the question of slopes in arbitrary directions, such as $\mathbf{i} + 2\mathbf{j}$ or $-\mathbf{j}$.

Let $\mathbf{u} = (u_1, u_2)$ be an arbitrary **unit vector** in \mathbb{R}^2 , i.e., $\|\mathbf{u}\| = 1$. We have already seen that the equation of the tangent plane in (a, b) of a function f may be written as

$$(\Delta x, \Delta y, \Delta z) = \lambda(1, 0, f_x(a, b)) + \mu(0, 1, f_y(a, b)), \quad \lambda, \mu \in \mathbb{R}.$$

where $\Delta x = x - a$, $\Delta y = y - a$ and $\Delta z = z - f(a, b)$. This allows us to easily describe the slope of f at the point (a, b) in the direction of \mathbf{u} : we must “measure” Δz when $\Delta x = u_1$ and $\Delta y = u_2$. Hence we take $\lambda = u_1$ and $\mu = u_2$ in order to find that

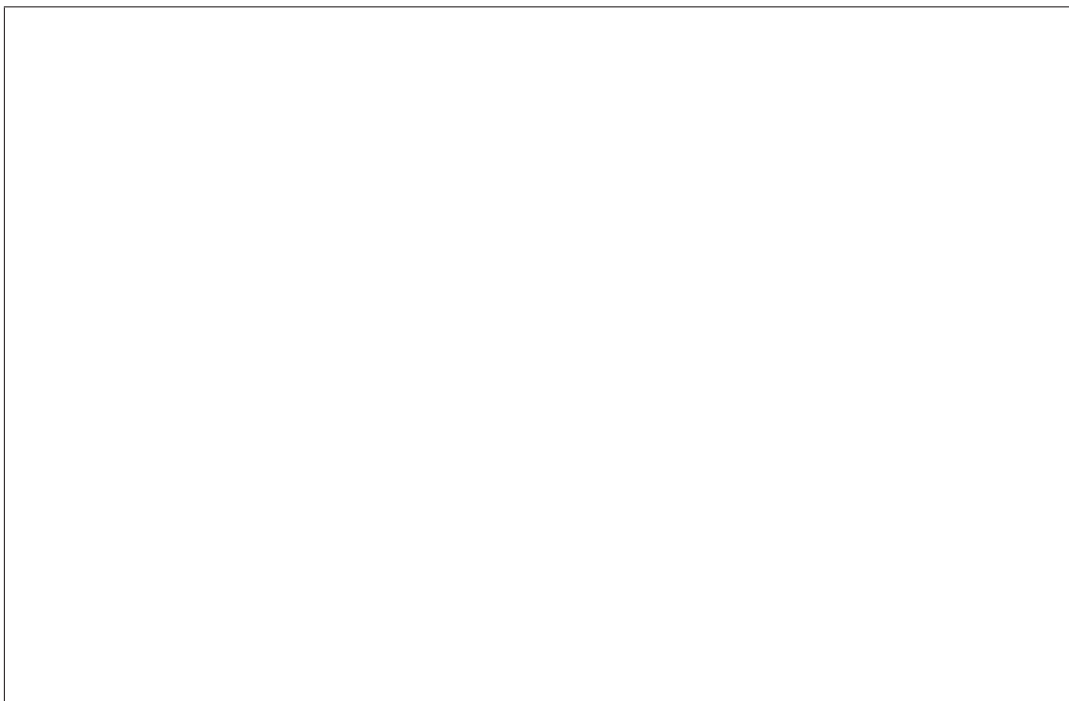
$$\begin{aligned} \Delta z &= u_1 f_x(a, b) + u_2 f_y(a, b) \\ &= (f_x(a, b), f_y(a, b)) \cdot \mathbf{u}. \end{aligned}$$

This is usually denoted as $f_{\mathbf{u}}(a, b)$ and known as the **slope** of f at the point (a, b) in the direction of \mathbf{u} , or the **directional derivative** of f at (a, b) in the direction of \mathbf{u} .

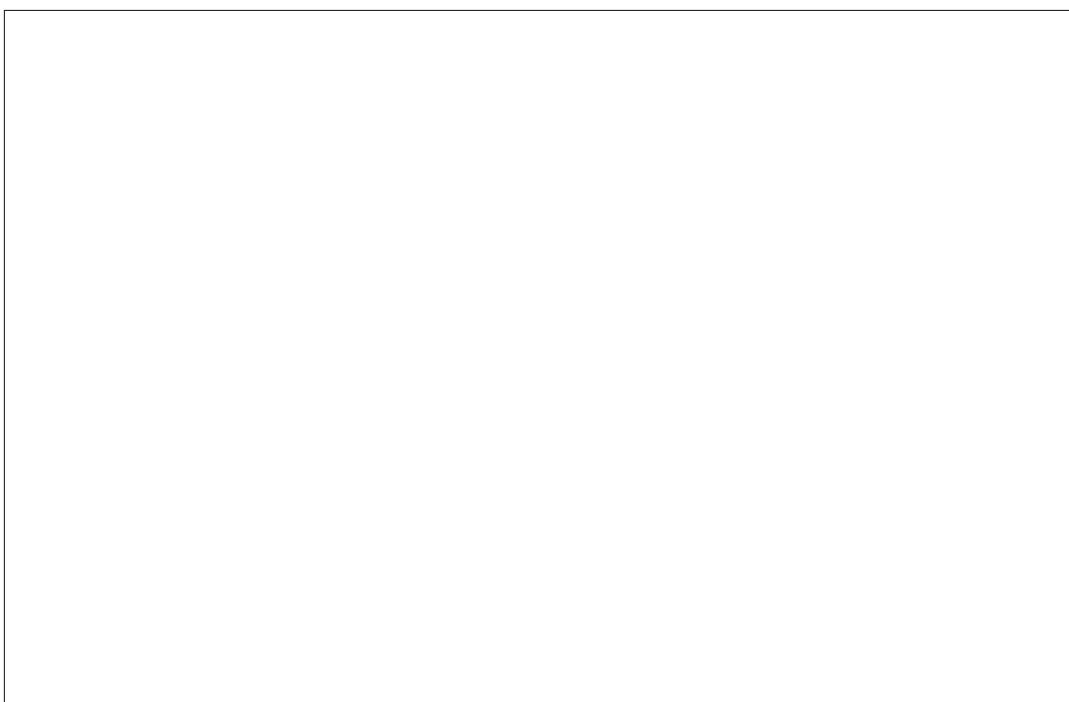
Because it is not always convenient to work with unit vectors, we more generally have that the directional derivative of f at (a, b) in the direction of an arbitrary nonzero vector \mathbf{u} is given by

$$f_{\mathbf{u}}(a, b) = (f_x(a, b), f_y(a, b)) \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Example: Find the directional derivative of $f(x, y) = 4 - x^2 - 4y^2$ at (a, b) in the $(1, 1)$ direction.



Example: If $f(x, y) = x^2 - 3y^2 + 6y$, find the slope at $(1, 0)$ in the direction $\mathbf{i} - 4\mathbf{j}$.



2.5.2 The gradient vector ∇f

The **gradient vector** or simply **gradient** of f is a vector with the partial derivatives as components

$$\text{grad } f = \nabla f = (f_x, f_y) = f_x \mathbf{i} + f_y \mathbf{j}.$$

Example: Find the gradient of $f(x, y) = x^2 - 3(y - 1)^2 + 3$.

Similarly, for a function $f(x, y, z, w)$ of four variables,

$$\text{grad } f = \nabla f = (f_x, f_y, f_z, f_w).$$

Note that directional derivative can be conveniently expressed in terms of the gradient as

$$f_{\mathbf{u}} = \nabla f \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|}.$$

Example: Find $f_{(1,-1)}(0, 1)$.

Example: Find the directional derivative of $g(x, y) = e^{x^2} \cos y$ at $(1, \pi)$ in the direction $-3\mathbf{i} + 4\mathbf{j}$.



2.5.3 Properties of the gradient vector

Consider the contour diagram of a plane $z = f(x, y) = mx + ny + c$. For $n \neq 0$ the contours have slope $-m/n$ in the xy -plane.



The gradient vector, $m\mathbf{i} + n\mathbf{j}$ ($m \neq 0$), has slope n/m and so is perpendicular to the contours. It points in the direction of increasing f . In fact, the direction in which it points is the direction of greatest slope.

Consider $f(x, y) = x^2 + y^2$. The contours are circles centered at the origin, and $\nabla f = 2x\mathbf{i} + 2y\mathbf{j}$ points radially out. Once again, ∇f points in the direction of greatest slope, perpendicular to the contour lines.

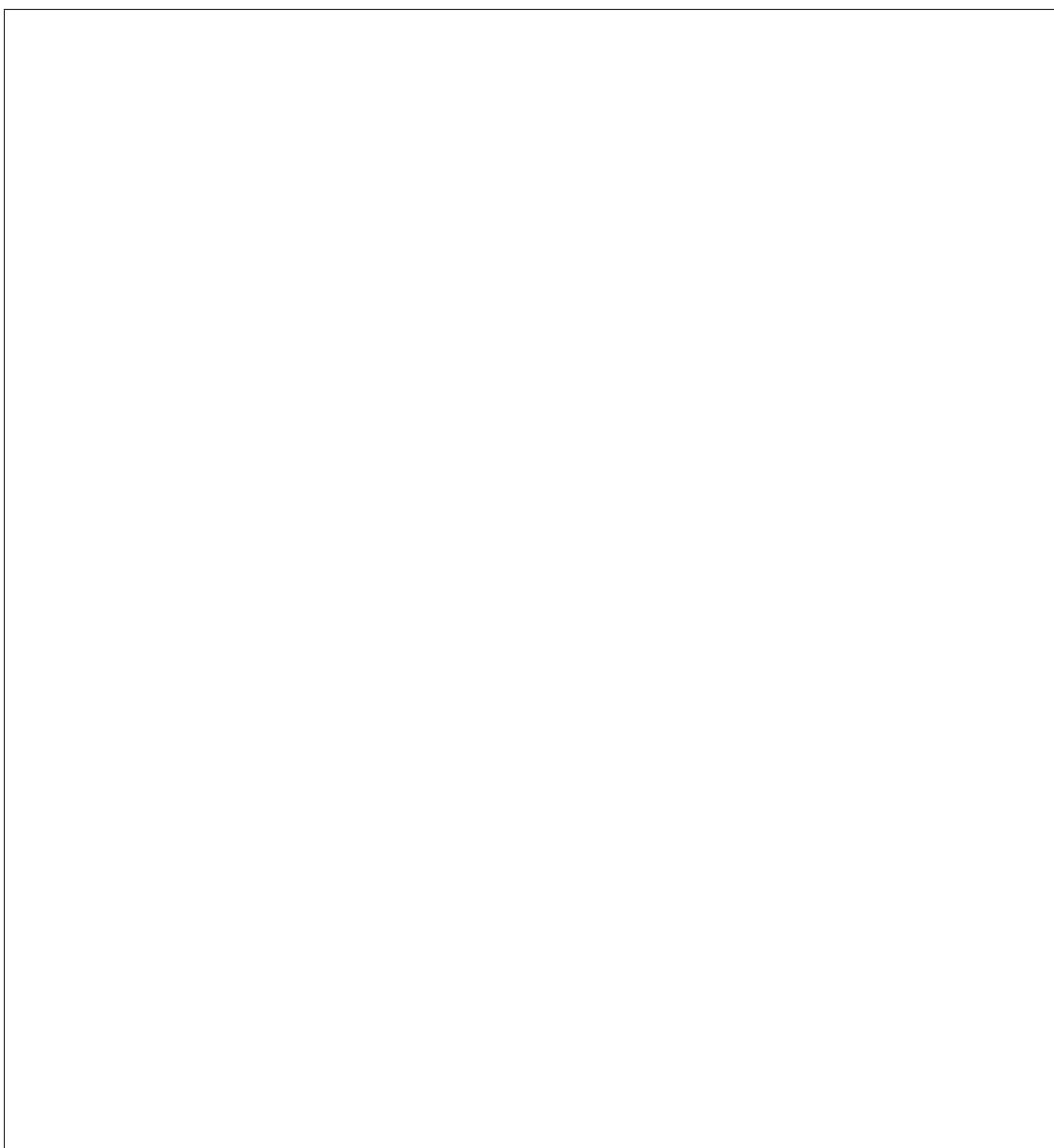


In general, two important properties of the gradient of a function are:

The gradient $\nabla f(a, b)$ is **perpendicular to the contour line through** (a, b) and points in the direction of increasing f . In fact, the direction and magnitude of **steepest slope** at (a, b) are given by $\nabla f(a, b)$ and $\|\nabla f(a, b)\|$.

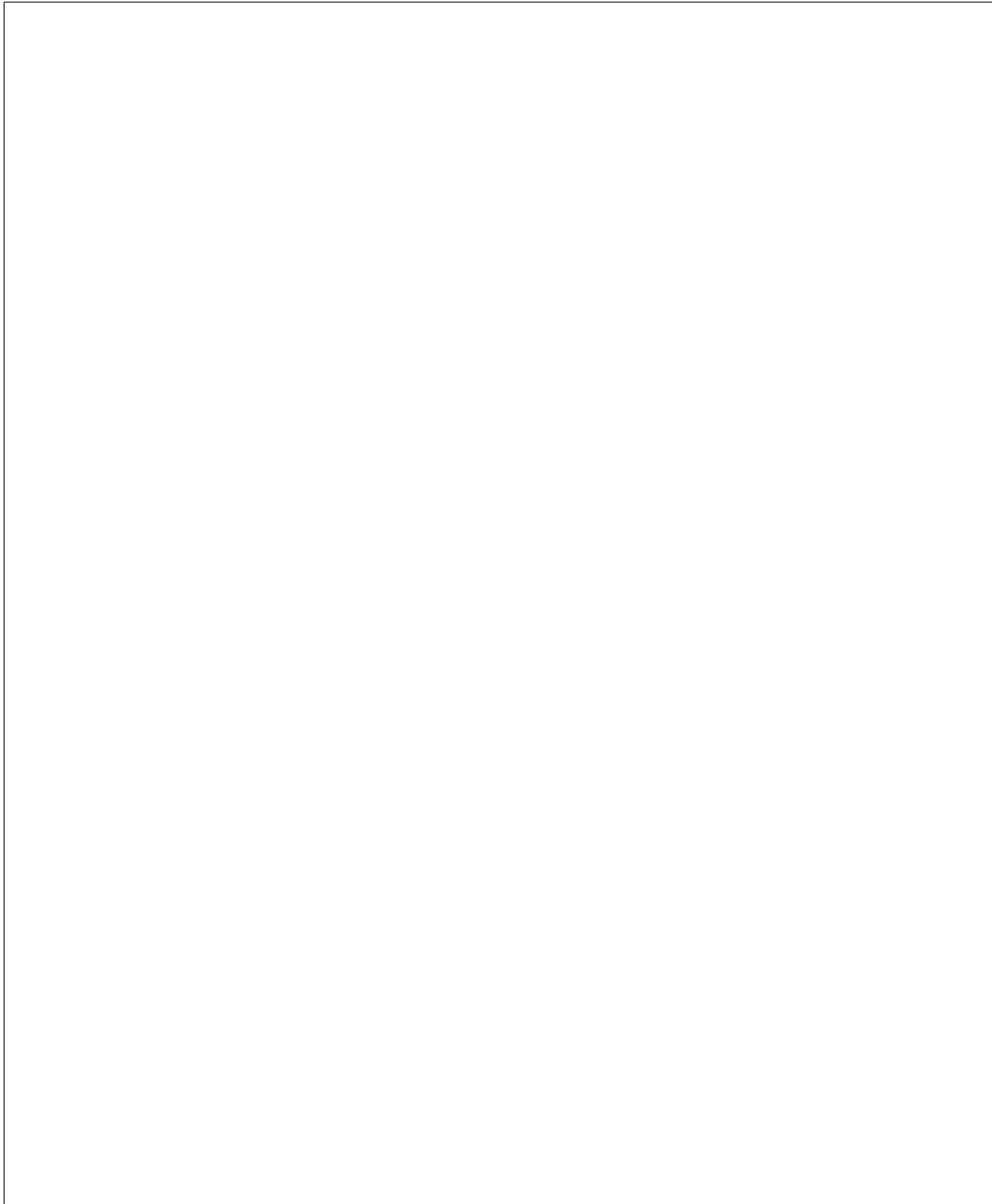
We can understand these two facts by considering the value of $\cos \theta$ in

$$f_{\mathbf{u}} = \nabla f \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|} = \|\nabla f\| \cos \theta.$$



Example: $T(x, y) = 20 - 4x^2 - y^2$ describes the temperature on the surface of a metal plate.

In which direction away from the point $(2, -3)$ does the temperature increase most rapidly?



Example: A team from a large British oil company is mapping the ocean floor to assist in the plugging of a leaking oil well in the Gulf of Mexico. Using sonar, they develop the model

$$D(x, y) = 1700 - 30x^2 - 50 \sin\left(\frac{\pi y}{2}\right),$$

where x and y are distance in kilometres, D is depth in metres, and $-2 \leq x \leq 2$ and $-2 \leq y \leq 2$.

(a) The well is located at $(1, 0.5)$. What is its depth?

(b) Determine the slope of the ocean floor in the positive x -direction and in the positive y -direction in the area considered by the clean-up team.

(c) Determine the magnitude and direction of greatest rate of change of depth at the position of the well.

2.5.4 Main points

- You should be able to find the slope of a function in a given direction.
- You should be able to find the direction and rate of maximal increase of a function.
- You should understand directional derivative and gradient.

2.6 The Chain Rule and Applications

The chain rule is covered in Stewart, Section 15.5.

2.6.1 Review for $f(x)$

For related functions of one variable, such as $y = f(u)$ and $u = g(x)$, we may use the chain rule to find $\frac{dy}{dx}$:

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx} \quad \text{or} \quad f'(x) = f'(u) g'(x).$$

Example: Find $y'(x)$ where $y = f(x) = (x^2 + 1)^5$.

Example: Suppose the radius of a cylinder decreases at a rate of $r'(t) = -2$ cm/s. How fast is the volume decreasing when $r = 1$ cm and $h = 2$ cm?

2.6.2 The chain rule for $f(x, y)$

Given $f(x, y)$ with x and y functions of t . Then

$$\frac{df}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta f}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{f(x(t + \Delta t), y(t + \Delta t)) - f(x(t), y(t))}{\Delta t}.$$

Now if f is smooth and Δf is small, we can relate it to Δx and Δy through the linear approximation

$$\Delta f \simeq f_x \Delta x + f_y \Delta y.$$

Hence

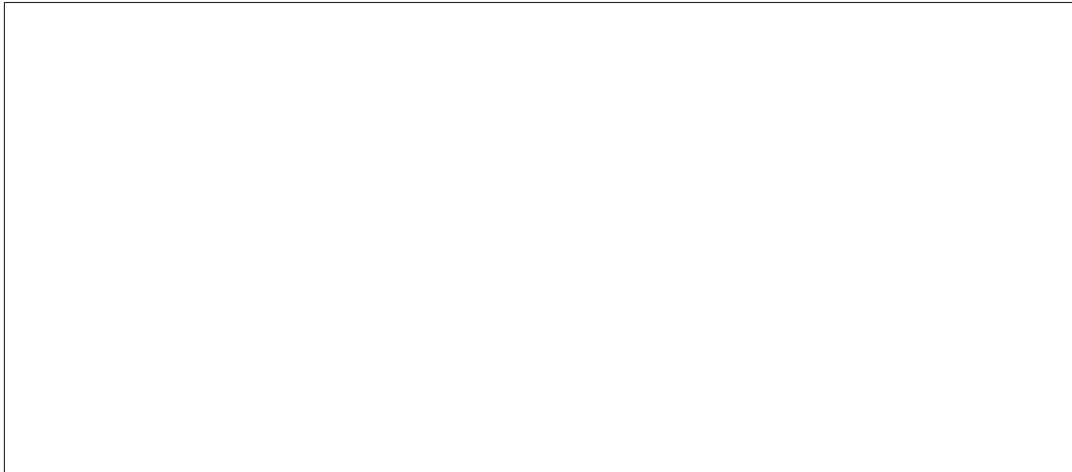
$$\frac{\Delta f}{\Delta t} \simeq f_x \frac{\Delta x}{\Delta t} + f_y \frac{\Delta y}{\Delta t}.$$

Now we let $\Delta t \rightarrow 0$, and provided $x(t)$ and $y(t)$ are smooth,

$$\frac{df}{dt} = f_x \frac{dx}{dt} + f_y \frac{dy}{dt}$$

which is the **chain rule** for $f(x(t), y(t))$.

Example: Continuing the previous example, suppose that not just the radius but also the height h is decreasing: $\frac{dh}{dt} = -1$ cm/s. What is the rate of change in volume?



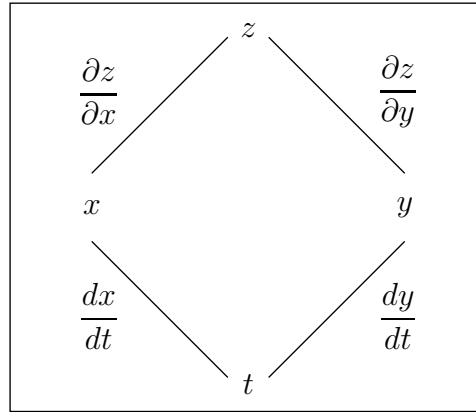
The chain rule can be extended to any number of dimensions, e.g.,

if $V(a(t), b(t), c(t)) = abc$ is the volume of a box then

$$\frac{dV}{dt} =$$

2.6.3 Larger Chain Rules

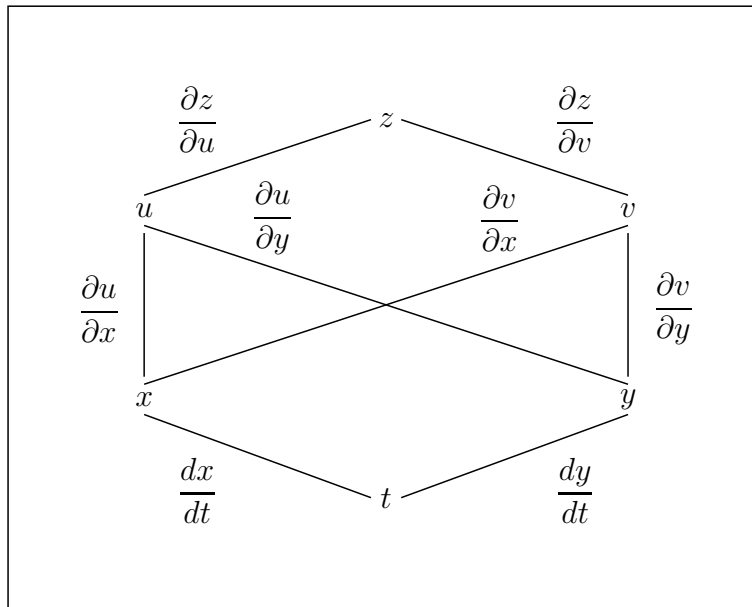
To keep check of what depends on what, you can draw a tree diagram for $z = f(x(t), y(t))$ as follows:



Hence

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \frac{dx}{dt} + \frac{\partial z}{\partial y} \frac{dy}{dt}.$$

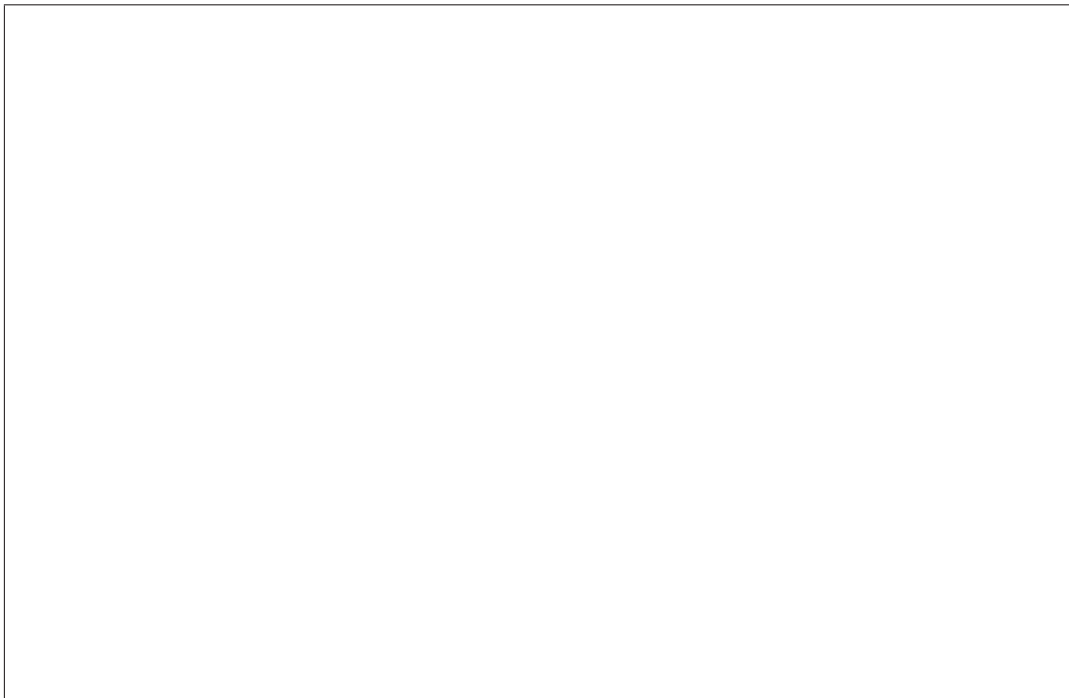
If $z = f(u(x(t), y(t)), v(x(t), y(t)))$ then the tree diagram is



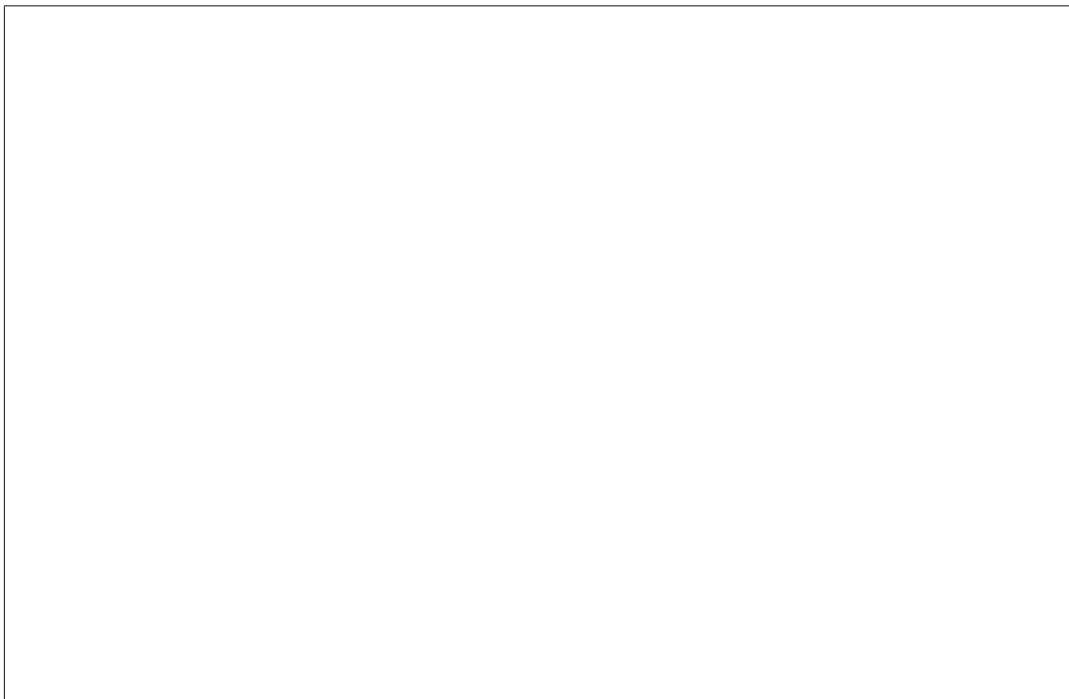
Hence

$$\begin{aligned} \frac{dz}{dt} &= \frac{\partial z}{\partial u} \left(\frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial y} \frac{dy}{dt} \right) + \frac{\partial z}{\partial v} \left(\frac{\partial v}{\partial x} \frac{dx}{dt} + \frac{\partial v}{\partial y} \frac{dy}{dt} \right) \\ &= \left(\frac{\partial z}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial x} \right) \frac{dx}{dt} + \left(\frac{\partial z}{\partial u} \frac{\partial u}{\partial y} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial y} \right) \frac{dy}{dt}. \end{aligned}$$

Example: Find the slope of $z = f(x, y) = \sin x \cos y$ in the radial direction at $(\pi, 2\pi)$. That is, find $\frac{\partial z}{\partial r}$.



Exercise: Given $h = f(x, y, z)$ and each of x , y and z depend on both s and t , find an expression for $\frac{\partial h}{\partial t}$.



2.6.4 Implicit Differentiation

Both $y + \sin(x + y) = 0$ and $x^{11} - y^7 + xy^2 + 1 = 0$ correspond to **curves** in \mathbb{R}^2 .

In neither of these examples can we explicitly solve for y as a function of x (or vice versa). This raises the following question: Can we compute $y'(x)$?

The answer to this question is affirmative and the method to do this goes by the name of **implicit differentiation**.

First we will try to find $y'(x)$ by “brute force”. It is only advisable to use this method if you have a PhD in applications of the chain rule.

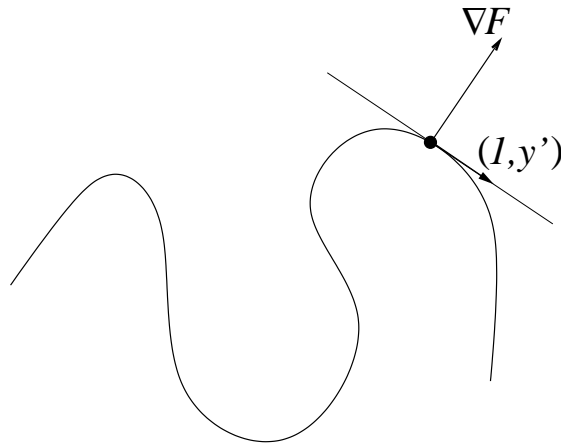
Example: Find $y'(x)$ on the curve $y + \sin(x + y) = 0$.



Important remark: Although we have found $y'(x)$ it is almost never possible to express the answer in terms of x only (as this requires y to be solved for x). Extreme care has thus to be taken in using the above formula for $y'(x)$. It only has meaning for points (x, y) that lie on the curve $y + \sin(x + y) = 0$ (such as $(0, 0)$ or $(-1 - \pi/2, 1)$).

So what if you do not have PhD in applications of the chain rule?

The idea is to view a given curve C in \mathbb{R}^2 (such as $y + \sin(x + y) = 0$) as the contour $z = 0$ of a function $F(x, y)$. That is, we define $F(x, y) = y + \sin(x + y)$. We already know that, for any given point (x, y) on C , $\nabla F(x, y)$ corresponds to a vector perpendicular to C in that point.



In other words,

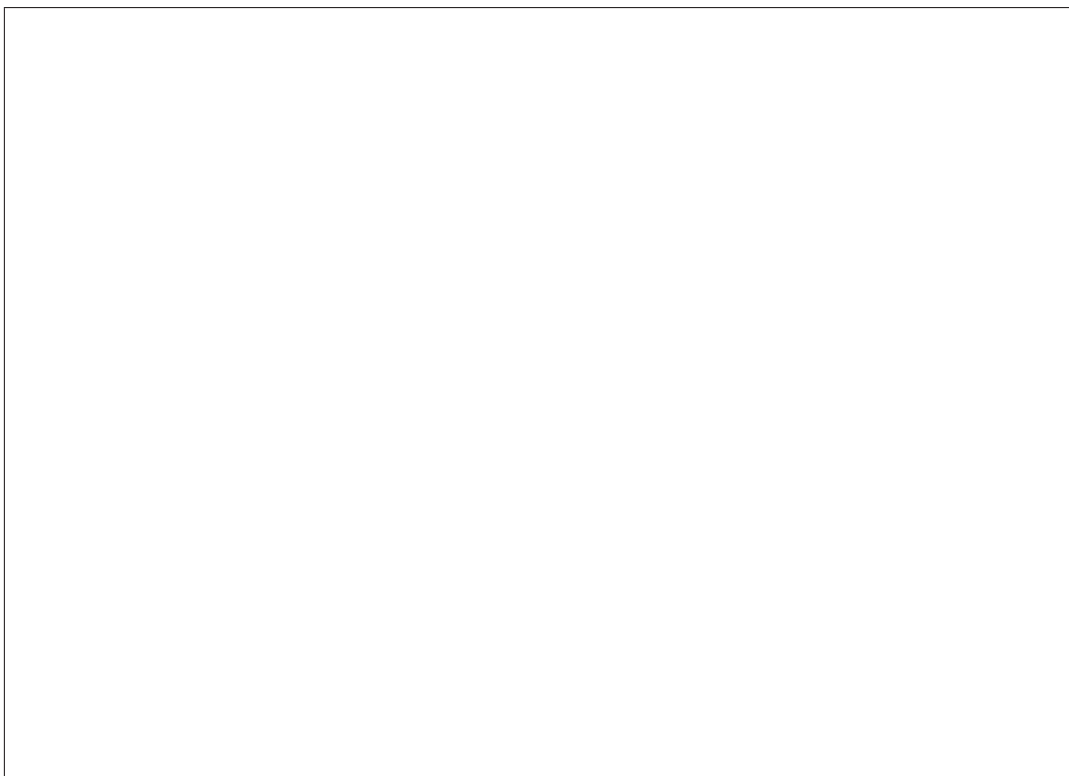
$$\nabla F(x, y) \cdot (1, y'(x)) = 0.$$

Solving this equation for $y'(x)$ gives

$$\frac{dy}{dx} = -\frac{F_x}{F_y}.$$

Exercise: Find $y'(x)$ on the curve $y + \sin(x + y) = 0$.

Example: The Folium of Descartes is given by $x^3 + y^3 = 6xy$. Find $y'(x)$.



Question: What happens at $x = y = 0$?

2.6.5 Main points

- You should be able to differentiate a multivariable function whose variables depend on multivariable functions.
- You should be able to find $\frac{dy}{dx}$ from implicit forms.

3 Max and Min Problems on Surfaces

The content of this chapter is found in Stewart, Sections 15.7 and 15.8.

3.1 Quadratic Approximation

In the previous chapter, we looked at the **first-order**, or **linear** approximation to a function. In this section we introduce the **second-order**, or **quadratic** approximation to a function.

3.1.1 Review for $f(x)$

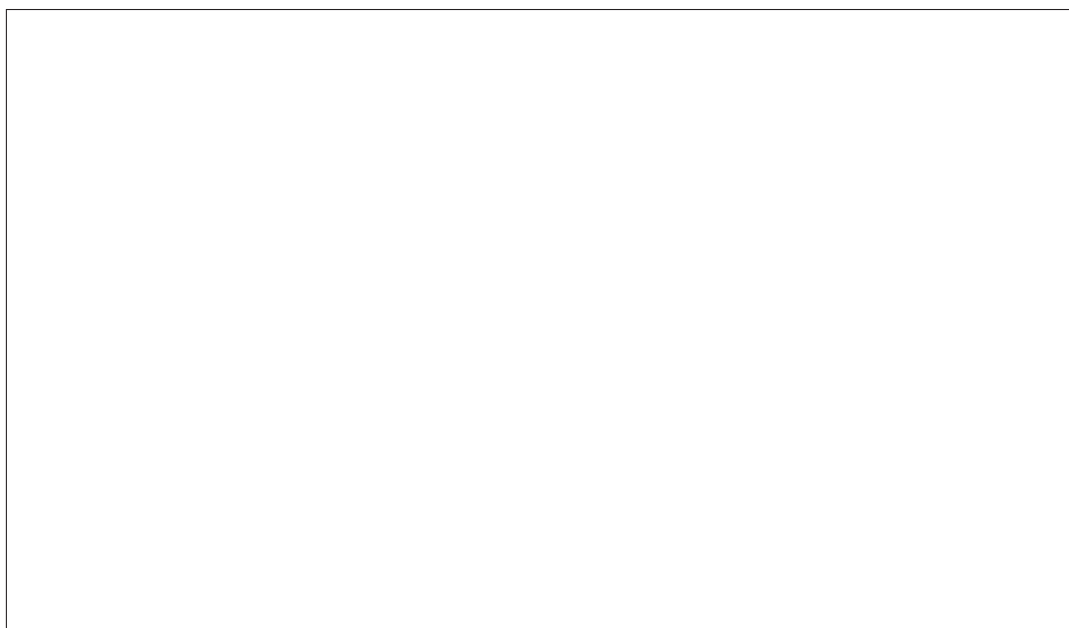
We saw in Section 2.4.1 that the value of $f(x)$ near $x = a$ can be approximated using a straight line:

$$f(x) \simeq f(a) + f'(a)(x - a).$$

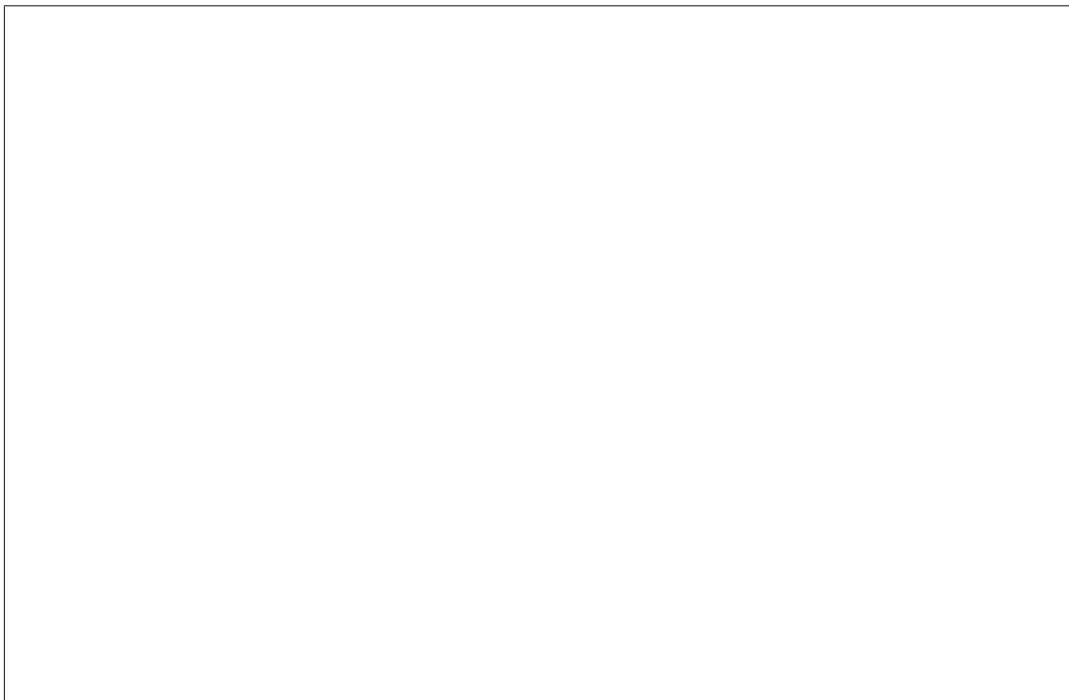
Adding a **second-order** term will improve this approximation

$$f(x) \simeq f(a) + f'(a)(x - a) + \frac{f''(a)}{2} (x - a)^2.$$

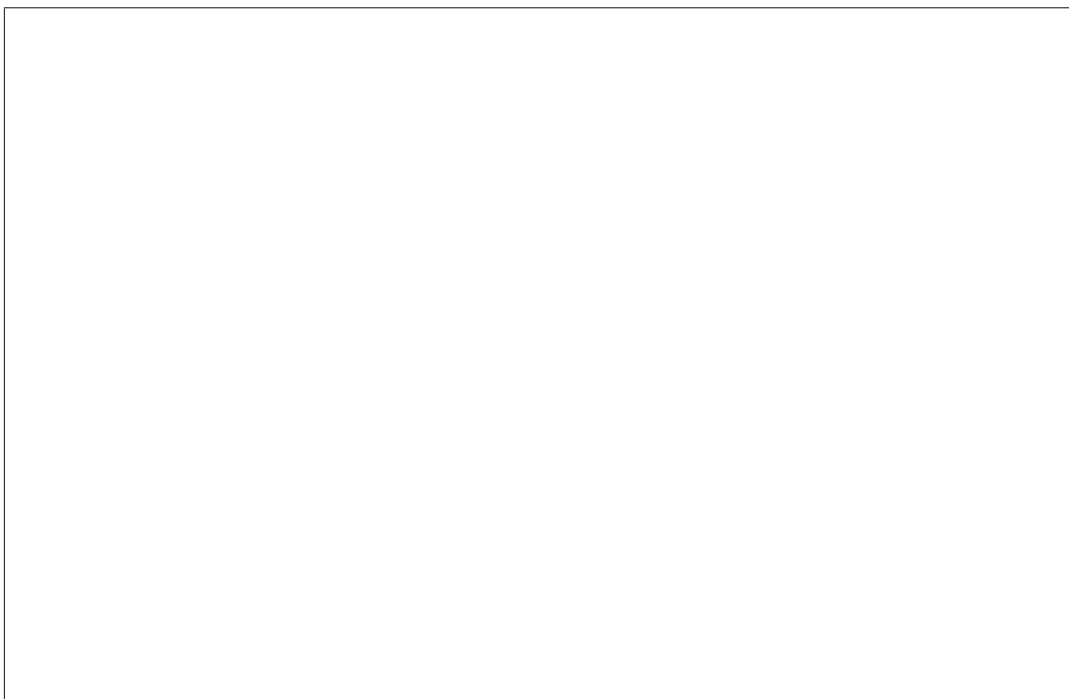
This **quadratic** (or **second-order**) approximation corresponds to the approximation of f by the parabola through $(a, f(a))$ with slope and second derivative at a matching those of f . Unless f is very wild around $x = a$ this is generally quite a good approximation for $x - a$ not too large.



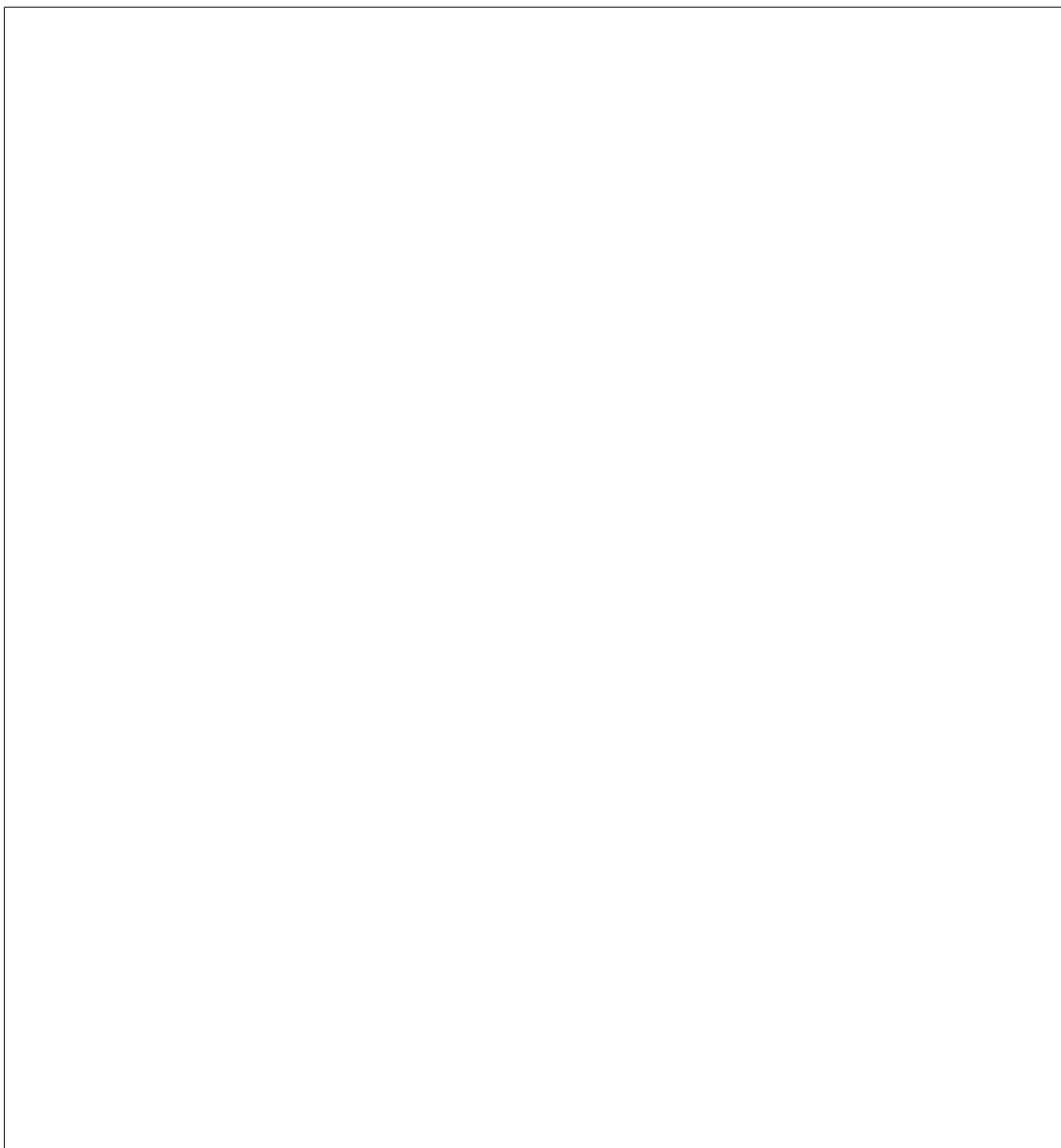
Example: Compute $e^{0.1}$ using a linear and a quadratic approximation.



Example: Approximate $\sqrt{26}$.



Example: In the same graph, sketch $f(x) = \cos x$ as well as its linear and quadratic approximations around 0.



3.1.2 Quadratic approximations of $f(x, y)$

This material is covered in Stewart p. 969.

The **quadratic** or **second-order** approximation to $f(x, y)$ around (a, b) is a function of the form

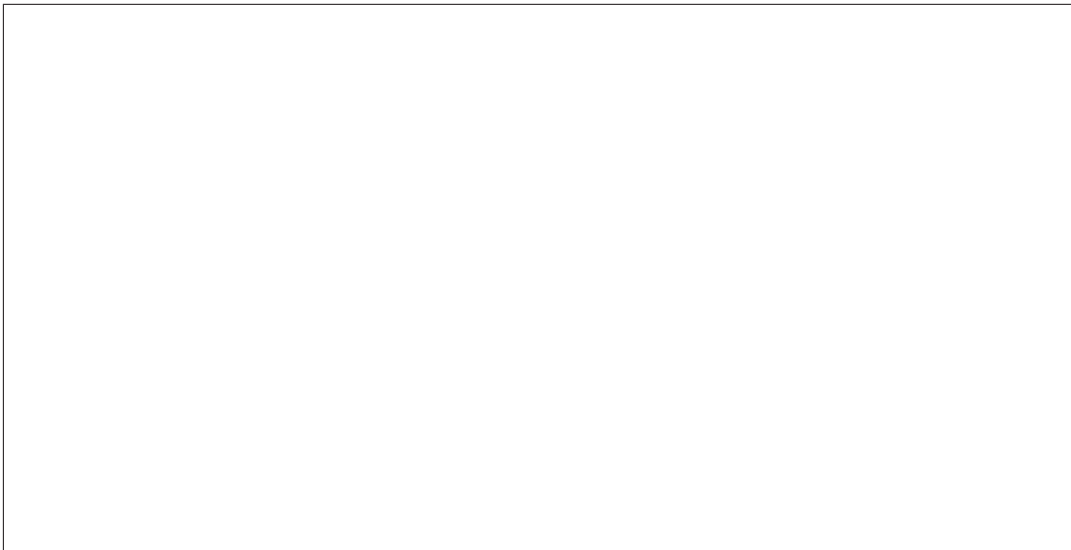
$$Q(x, y) = h + mx + ny + ax^2 + bxy + cy^2$$

such that $Q(a, b) = f(a, b)$ and such that all first order and second order partial derivatives of f and Q agree at (a, b) .

It is not hard to verify that this gives

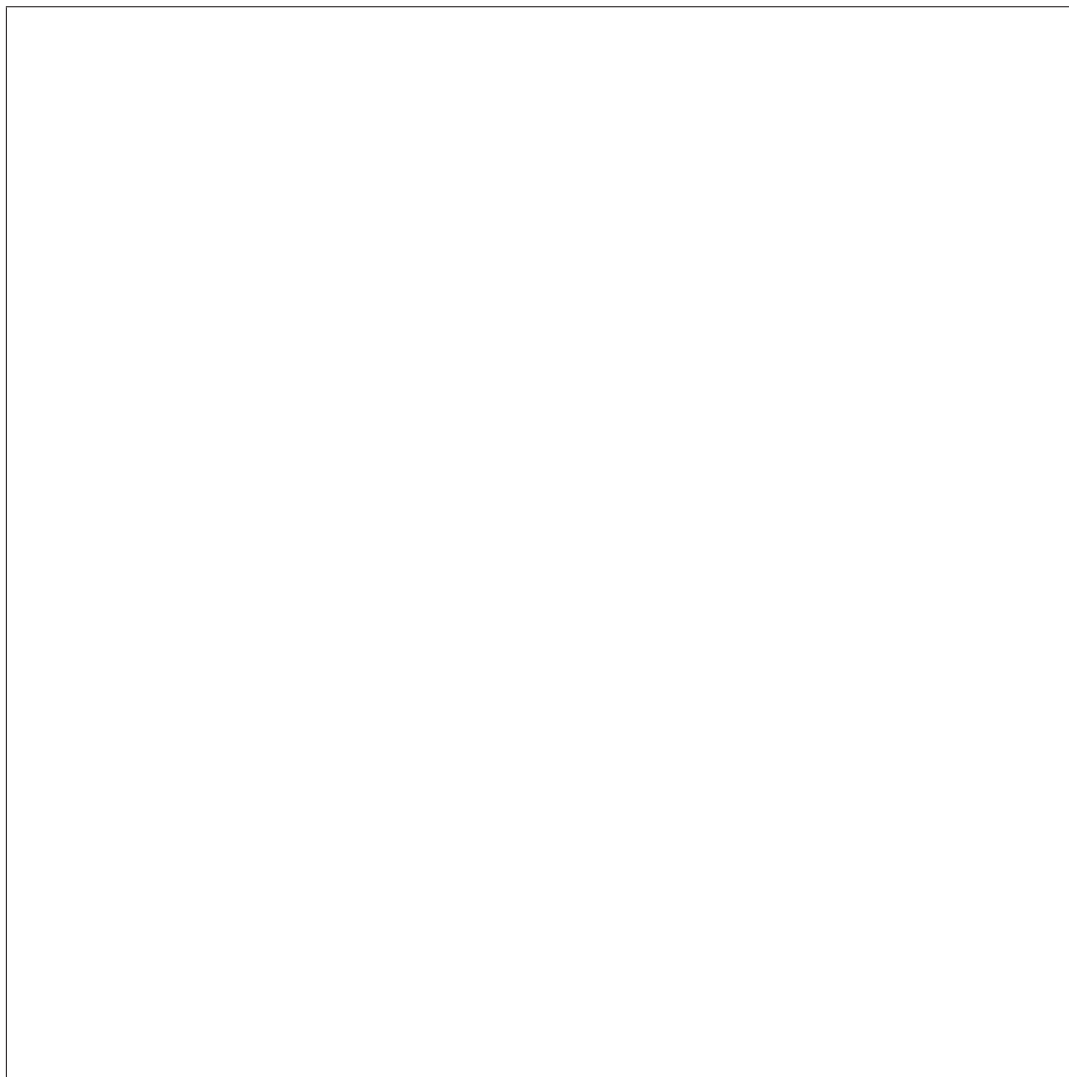
$$\begin{aligned} Q(x, y) &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &\quad + \frac{1}{2}f_{xx}(a, b)(x - a)^2 + f_{xy}(a, b)(x - a)(y - b) + \frac{1}{2}f_{yy}(a, b)(y - b)^2. \end{aligned}$$

Example: Check that $Q_{xx}(a, b) = f_{xx}(a, b)$ and $Q_{xy}(a, b) = f_{xy}(a, b)$.



Example: Find the quadratic approximation around $(0, 0)$ of

$$f(x, y) = 1 - x^2 + y^2 + xy + x^3 + x^2y^2.$$



If $|x|, |y| < 0.1$ then we are ignoring terms on the order of $(0.1)^3 = 0.001$.

In this example the function is a polynomial in x and y and we can immediately read off the quadratic approximation. In the next example we really need to use the formula (or do we?).

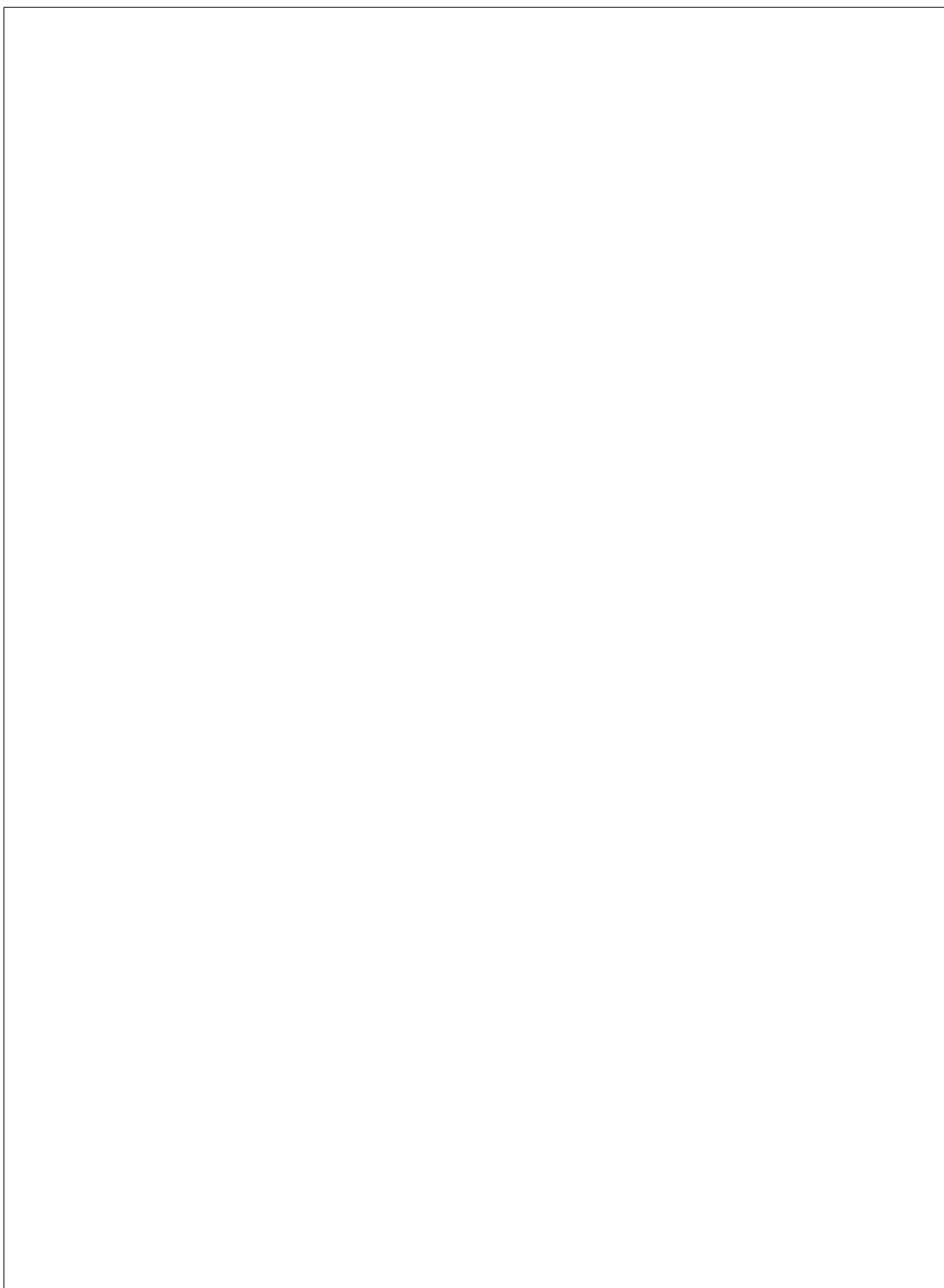
Example: Find the linear and quadratic approximations around $(0, 0)$ of

$$f(x, y) = e^{-x^2 - y^2}.$$



Important remark: Using quadratic approximations is a good idea if you want to know the value of a complicated function at a given point, close to another point where the value is actually known. It never is a good idea to use approximations for simple functions such as polynomials. The example below nicely illustrates this point.

Example: Find the quadratic approximation to $f(x, y) = x^2y$ around $(1, 2)$.



Example: Consider, again, a barrel with base of radius 1 and height 2.

$$\begin{aligned}
 V(r, h) = \pi r^2 h &\quad \Rightarrow \quad V(1, 2) = 2\pi \\
 \frac{\partial V}{\partial r} = 2\pi r h &\quad \Rightarrow \quad \frac{\partial V}{\partial r}(1, 2) = 4\pi \\
 \frac{\partial V}{\partial h} = \pi r^2 &\quad \Rightarrow \quad \frac{\partial V}{\partial h}(1, 2) = \pi \\
 \frac{\partial^2 V}{\partial r^2} = 2\pi h &\quad \Rightarrow \quad \frac{\partial^2 V}{\partial r^2}(1, 2) = 4\pi \\
 \frac{\partial^2 V}{\partial r \partial h} = 2\pi r &\quad \Rightarrow \quad \frac{\partial^2 V}{\partial r \partial h}(1, 2) = 2\pi \\
 \frac{\partial^2 V}{\partial h^2} &= 0.
 \end{aligned}$$

Hence the quadratic approximation around $(1, 2)$ is

$$\begin{aligned}
 V(r, h) &\simeq Q(r, h) \\
 &= 2\pi + 4\pi(r - 1) + \pi(h - 2) + 2\pi(r - 1)^2 + 2\pi(r - 1)(h - 2).
 \end{aligned}$$

Let $\Delta r = r - 1$ and $\Delta h = h - 2$. If $\Delta r = 0.05$ and $\Delta h = 0.1$, what is the approximate error in volume ΔV ?

3.1.3 Main points

- You should be able to find the quadratic approximation given a function of two variables.
- In Matlab, you should be able to plot both the linear and quadratic approximations of a function, and make a comparison of the errors.
- The quadratic approximation is superior to the linear approximation.

3.2 Critical Points and Optimization

3.2.1 Review for $f(x)$

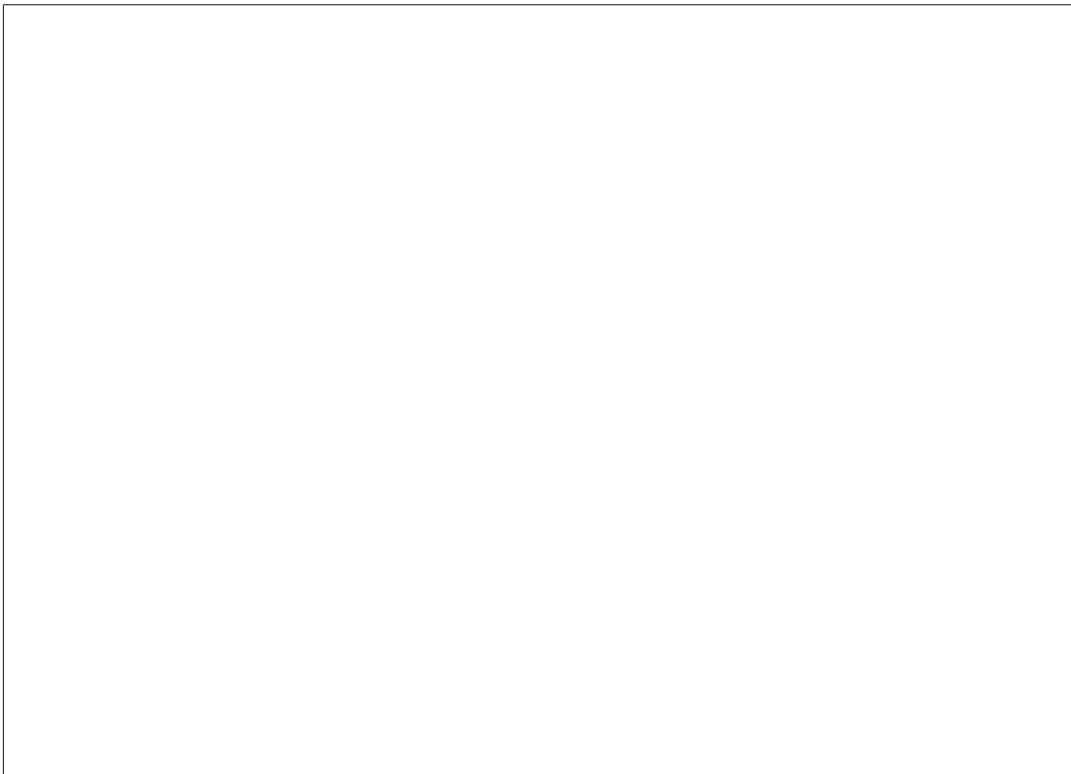
Smooth functions f of one variable have **local** maxima and minima where $f(x)$ has zero slope, i.e., where the first derivative of f is zero:

$$\frac{df}{dx} = 0.$$

The **second derivative test** tells us whether $f(x)$ is concave up (a minimum) or concave down (a maximum):

$$\begin{aligned}\frac{d^2f}{dx^2} > 0 &\Rightarrow \text{local minimum,} \\ \frac{d^2f}{dx^2} < 0 &\Rightarrow \text{local maximum.} \\ \frac{d^2f}{dx^2} = 0 &\Rightarrow ?\end{aligned}$$

Example: Find the local maxima and minima of $f(x) = x^2 - x^4$.



Note, however, that the local minimum in the above example is not a global minimum. At the local minimum, $x = 0$ and $f(0) = 0$. But for $|x| > 1$, we have $f(x) < 0$.

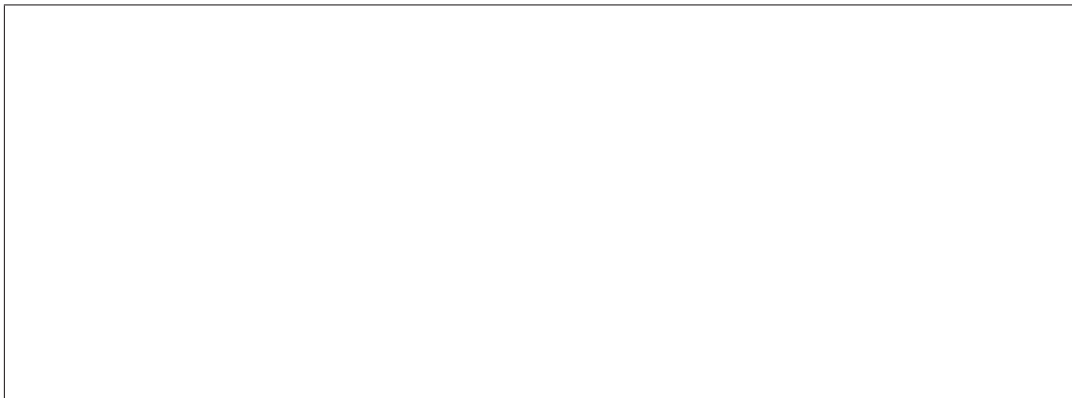
If we define the domain as $[-2, 2]$, the **global** minimum occurs on the boundary at $f(\pm 2) = 4 - 16 = -12$.

The global maximum, however, is at the local maximum, i.e. $f\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

The global maximum or minimum of a continuous function $f(x)$ occurs either at a local minimum, a local maximum, or on the boundary of its domain.

In some circumstances there may be local maxima or minima in points where the derivative $\frac{df}{dx}$ is not defined (at cusps, for instance).

Example: $f(x) = |x - 3|$.



A more general definition of a **local minimum or maximum** is:

- A local minimum occurs at a point a where $f(x) \geq f(a)$ for all x sufficiently close to a .
- A local maximum occurs at a point a where $f(x) \leq f(a)$ for all x sufficiently close to a .

If f is continuous, a local maximum or minimum occurs either when $f'(x) = 0$ or when $f'(x)$ is undefined.

3.2.2 Critical points for $f(x, y)$

In this section we will find and classify critical points of a function of two variables, see Stewart, Section 15.7. (There are also some useful Matlab functions which perform minimisation, such as **fmin** and **fminsearch**).

First we define **local maxima** and **local minima**:

- $f(x, y)$ has a *local maximum* at (a, b) if $f(a, b) \geq f(x, y)$ for all (x, y) sufficiently close to (a, b)
- $f(x, y)$ has a *local minimum* at (a, b) if $f(a, b) \leq f(x, y)$ for all (x, y) sufficiently close to (a, b) .

If f is differentiable, a local maximum or minimum can only occur in points where the corresponding tangent plane is horizontal, i.e., where

$$\nabla f = \mathbf{0}.$$

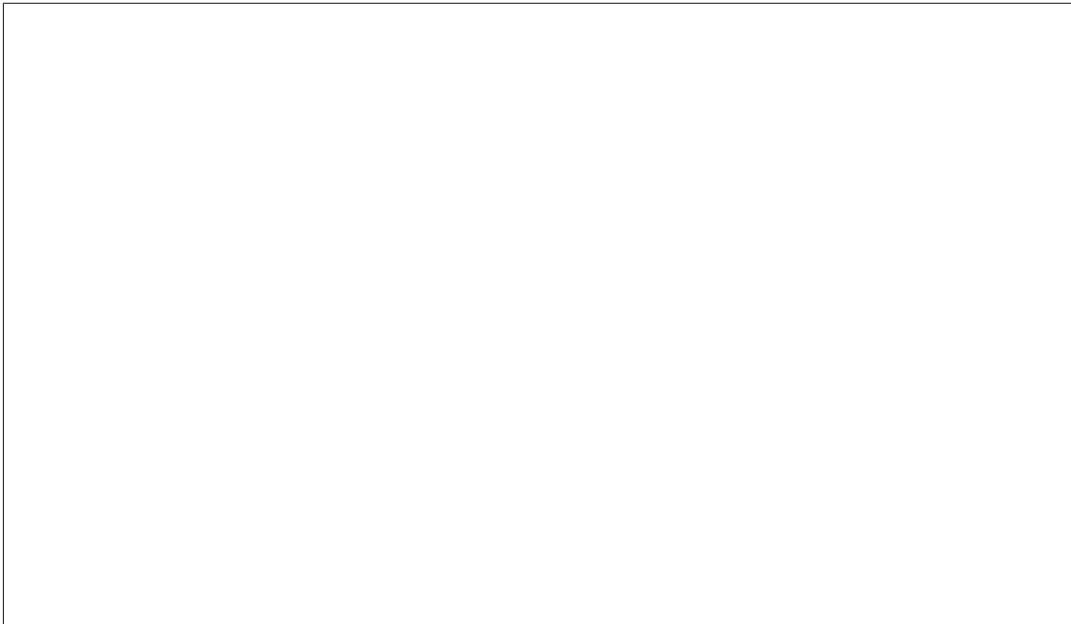
Note that if either one of $\frac{\partial f}{\partial x}$ or $\frac{\partial f}{\partial y}$ is undefined then ∇f is undefined.

Definition: A critical point of $f(x, y)$ is a point where either $\nabla f = (0, 0)$ or where ∇f is undefined.

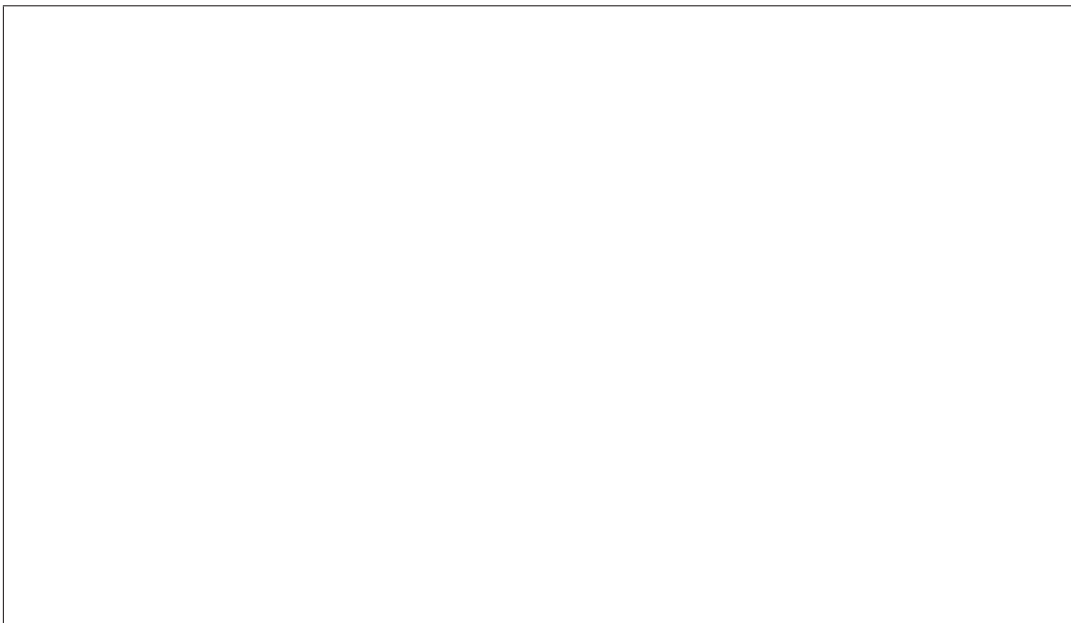
Local maxima and minima occur at critical points, but not all critical points are local maxima or minima.

Remark: (For the mathematically inclined.) The rather vague description “sufficiently close” can be made much more precise. In the simplest case, when $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, it goes as follows. The function f has a *local maximum* at (a, b) if there exists an $\epsilon > 0$ such that $f(a, b) \geq f(x, y)$ for all (x, y) such that $\|(x, y) - (a, b)\| < \epsilon$.

Example: Find the critical points of $f(x, y) = x^2 + 3x + y^2$.



Example: Find the critical points of $f(x, y) = x^2 + 3x - y^2$.



Definition. A function f has a **saddle point** at P_0 if P_0 is a critical point of f and if within any distance of P_0 (no matter how small) there are points P_1 and P_2 with

$$f(P_1) > f(P_0) \quad \text{and} \quad f(P_2) < f(P_0).$$

Important remark: Not all saddle points look like a “horse saddle”, and in one of your Matlab sessions you will be asked to look at a so-called monkey saddle, described by an equation of the form $z = x^3 - 3xy^2$.

Example: Find all the critical points of

$$f(x, y) = \left(\frac{1}{2} - x^2 + y^2\right)e^{(1-y^2-x^2)},$$

and use a quadratic approximation to investigate the nature of the critical points.

$$\frac{\partial f}{\partial x} =$$

$$\frac{\partial f}{\partial y} =$$

$$\nabla f = 0 \quad \Rightarrow \quad \frac{\partial f}{\partial x} = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = 0$$

\Rightarrow

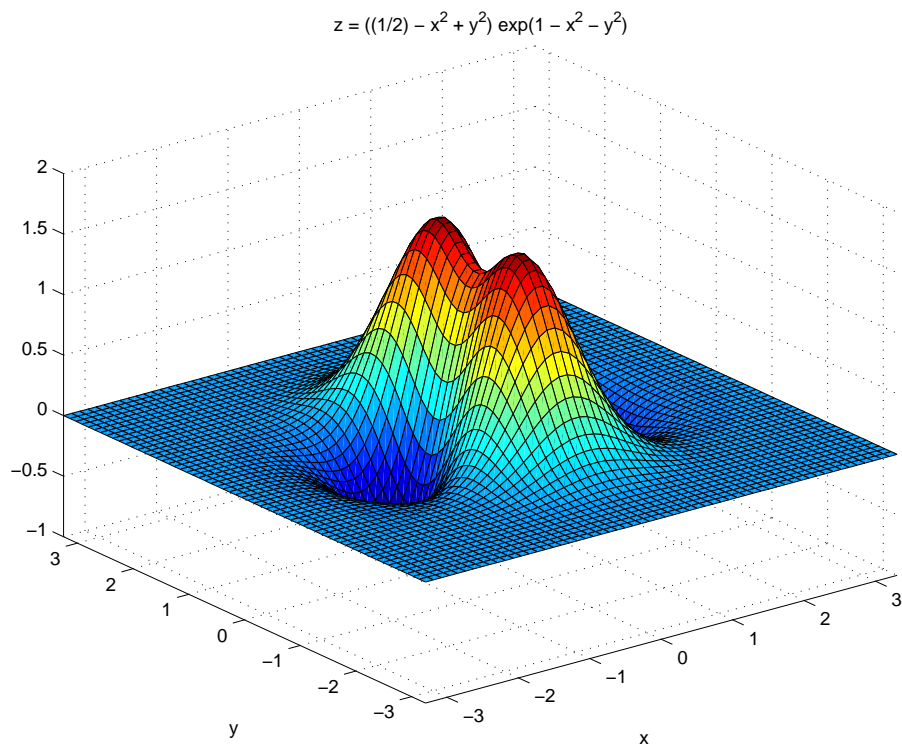
There are four cases to consider:

- (1) $x = 0$ and $y = 0$;
- (2) $x = 0$ and $\frac{1}{2} + x^2 - y^2 = 0$;
- (3) $\frac{3}{2} - x^2 + y^2 = 0$ and $y = 0$;
- (4) $\frac{3}{2} - x^2 + y^2 = 0$ and $-\frac{1}{2} - x^2 + y^2 = 0$.

Solving each of these gives a total of five critical points

$$(1) \quad (0, 0), \quad (2) \quad (0, \pm\frac{1}{2}\sqrt{2}), \quad (3) \quad (\pm\frac{1}{2}\sqrt{6}, 0).$$

Note that case (4) has no solution.



Now

$$\frac{\partial^2 f}{\partial x^2} =$$

so that

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2}(0, 0) &= \\ \frac{\partial^2 f}{\partial x^2}\left(0, \pm\frac{1}{2}\sqrt{2}\right) &= \\ \frac{\partial^2 f}{\partial x^2}\left(\pm\frac{1}{2}\sqrt{6}, 0\right) &= \end{aligned}$$

Next,

$$\frac{\partial^2 f}{\partial x \partial y} =$$

so that

$$\begin{aligned}\frac{\partial^2 f}{\partial x \partial y}(0, 0) &= \\ \frac{\partial^2 f}{\partial x \partial y}\left(0, \pm \frac{1}{2}\sqrt{2}\right) &= \\ \frac{\partial^2 f}{\partial x \partial y}\left(\pm \frac{1}{2}\sqrt{6}, 0\right) &= \end{aligned}$$

Finally,

$$\frac{\partial^2 f}{\partial y^2} =$$

so that

$$\begin{aligned}\frac{\partial^2 f}{\partial y^2}(0, 0) &= \\ \frac{\partial^2 f}{\partial y^2}\left(0, \pm \frac{1}{2}\sqrt{2}\right) &= \\ \frac{\partial^2 f}{\partial y^2}\left(\pm \frac{1}{2}\sqrt{6}, 0\right) &= \end{aligned}$$

The quadratic approximation of f at $(0, 0)$ is thus

$$Q(x, y) =$$

The point $(0, 0)$ is a saddle point.

The quadratic approximation at $(0, \pm\frac{1}{2}\sqrt{2})$ is

$$Q(x, y) =$$

The points $(0, \pm\frac{1}{2}\sqrt{2})$ are local maxima.

The quadratic approximation at $(\pm\frac{1}{2}\sqrt{6}, 0)$ is

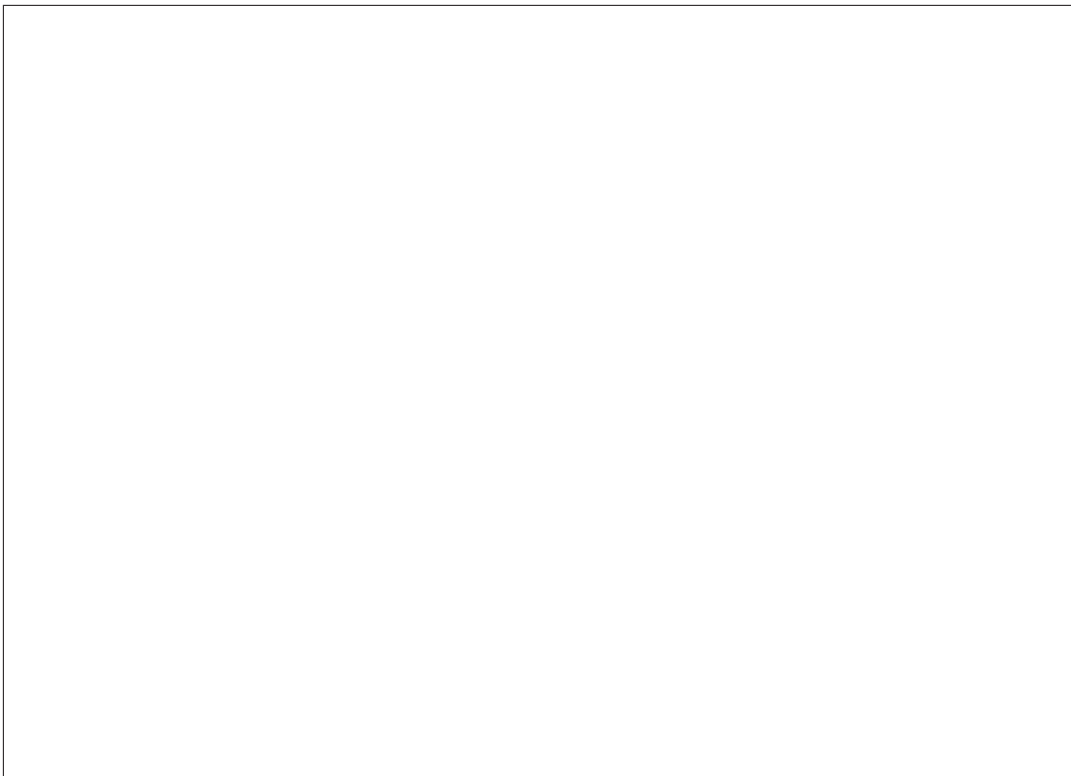
$$Q(x, y) =$$

The points $(\pm\frac{1}{2}\sqrt{6}, 0)$ are local minima.

The plus or minus signs in front of x^2 and y^2 in the quadratic approximation often are a good indicator as to whether f has a maximum, minimum or saddle at a given critical point.

However, when $Q(x, y)$ has cross-terms such as xy one needs to be extremely careful.

Example: Examine the critical point $(0, 0)$ of $f(x, y) = x^2 + xy + y^2$.



So f has a local minimum at $(0, 0)$.

Next consider

$$f(x, y) = x^2 + xy + \frac{y^2}{8} = \left(x + \frac{1}{2}y\right)^2 - \frac{1}{8}y^2.$$

This function has a saddle point at $(0, 0)$.

Question: Can we find a general rule to determine the nature of critical points?

Let Q be the quadratic function

$$\begin{aligned} Q(x, y) &= ax^2 + bxy + cy^2 + dx + ey + f \\ &= a \left(\left(x + \frac{by}{2a} \right)^2 - \frac{b^2y^2}{4a^2} + \frac{cy^2}{a} \right) + dx + ey + f \\ &= a \left(\left(x + \frac{by}{2a} \right)^2 + \frac{(4ac - b^2)}{4a^2} y^2 \right) + dx + ey + f. \end{aligned}$$

Then

- $a > 0$ and $4ac - b^2 > 0$ implies that f has a local minimum;
- $a < 0$ and $4ac - b^2 > 0$ implies that f has a local maximum;
- $4ac - b^2 < 0$ implies that f has a saddle.

Applying this argument to our quadratic approximation formula for f around (a, b) :

$$\begin{aligned} Q(x, y) &= f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \\ &\quad + \frac{1}{2}f_{xx}(a, b)(x - a)^2 + f_{xy}(a, b)(x - a)(y - b) + \frac{1}{2}f_{yy}(a, b)(y - b)^2 \end{aligned}$$

gives what is known as the **second-derivative test**.

Theorem: Assume that f and its first- and second-order partial derivatives are all continuous at (a, b) and $\nabla f(a, b) = (0, 0)$. Define the 2×2 determinant

$$D = \begin{vmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{vmatrix}.$$

Then:

- $f_{xx}(a, b) > 0$ and $D > 0$ implies that (a, b) is a local minimum;
- $f_{xx}(a, b) < 0$ and $D > 0$ implies that (a, b) is a local maximum;
- $D < 0$ implies that (a, b) is a saddle point;
- if $D = 0$ the second-derivative test is inconclusive.

Remark: (For the connoisseurs.) The matrix

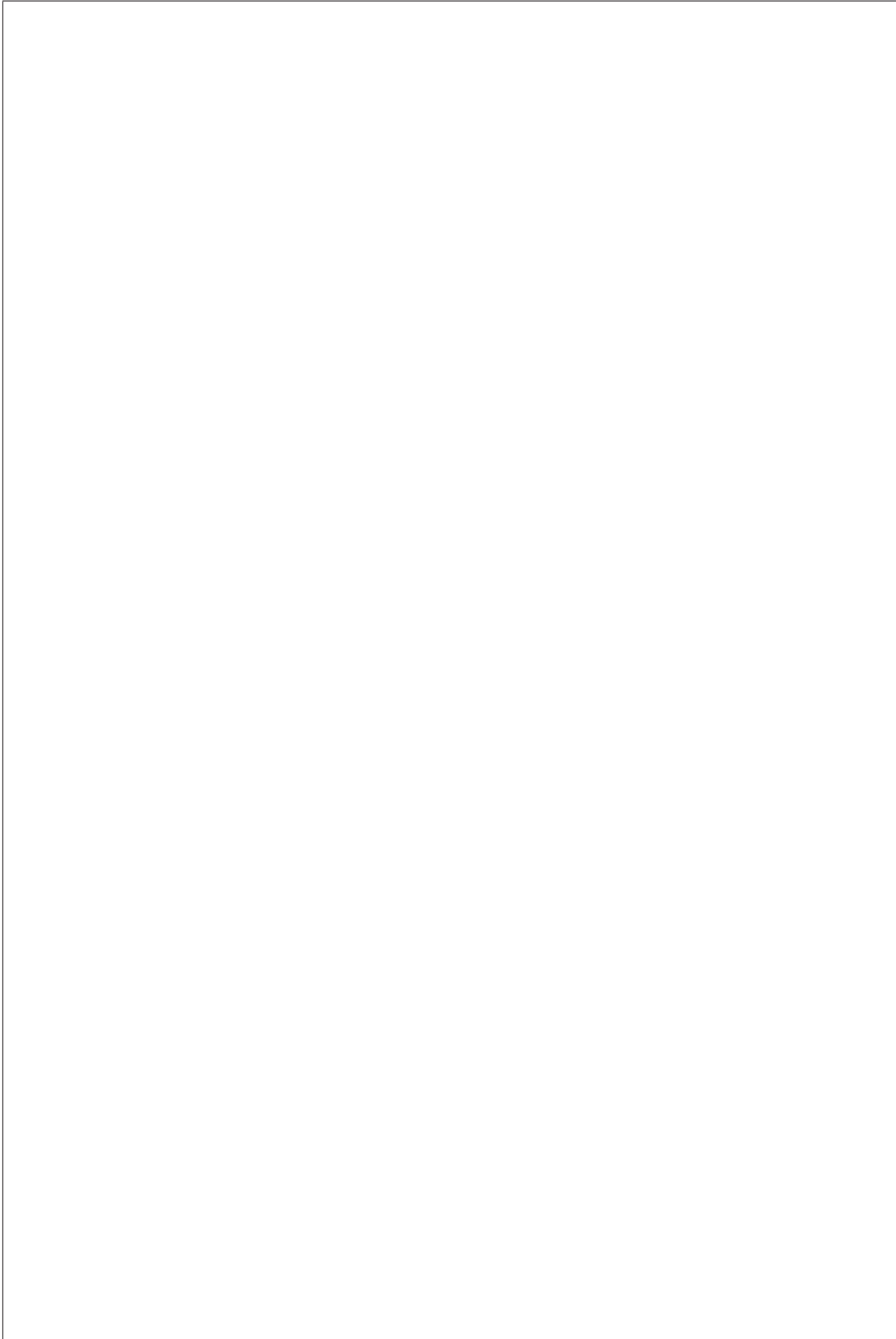
$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

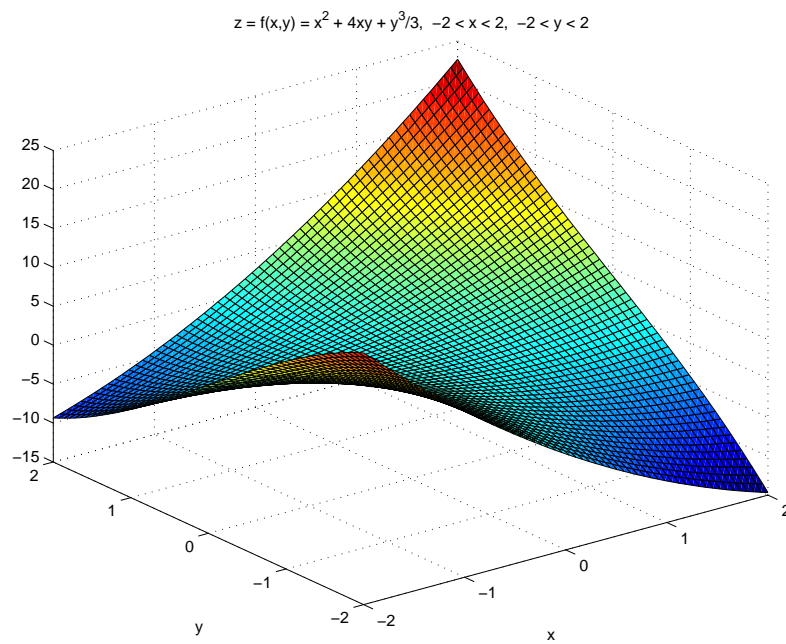
is known as the **Hessian** and its determinant as the **discriminant**.

The beauty of this test is that we do not even have to write down the quadratic approximation.

Example: Find the critical points of $f(x, y) = e^{-x^2+y^2}$ and classify them.

Example: Find and classify the critical points of $f(x, y) = x^2 + 4xy + \frac{1}{3}y^3$.



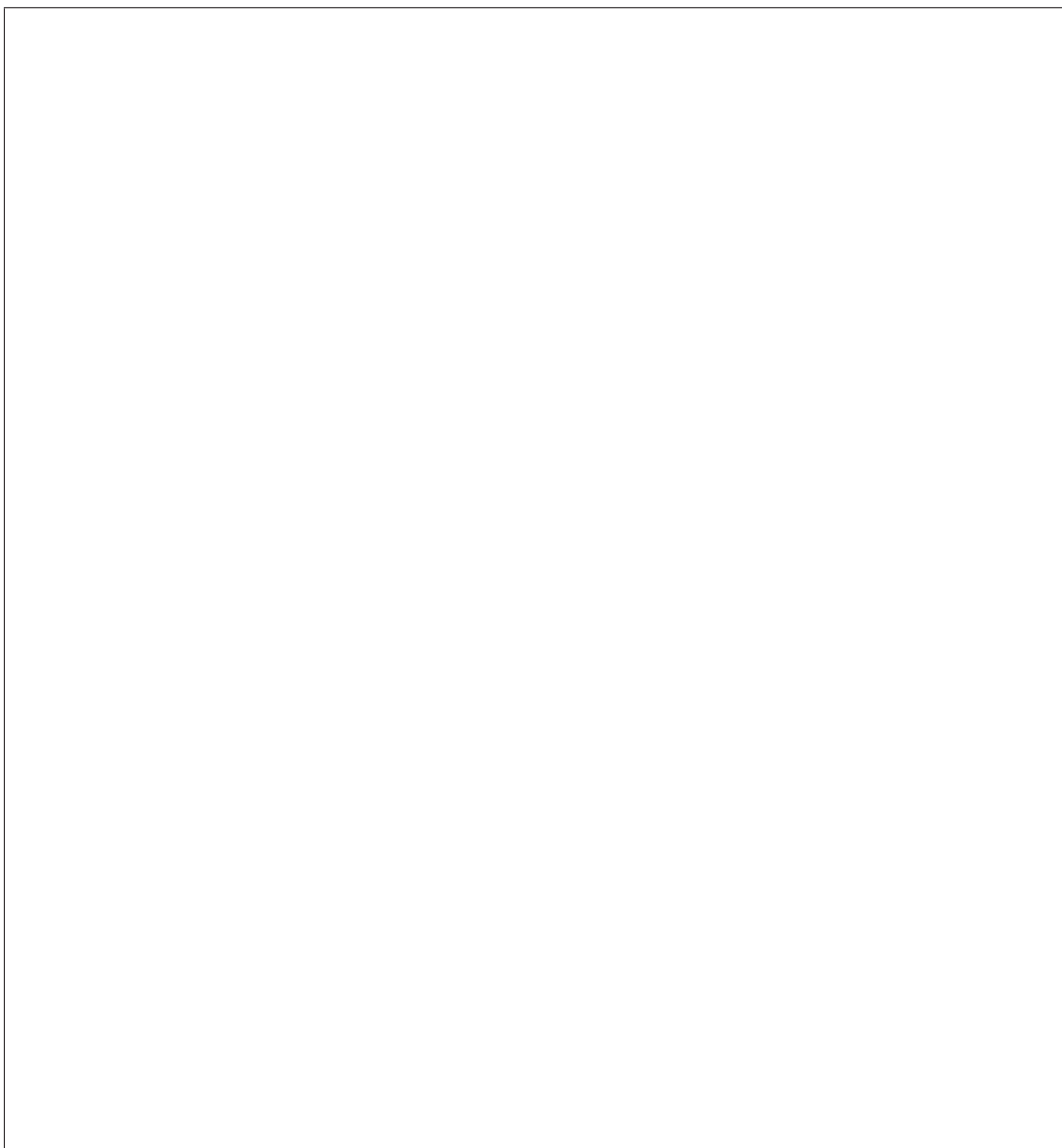


3.2.3 Global maxima and minima

To find the global maximum or minimum values of f on a closed and bounded domain D , we need to compare both

- the values of f at the critical points (where $\nabla f = 0$ or undefined), and
- the extreme values of f on the boundary of D .

Example: What are the global maximum and minimum values of the function $f(x, y) = x^2 + 4xy + \frac{1}{3}y^3$ on the domain $[-2, 2] \times [-2, 2]$?





3.2.4 The Matlab command `fminsearch`

The *Nelder–Mead* algorithm is a search method built into the Matlab function `fminsearch`. One advantage of this method is that it works with non-smooth functions as well, so it extends beyond traditional calculus methods.

The syntax is that `fminsearch` will try to minimise a function f that you must store as a function in an M-file. Then you give `fminsearch` a starting point and let it run.

For example, to minimise $f(x, y, z) = (x - 1)^2 + (y - 2)^2 + (z - 3)^2$, use the following code:

```
function output = f(x)
output = (x(1)-1)^2 + (x(2)-2)^2 + (x(3)-3)^2;
```

Note that `x(1)` corresponds to x , `x(2)` to y and `x(3)` to z . This is because `fminsearch` requires that f takes a single vector as input (in this example the three-dimensional vector \mathbf{x}).

Now, to run the minimisation, call

```
fminsearch('f', [0,5,10])
```

where $x = 0, y = 5, z = 10$ is the initial guess.

For functions with several local minima, `fminsearch` can get stuck within a local minimum. Starting it with different initial guesses or restarting it when it thinks it has converged may help it to look for “better” minima. How would you use `fminsearch` to find a maximum value of f ?

3.2.5 Main points

- You should know that the local minimum or maximum is obtained when either all the partial derivatives are zero or some of the partial derivatives are not defined.
- You should know how to find critical points and classify them using either the quadratic approximation, or the second derivative test.
- You should be able to find global maxima and minima of a function on a bounded domain.
- You should be competent finding extrema with `fminsearch`.

3.3 Constrained Optimisation and Lagrange Multipliers

3.3.1 Lagrange multipliers

In practical problems one often needs to maximise or minimise a function subject to certain constraints, see also Stewart, Section 15.8.

Example: Find the minimum value of $x^2 + y^2$ subject to the constraint $x + y = 1$.

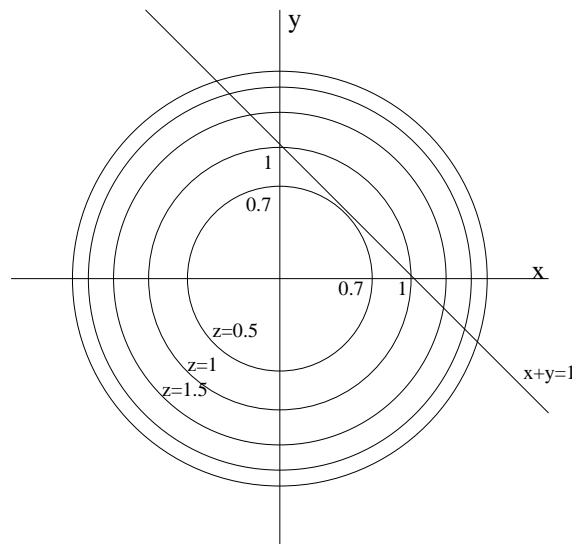
We want to minimise the function f given by $f(x, y) = x^2 + y^2$, subject to the constraint $x + y = 1$.

In this baby-example we can explicitly solve the **constraint-equation**: $y = 1 - x$, so

$$F(x) := f(x, 1 - x) = x^2 + (1 - x)^2 = 2x^2 - 2x + 1,$$

which is a function of x alone. The critical points of F occur when $F'(x) = -2 + 4x = 0$, yielding $x = 1/2$. Since $F''(x) = 4 > 0$ it follows that F has a minimum at $x = 1/2$. Therefore, f subject to the constraint $x + y = 1$ achieves its minimum value at $x = y = 1/2$, with actual value also $1/2$.

There is a much better way to solve the above problem not requiring the explicit solution of the constraint-equation. The trick is to define a second function, say g , such that the constraint-equation corresponds to $g(x, y) = 0$. In our example, $g(x, y) = x + y - 1$. Now graph the contour plot of $f(x, y) = x^2 + y^2$ and add to this graph the single contour $g(x, y) = 0$:



The minimum occurs where the contour $g(x, y) = 0$ touches one of the contours of f , which means that ∇f and ∇g are parallel. Hence

$$\nabla f = \lambda \nabla g$$

for some λ , known as the **Lagrange multiplier**.

Example: Find the minimum value of $x^2 + y^2$ subject to the constraint $x + y = 1$.



This method of Lagrange multipliers can be extended to any number of dimensions.

Example: Minimise $x^2 + y^2 + z^2$ subject to the constraint $x + y + z = 1$.



Important remark: The above problem again illustrates that it pays to think before doing a calculation. Why could we immediately have guessed the answer?

Example: Let $V(x, y, z) = xyz$ describe the volume of a box with vertices at $(0, 0, 0)$, $(x, 0, 0)$, $(0, y, 0)$, \dots , (x, y, z) . Maximise the volume of the box subject to the constraint that the vertex (x, y, z) lies in the plane $z + 2x + 3y = 6$.



Example: The distance between a point and a plane.

We now return to a problem considered Chapter 1 — that of finding the distance between a point and a plane.

Let $P = (x_1, y_1, z_1)$ be a point and $\Pi : ax + by + cz = d$ a plane. We want to compute the distance between P and Π .

If we take an arbitrary point $Q = (x, y, z)$ in \mathbb{R}^3 then the square of the distance between P and Q is given by

$$D^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 =: f(x, y, z).$$

We want to minimise this distance subject to the constraint that Q lies in the plane Π , i.e., subject to the constraint $ax + by + cz = d$. Following the method of Lagrange multipliers we thus define $g(x, y, z) = ax + by + cz - d$ and compute $\nabla f = \lambda \nabla g$ to get

$$\begin{cases} 2(x - x_1) = \lambda a \\ 2(y - y_1) = \lambda b \\ 2(z - z_1) = \lambda c. \end{cases} \quad (*)$$

First we use $(*)$ to find

$$D^2 = f(x, y, z) = (\lambda a/2)^2 + (\lambda b/2)^2 + (\lambda c/2)^2 = \left(\frac{\lambda}{2}\right)^2 (a^2 + b^2 + c^2)$$

from which we still need to eliminate λ . Second we use $(*)$ to find

$$x = \lambda a/2 + x_1, \quad y = \lambda b/2 + y_1, \quad z = \lambda c/2 + z_1.$$

Substituting this into the constraint equation $ax + by + cz - d = 0$ we find that

$$\frac{\lambda}{2} = \frac{d - ax_1 - by_1 - cz_1}{a^2 + b^2 + c^2}.$$

Therefore,

$$D^2 = \left(\frac{\lambda}{2}\right)^2 (a^2 + b^2 + c^2) = \frac{(d - ax_1 - by_1 - cz_1)^2}{a^2 + b^2 + c^2}$$

as before.

Exercise: Use the above equations to determine the point in Π that lies closest to P .

3.3.2 Main points

- You should be able to apply the method of Lagrange multipliers to solve problems in constraint optimization.
- You should understand the method of Lagrange multipliers geometrically in terms of gradient vectors.

4 Ordinary Differential Equations

4.1 Introduction

4.1.1 Main points

In this section we shall cover the following questions and concepts:

- What is a differential equation?
- How do differential equations arise?
- What does a solution to a differential equation mean?
- What is an initial value problem?
- How to solve a differential equation of the form $\frac{dy}{dx} = f(x)$.
- What is meant by the order of a differential equation?

In almost all fields of scientific endeavour, quantifying time-dependent occurrences almost inevitably comes down to a **differential equation** (DE). A DE is an equation which contains one or more derivatives of an unknown function. There are two types of DEs:

- **ordinary** differential equations (ODEs), where the unknown function is a function of only one variable, and
- **partial** differential equations (PDEs) where the unknown function is a function of more than one variable. We will not consider PDEs in this course.

4.1.2 Examples of ODEs

- Unbounded population growth: $P'(t) = kP$.
- Bounded population growth: $P'(t) = kP(1 - P/\theta)$.
- Motion due to gravity: $my''(t) = -mg$.
- Spring system: $mx''(t) = -kx$.

In ODEs, one often takes t instead of x for the independent variable, where t denotes time. Also, derivatives, such as $x'(t)$ and $x''(t)$ with respect to time are often written as \dot{x} and \ddot{x} respectively.

4.1.3 How do ODEs arise?

Population modelling. Assume a population grows at a constant rate proportional to the size of the population. If $P = P(t)$ stands for the population at time t , then the model states that

$$\frac{dP}{dt} \propto P,$$

i.e.,

$$\frac{dP}{dt} = kP,$$

where k is the growth constant. If $k > 0$ the population is growing (think humans) and if $k < 0$ the population is decreasing (think tigers).

Question: What happens for $k = 0$?

Newton's eureka moment. Newton's **second law of motion** states that

$$\text{mass} \times \text{acceleration} = \text{force}$$

Let $y = y(t)$ be the vertical displacement from ground at time t of an apple of mass m , soon to land on Newton's head. Then the acceleration is $\frac{d^2y}{dt^2}$ so that

$$m \frac{d^2y}{dt^2} = -mg,$$

where g is the constant of acceleration at the surface of the earth due to gravity.

4.1.4 Solution to an ODE

Example: Show that $y = y(x) = A \exp(x^2/2)$ is a **solution** to the ODE $y' = xy$.



When asked to **solve** an ODE, you are expected to find *all* possible solutions.

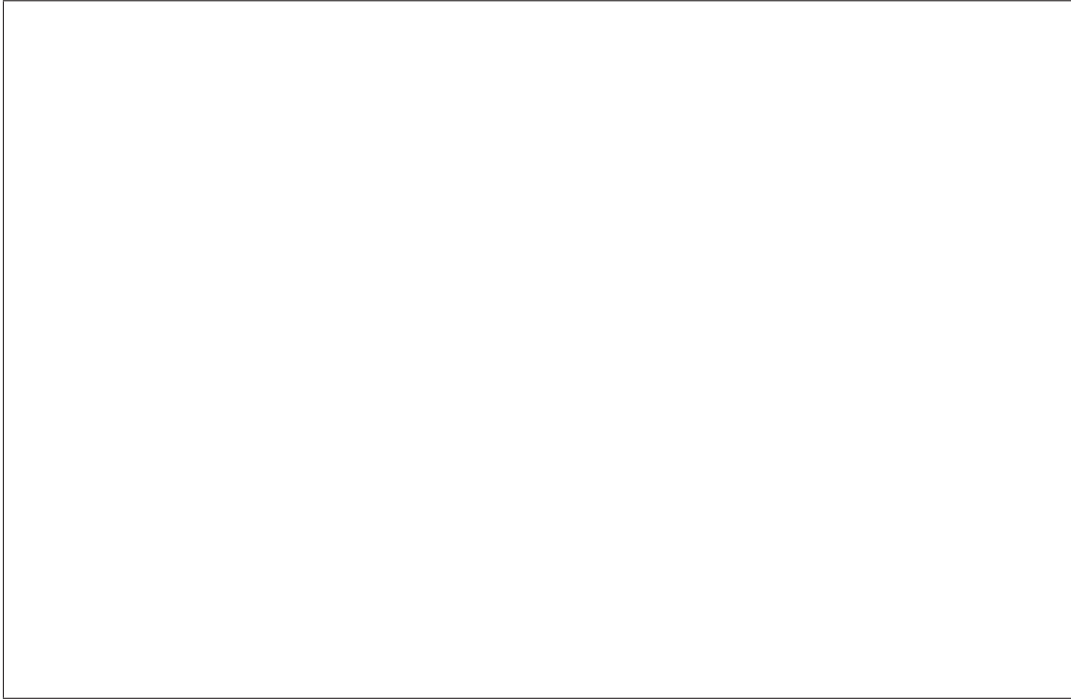
You should already know how to solve ODEs of the form

$$\frac{dy}{dx} = f(x) \quad \text{or} \quad \frac{d^2y}{dx^2} = g(x).$$

Question: What can we do with the constants of integration?

4.1.5 Initial value problem: fluid flow

A flow-meter in a pipeline measures flow-through as $2 + \sin t$ litres/second. How much fluid passes through the pipeline from time zero up to time T ?



The problem of solving an ODE subject to some initial conditions of the form $y(t_0) = y_0$, $y'(t_0) = z_0$, etc., is called an **Initial Value Problem** (IVP).

4.1.6 The order of an ODE

The order of the highest-order derivative in an ODE defines the **order** of the ODE.

- $\frac{dP}{dt} = kP$
- $m\ddot{x} = -kx$
- $x(y'')^2 + y'y''' + 4y^5 = yy'$
- $y' = xy$

4.1.7 Dr. VO's bike at constant speed

Find the position of Dr. VO's bike (at time t) travelling at a constant speed of 60 km/h along a perfectly straight downhill road.

You won't need any knowledge of ODEs to solve this simple problem, but let us ignore this for the moment. This example only serves to illustrate some terminology, and to promote cycling.

If $x = x(t)$ is the distance travelled by Dr. VO at time t then the corresponding ODE is

$$\frac{dx}{dt} = 60. \quad (*)$$

We can directly integrate to find $x(t)$:

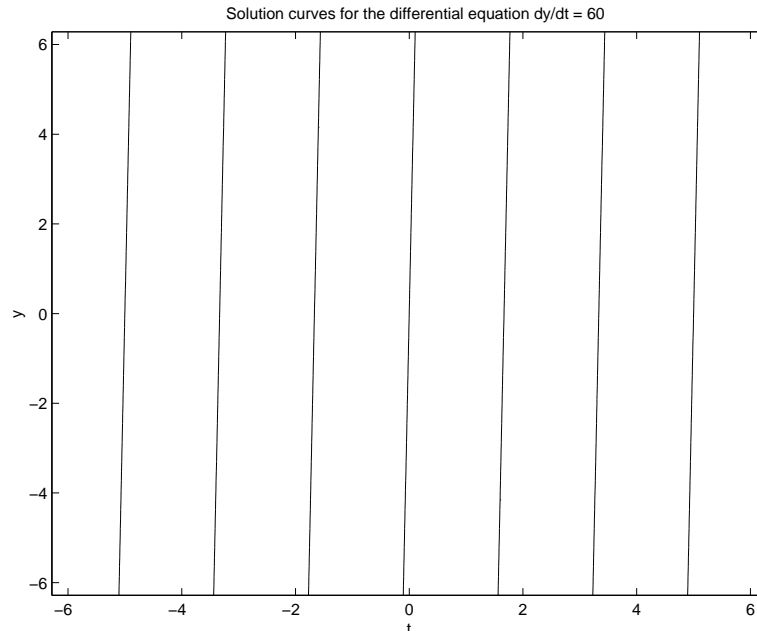
$$\frac{dx}{dt} = 60 \quad \Rightarrow \quad \int dx = \int 60 dt.$$

This of course yields the infinite family of solutions

$$x(t) = 60t + C.$$

To determine the constant C we need more information, such as Dr. VO's initial position at time 0, in which case we would be solving an IVP.

For different values of C we get different solutions, and below we graph some of these. If $C = 0$ then $y = 60t$. All the other solutions are parallel to this line.



The curves $y = 60t + C$ are called **solution curves** to the ODE (*). Note that in this particular case all curves are straight lines with slope 60.

4.2 Motion of Projectiles

4.2.1 Main points

At the end of this section you should be able to:

- Derive, using Newton's second law, equations of motion in a force field (such as gravity).
- Solve these equations in certain simple cases.
- Understand the difference between an analytical and numerical solution.

4.2.2 Gravity in one dimension

Consider an apple falling under gravity. Find an expression for the height y of the apple.



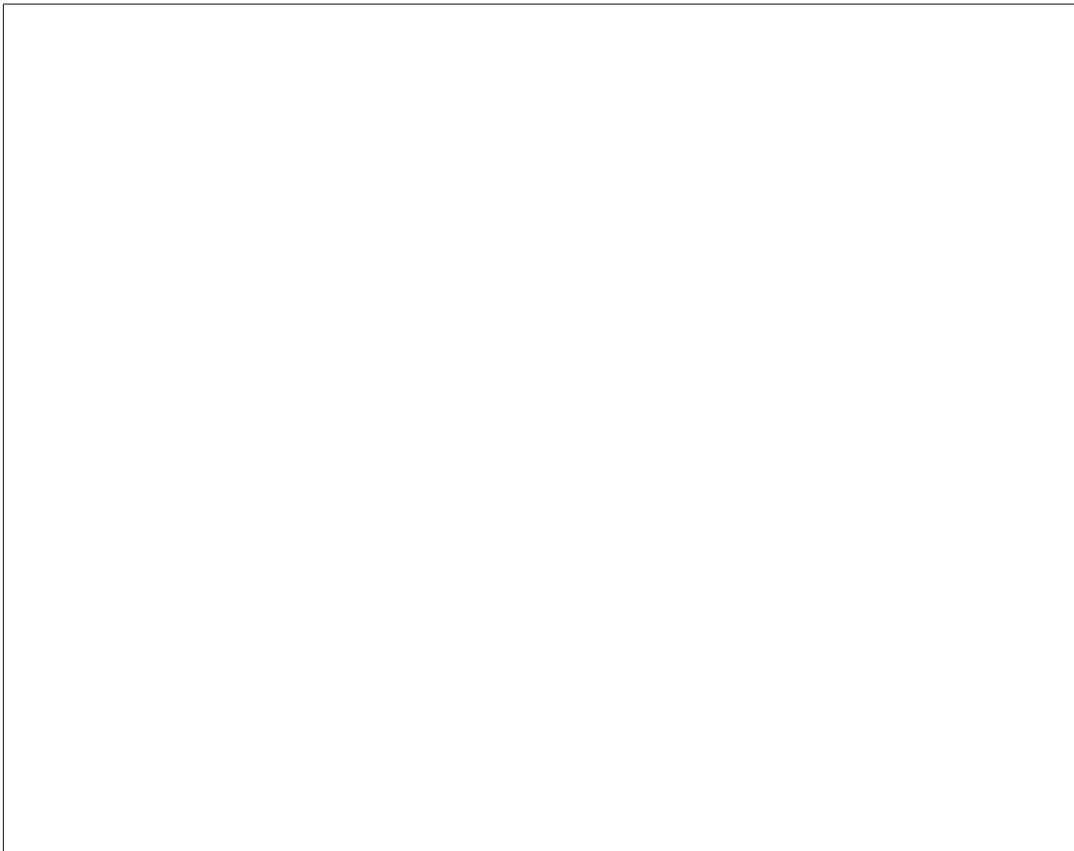
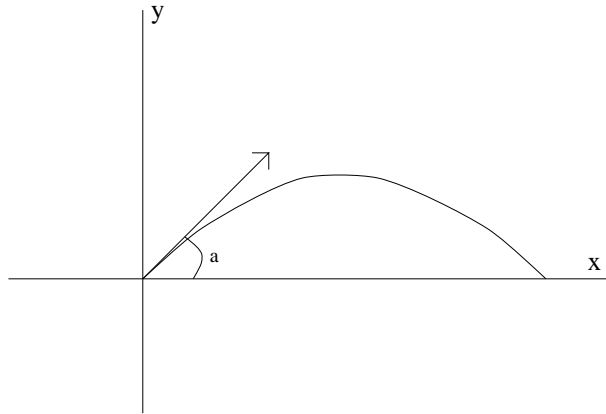
4.2.3 Apples in two dimensions

Now suppose you are throwing apples over your fence to your creepy neighbour with the aim of hitting him on the head. You may assume your neighbour is sunbathing (so his head is at ground-level) 15 metres away from where you are standing.

Find an expression for the position $(x(t), y(t))$ of the apple.

4.2.4 Example: hitting a target

- (a) If you throw at 20m/s, at what angle do you need to throw in order to hit a neighbour 15 metres away?
- (b) What is the equation to the curve that the apple traces in space?





4.2.5 Realistic models

More realistic models of projectile motion deal with air resistance.

- For fast moving objects (e.g., Dr. (The Flash) Murrance on his bike) a good model includes air resistance proportional to square of the velocity

$$\frac{d^2y}{dt^2} = -g - f \frac{dy}{dt} \left| \frac{dy}{dt} \right|.$$

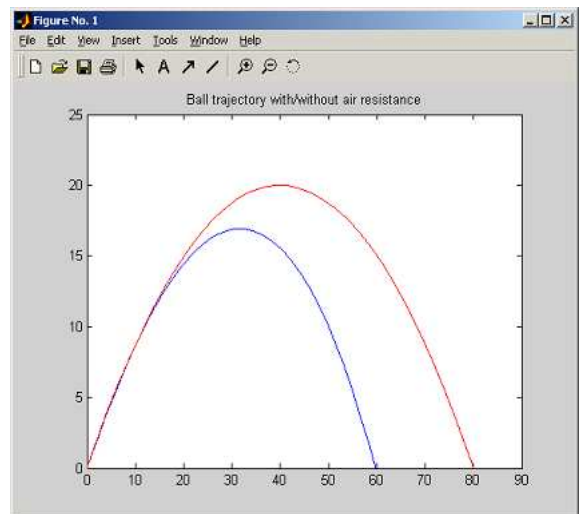
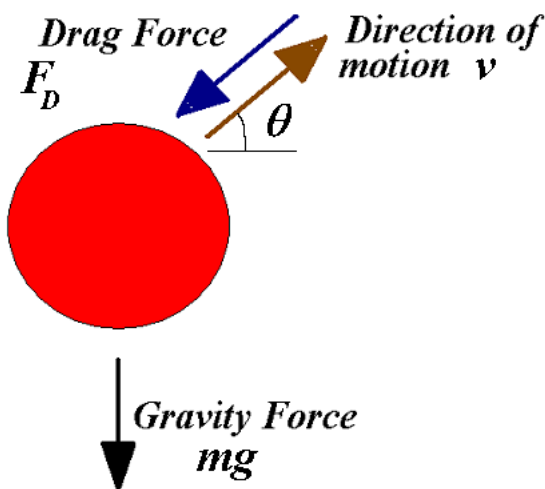
- For slower objects (throwing apples), a good enough model has air resistance proportional to velocity.

Realistic models may also include the fact that gravity diminishes as you move away from the earth's surface (Newton's inverse square law).

4.2.6 A model with air resistance

Consider the motion of a ball subject to air resistance proportional to the *square* of the speed. Note that the air resistance vector is always directed *against* the direction of motion.

Decompose the coordinates into x and y directions, and again apply Newton's law $F = ma$.



Note that

$$\sin \theta = \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} \quad \text{and} \quad \cos \theta = \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}}.$$

Hence $F = ma$ in the y direction gives

$$\begin{aligned} m\ddot{y} &= -mg - F_D \sin \theta \\ \Rightarrow m\ddot{y} &= -mg - kv^2 \sin \theta \\ \Rightarrow m\ddot{y} &= -mg - k(\dot{x}^2 + \dot{y}^2) \times \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} \\ \Rightarrow \ddot{y} &= -g - \frac{k}{m} \dot{y} \sqrt{\dot{x}^2 + \dot{y}^2}. \end{aligned}$$

In much the same way

$$\ddot{x} = -\frac{k}{m} \dot{x} \sqrt{\dot{x}^2 + \dot{y}^2}.$$

This is a coupled system of ODEs which is extremely difficult to solve analytically. However, numerical solutions are easy to obtain, and you can undertake this a little later in the semester. A sample trajectory in the xy -plane is shown on the previous page.

Definitions

- To solve an ODE (or IVP) **analytically** means to give a solution curve in terms of continuous functions defined over a specified interval, where the solution is obtained exactly by analytic means (e.g., by integration). The solution satisfies the ODE (and initial conditions) on direct substitution.
- To solve an ODE (or IVP) **numerically** means to use an algorithm to generate a sequence of points which approximates a solution curve.

Important remark: As we have already seen, the ODE

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

can be solved analytically. The **solution** simply is

$$y = \int f(t) dt.$$

Now this may seem like a cop-out because all we are saying is that the solution is given by the anti-derivative of $f(t)$, and the above solution is as informative as the actual ODE. In practice one hopes to be able perform the above integral to get a more explicit form of the solution. Depending on the form of f this may either be done exactly or by numerical means.

4.3 Slope Fields and Equilibrium Solutions

4.3.1 Main points

By the end of this section, you should understand the following concepts:

- How to generate and interpret slope fields.
- What is meant by an equilibrium solution and how to find one.
- How to determine the stability of an equilibrium solution.
- How uniqueness of solutions to an IVP implies that solutions cannot cross.

We have seen that in order to solve

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

we only need to integrate. However, for the general first order ODE

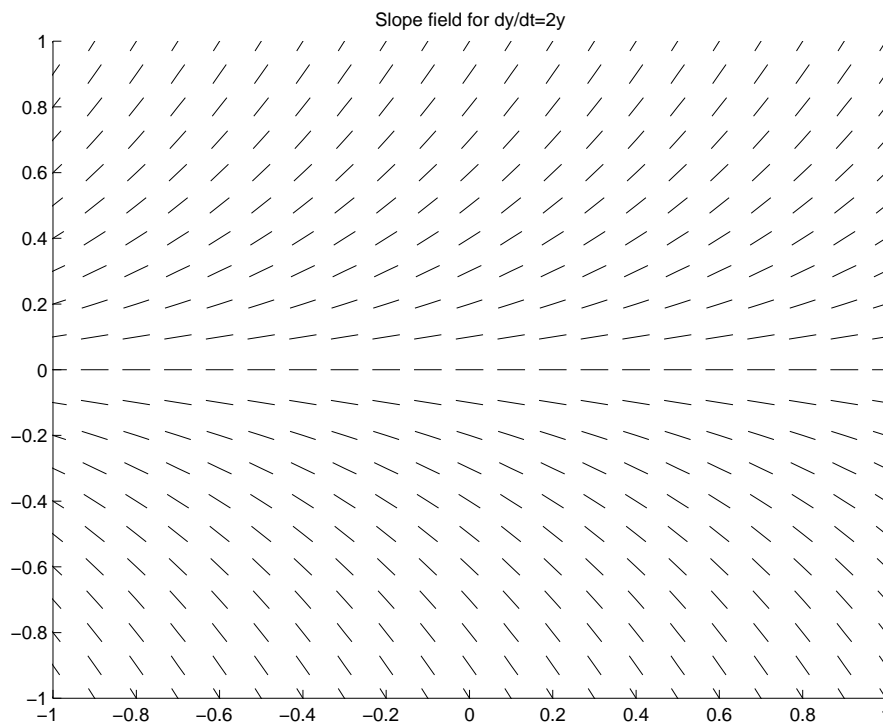
$$\frac{dy}{dt} = f(t, y)$$

this no longer works. Nonetheless, the ODE gives a qualitative picture of the solution by noting that at $(y, t) = (a, b)$ the slope of $y(t)$ is $f(a, b)$. So what one can do is as follows:

- In the ty -plane, at $(t, y) = (a, b)$ draw a small straight line with slope $f(a, b)$.
- Repeat the process for many different values of (a, b) .
- The resulting diagram is called the **slope field**.

Note the slope field can be generated without having to solve the ODE.

4.3.2 Example: the slope field of $y' = 2y$



From the slope field of $y' = 2y$ we can see that $y = 0$ is one of the solution curves. It is a constant or **equilibrium solution**.

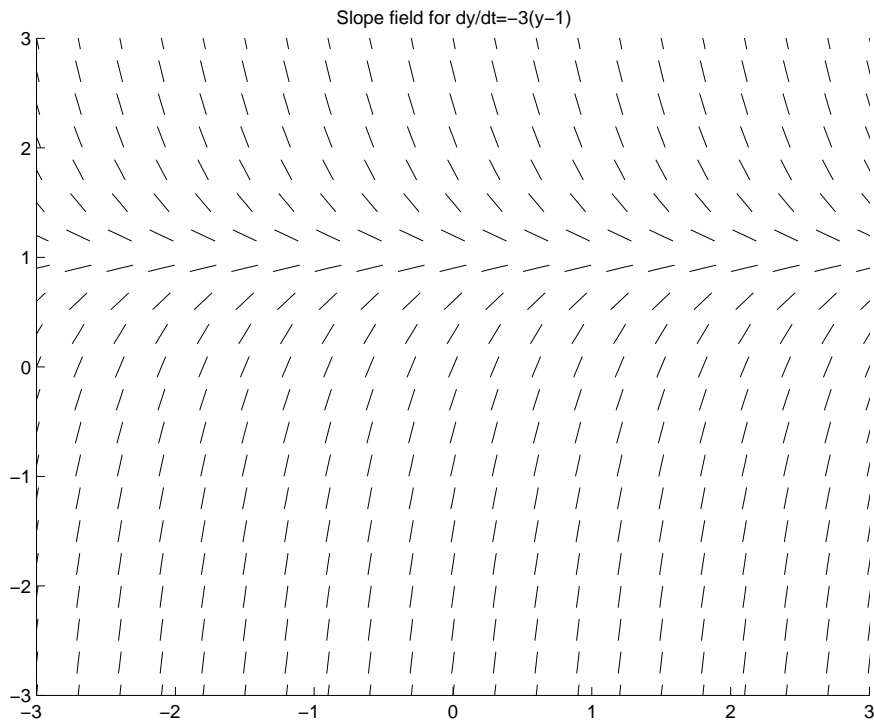
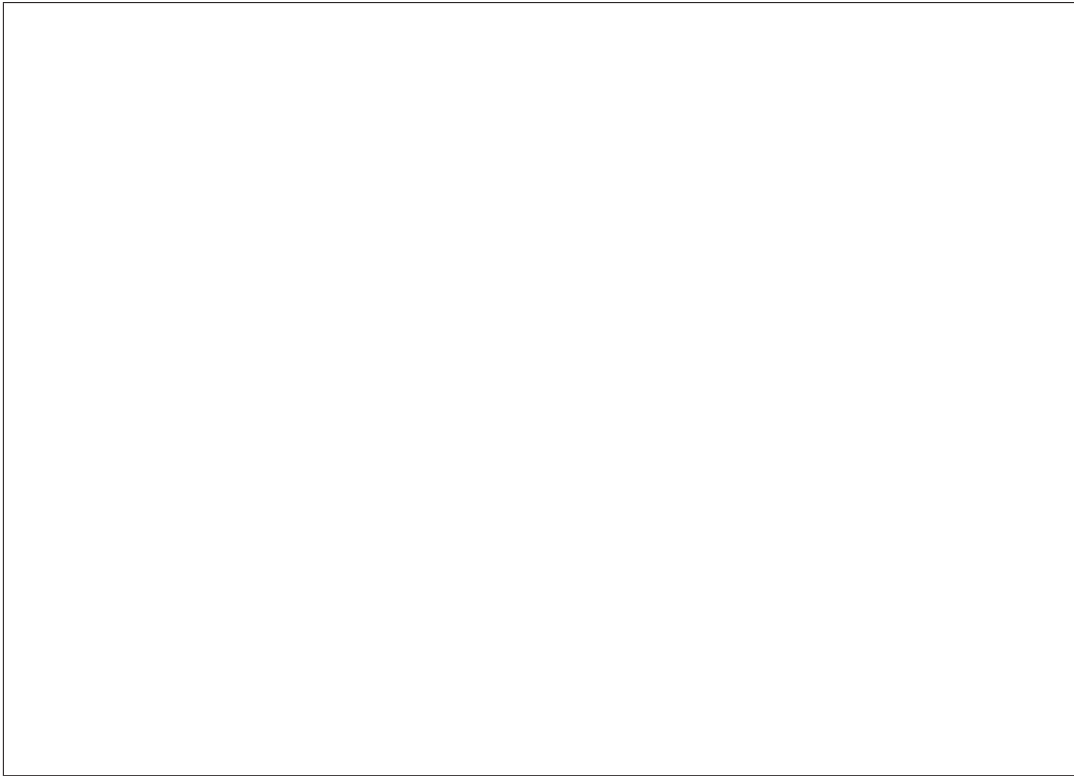
4.3.3 Equilibrium solutions

An **equilibrium solution** is a constant solution to the ODE

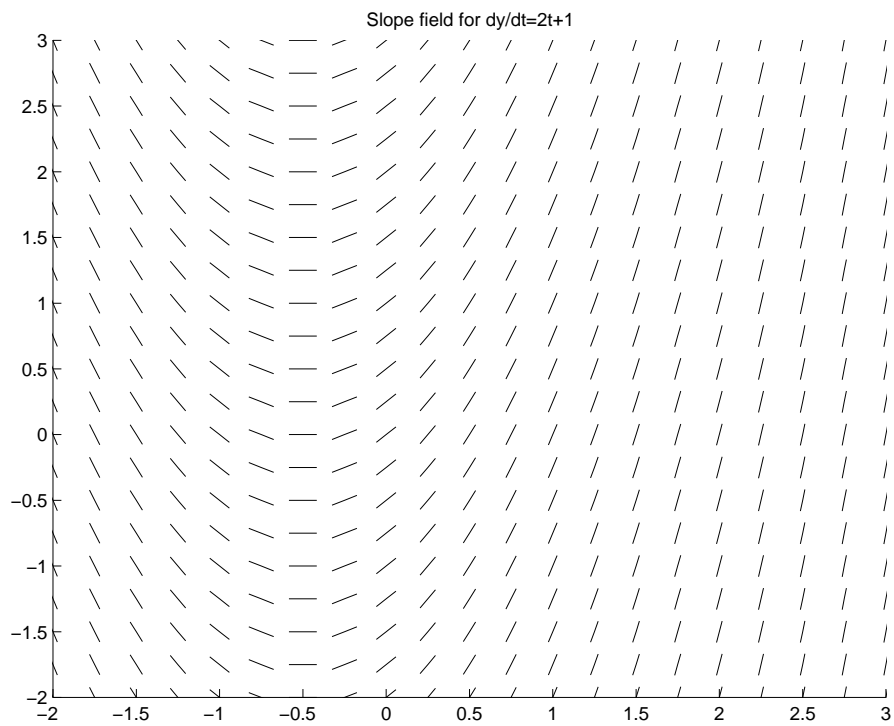
$$\frac{dy}{dt} = f(t, y).$$

The graph of an equilibrium solution is a horizontal line, so the slope is zero, i.e., $y' = 0$. This can only happen if $f(t, y) = 0$ has a solution.

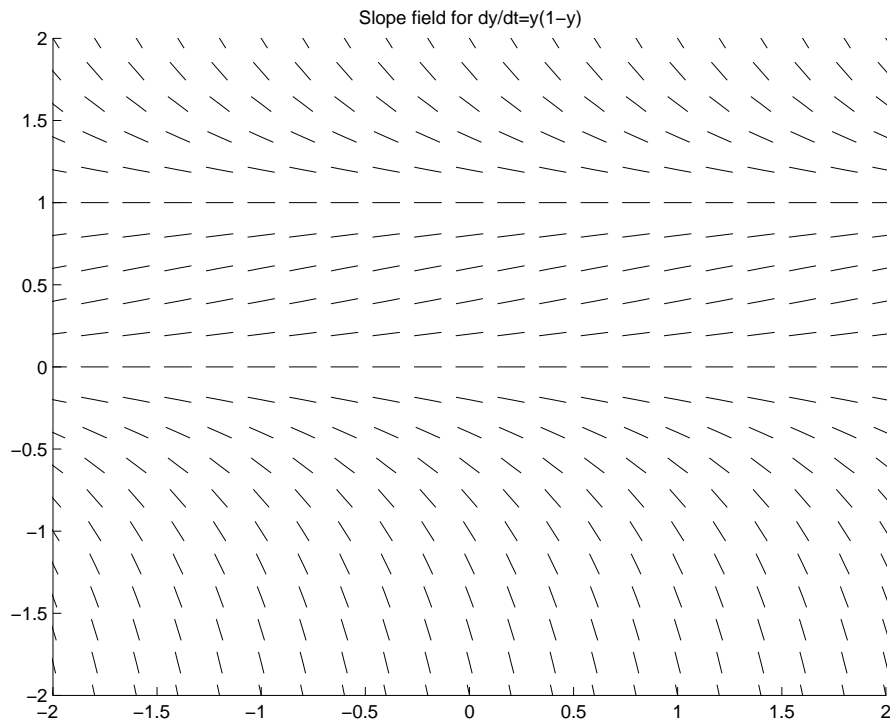
Example: Find the equilibrium solutions of $y'(t) = -3(y - 1)$.



Example: Find the equilibrium solutions of $y'(t) = 2t + 1$.



Example: Find the equilibrium solutions of $y' = y(1 - y)$.



4.3.4 Stability of equilibrium solutions

A pencil sitting balanced vertically is in an equilibrium state. But make one small perturbation and it will topple over. This is an **unstable** equilibrium. On the other hand, a pendulum hanging vertically is also in an equilibrium state. But if you perturb it slightly, it will eventually (with friction) return to its equilibrium. This is a **stable** equilibrium.

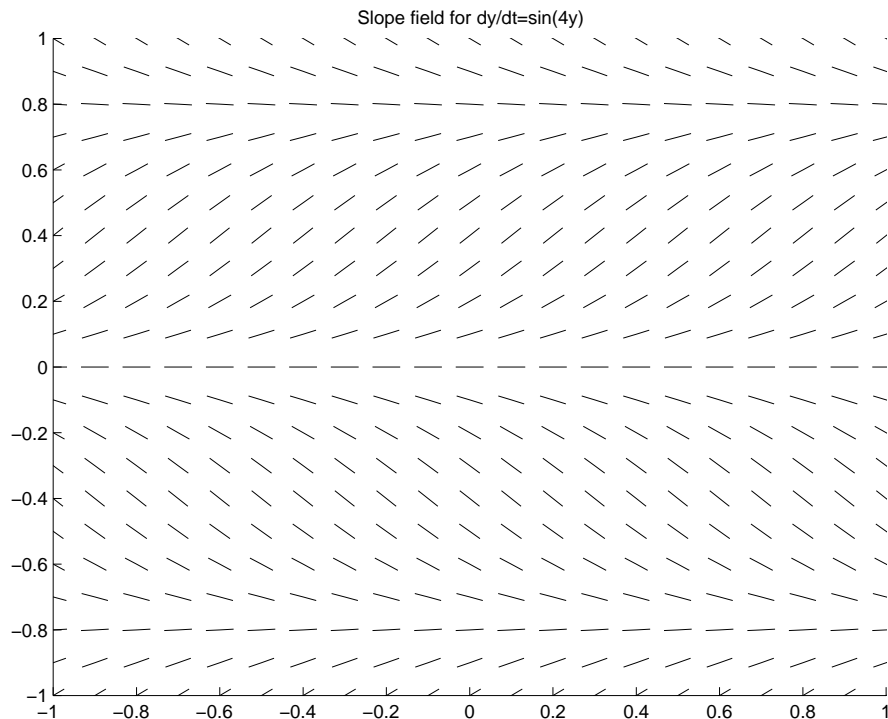
From the slope field, you can decide if an equilibrium solution is stable or not. Formally, an equilibrium solution $y(t) = y_0$ to the differential equation $y' = f(t, y)$ is stable if the initial value problem:

$$\frac{dy}{dt} = f(t, y), \quad y(0) = y_0 \pm \epsilon$$

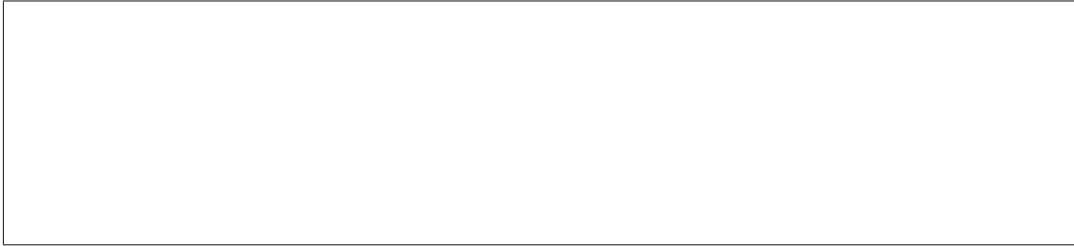
has a solution $y(t)$ which satisfies

$$\lim_{t \rightarrow \infty} y(t) = y_0.$$

In other words, if you start sufficiently close to a stable equilibrium solution, then you will approach that equilibrium solution.



Example: In the previous diagram, which equilibrium solutions are stable and which are unstable?

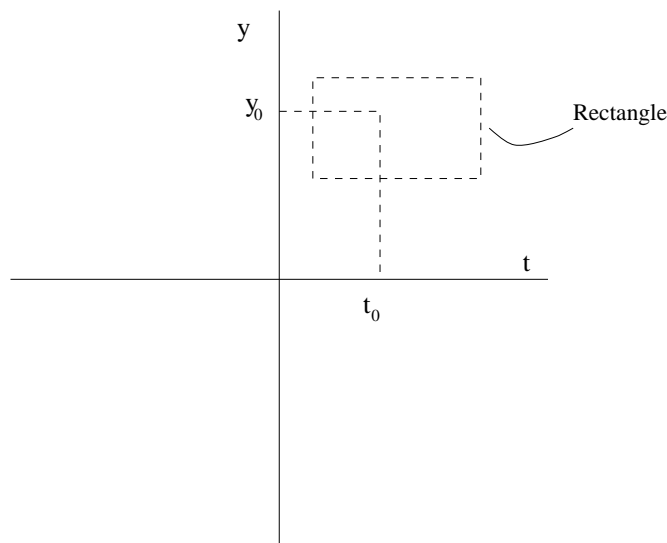


4.3.5 Existence and uniqueness of solutions

Consider the IVP

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0.$$

Then, if $f(t, y)$ is “smooth” in some rectangle about (t_0, y_0) , there exists a **unique** solution $y = y_1(t)$ at least in some small neighbourhood of (t_0, y_0) .



It is an important point that **equilibrium solutions cannot be crossed** (by other solution curves). In fact, no solutions curves can cross each other because this would mean that in some point(s) $y'(t)$ has more than one value. Equilibrium solutions therefore **partition** the solution space.

4.3.6 Long-term behaviour

True or False: Assume the ODE

$$\frac{dy}{dt} = -3(y - 1)$$

has initial condition $y(t_0) = c$ with $c < 1$. Then $y(t) < 1$ for all $t > t_0$.



Example: The only equilibrium solution of $y' = y^2$ is $y = 0$.

For $y > 0$, $y' > 0$ so that $y(t)$ is increasing and hence any solution with $y > 0$ tends away from $y = 0$.

For $y < 0$, $y' > 0$ so that $y(t)$ is increasing and hence any solution with $y < 0$ approaches $y = 0$.

Therefore $y = 0$ is neither stable nor unstable.

4.4 Euler's Method for Solving ODEs Numerically

4.4.1 Main points

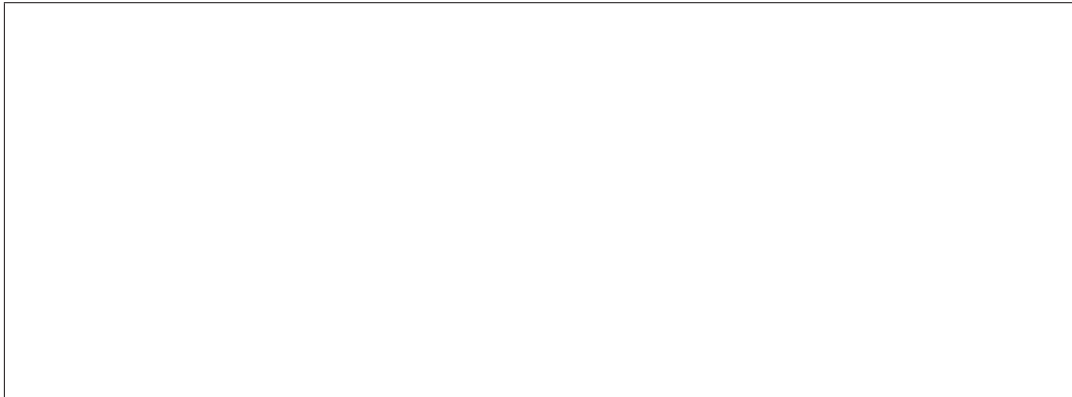
At the end of this section you should know:

- How slope fields lead to Euler's method.
- How to approximate a solution to an IVP using Euler's method by hand.
- How to implement the Euler method in Matlab.
- How errors can creep in to the approximate solution.

Stewart, Section 10.2 covers the Euler method. The method gives a simple approximate solution to an ODE and is closely related to the notion of a slope field.

4.4.2 Revision: equation of a straight line

What is the equation of the straight line with slope m that passes through the point (a, b) ?



Euler's method uses tangent lines as approximations to solution curves. The tangent line to a solution curve of $y' = f(t, y)$ at (t_0, y_0) is

$$y = y_0 + f(t_0, y_0)(t - t_0)$$

which approximates the curve when t is close to t_0 .

Now imagine a family of solution curves to the differential equation. We can calculate an approximate value for y at some later time by taking lots of small steps in time. At each step we will use the tangent line to the solution through that point.

Let

$$t_1 = t_0 + \Delta t, \quad t_2 = t_1 + \Delta t, \quad t_3 = t_2 + \Delta t, \quad \dots, \quad t_n = t_{n-1} + \Delta t.$$

To find approximate y values at these times, use the following sequentially

$$y_1 = y_0 + f(t_0, y_0)\Delta t$$

$$y_2 = y_1 + f(t_1, y_1)\Delta t$$

$$y_3 = y_2 + f(t_2, y_2)\Delta t$$

and so on.

Example: Use Euler's method with $\Delta t = 0.2$, to find an approximate solution to $y(0.6)$ for the IVP

$$\frac{dy}{dt} = 2t, \quad y(0) = 0.$$

Compare your answer with the actual value.

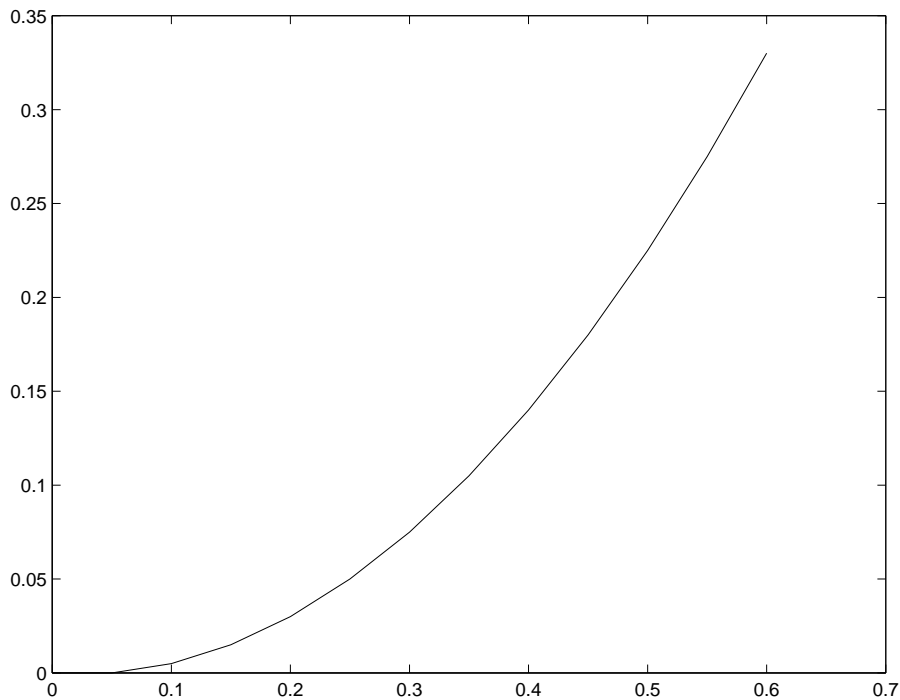
The method is only accurate if you make Δt small, which means a large number of steps is usually required. For this you might want to use Matlab. Define a t vector and an initial y value, then use the **for** command to iterate. In the example $\Delta t = 0.05$ and 12 steps have been chosen to take t to 0.6.

```
t=(0:0.05:0.6);
y(1)=0;
for i=1:12
    y(i+1)=y(i)+2*t(i)*0.05;
end
y(13)
plot (t,y,'-')
```

Using Matlab we find $y(0.6) = 0.33$ which is much much better than the previous estimate value for $y(0.6)$.

Note that we start with $y(1) = 0$, whereas in the theoretical work we write $y(0) = 0$. Matlab does not allow you to put an index of zero. Keep this in mind in case you ever get the Matlab error message:

??? Index into matrix is negative or zero.



4.4.3 Euler's method using Matlab; an example

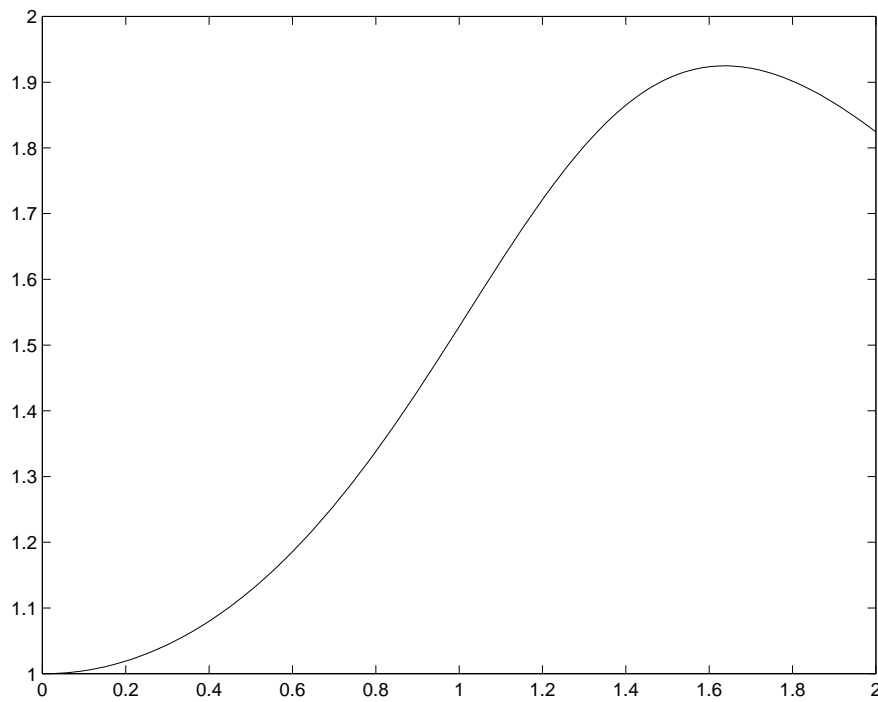
Consider the IVP

$$\frac{dy}{dx} = \sin(xy), \quad y(0) = 1.$$

Matlab code to estimate $y(2)$, using Euler's method with step size $\Delta x = 0.01$ is:

```
x=(0:0.01:2);  
y(1)=1;  
for i=1:200  
    y(i+1)=y(i)+sin(x(i)*y(i))*0.01;  
end  
y(201)  
plot (x,y, '-')
```

Using Matlab we find $y(2) = 1.8243$.

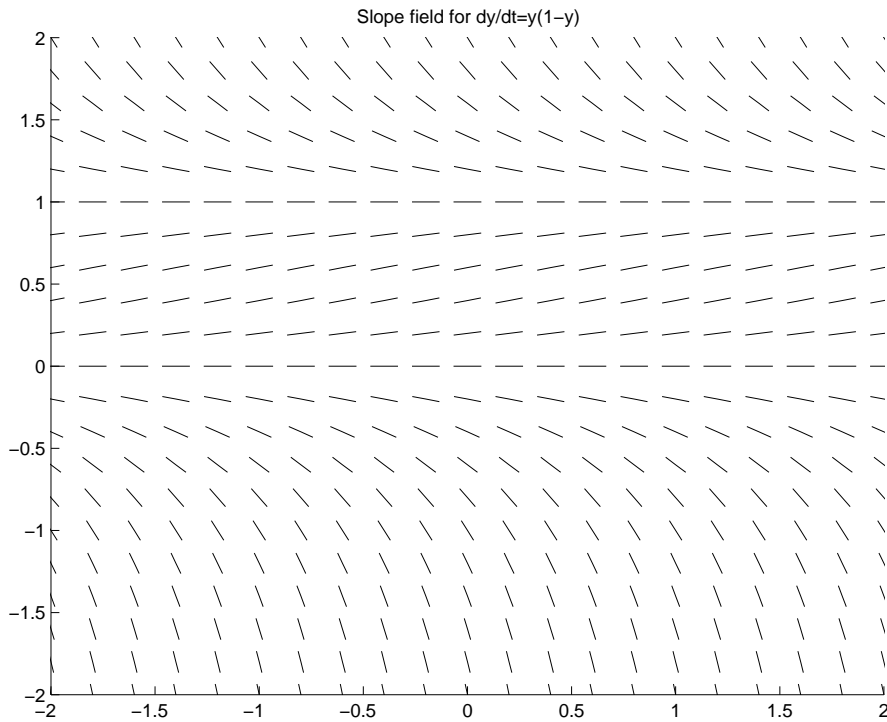


4.4.4 Demonstrating errors in Euler's method

Use Euler's method to approximate the solution curve of the initial value problem

$$y' = y(1 - y), \quad y(-2) = 2.$$

Use the slope field below as a guide to draw in the approximate solution curve. What goes wrong in this case?



4.5 Separable ODEs

By the end of this section, you should understand the following concepts:

- How to identify a first order separable ODE.
- How to solve a separable ODE.
- The difference between a general and special solution.

Separable first order ODEs are one of several classes of ODE we will study in MATH1052. It is very important that you become skilled at identifying and solving this type of ODE.

4.5.1 Definition

A first order ODE is called **separable** if it can be written in the form

$$\frac{dy}{dx} = f(x)g(y).$$

Which of the following ODEs are separable?

- $\frac{dy}{dt} = y(1 - y),$
- $\frac{dy}{dx} = e^{x+y},$
- $y' = e^{(x+y)^2},$
- $\dot{y} = \frac{ty + y}{t^2},$
- $\frac{dy}{dt} = ty + y^2.$

4.5.2 Solving separable ODEs

1. Rewrite the equation as $\frac{1}{g(y)} \frac{dy}{dx} = f(x)$.
2. Integrate both sides with respect to x : $\int \frac{1}{g(y)} \frac{dy}{dx} dx = \int f(x) dx$.
3. Note the integral on the left is a “substitution”, so that we can replace the last equation by $\int \frac{dy}{g(y)} = \int f(x) dx$.
4. If one is lucky (or very smart, or both) one or both of the integrals can actually be evaluated.
5. If one is even more lucky (or even smarter) one can then finally explicitly express y as a function of x .

Important remark: The final two steps will not always be possible. Whenever you are asked to solve a separable ODE (IVP) in this course you are supposed to always “go as far as possible”. For good reasons we assume you are infinitely smart; not being able to compute simple integrals is never a valid excuse.

Example: Solve the ODE $\frac{dy}{dx} = \frac{x}{y}$.



The solution we found for the previous problem contains a constant C and is therefore known as **the general solution**.

In the case of an IVP this constant will be fixed and we find what is known as a particular solution

Example: Solve the IVP $\frac{dy}{dx} = \frac{x}{y}$, $y(0) = 3$.



Example: Solve the IVP $\frac{dy}{dx} = \frac{\sin x}{y}$, $y(0) = 1$.



Important remark: In the two previous examples of an IVP we could first solve the ODE by explicitly evaluating both integrals and by then fixing the constant using the initial value. If however one or both of the integrals cannot be evaluated we can still write down an actual solution, be it a rather implicit one. Indeed, the solution to the IVP $y'(x) = f(x)g(y)$, $y(x_0) = y_0$ is given by

$$\int_{y_0}^y \frac{dt}{g(t)} = \int_{x_0}^x f(t) dt.$$

Example: Again solve the IVP $\frac{dy}{dx} = \frac{\sin x}{y}$, $y(0) = 1$, this time using the above formula.



4.5.3 Singular solutions

Hopefully you by now have complained to your lecturer for committing acts of gross indecency. After all, in step 1 of our recipe for solving separable ODEs they rewrote $y'(x) = f(x)g(y)$ as

$$\frac{1}{g(y)} \frac{dy}{dx} = f(x),$$

which is highly illegal if $g(y) = 0$ for some y .

So our method is only valid provided $y = y(x) \neq 0$.

If there is an a such that $g(a) = 0$ then the ODE will have the equilibrium solution $y(x) = a$. (Check!) Such a solution is also known as a **singular solution** because it won't generally arise from our earlier recipe. Whenever you are asked to solve a separable ODE you will have to check for singular/equilibrium solutions.

Example: Solve $\frac{dy}{dx} = y(y + 1)$.



Example: Solve $\frac{dy}{dx} = y^2$.



Example: Solve the ODE $y'(x) = -x/y$ and sketch the solution curves.



4.6 Applications, Heat Equation, Population Growth

By the end of this section, you should know:

- How to apply Newton's law of cooling and solve the resulting IVP.
- How to model population-growth and solve the resulting IVP.
- How to interpret the obtained solutions.

Here, and in Section 10.4 of Stewart, some simple exponential models and extensions of these are presented.

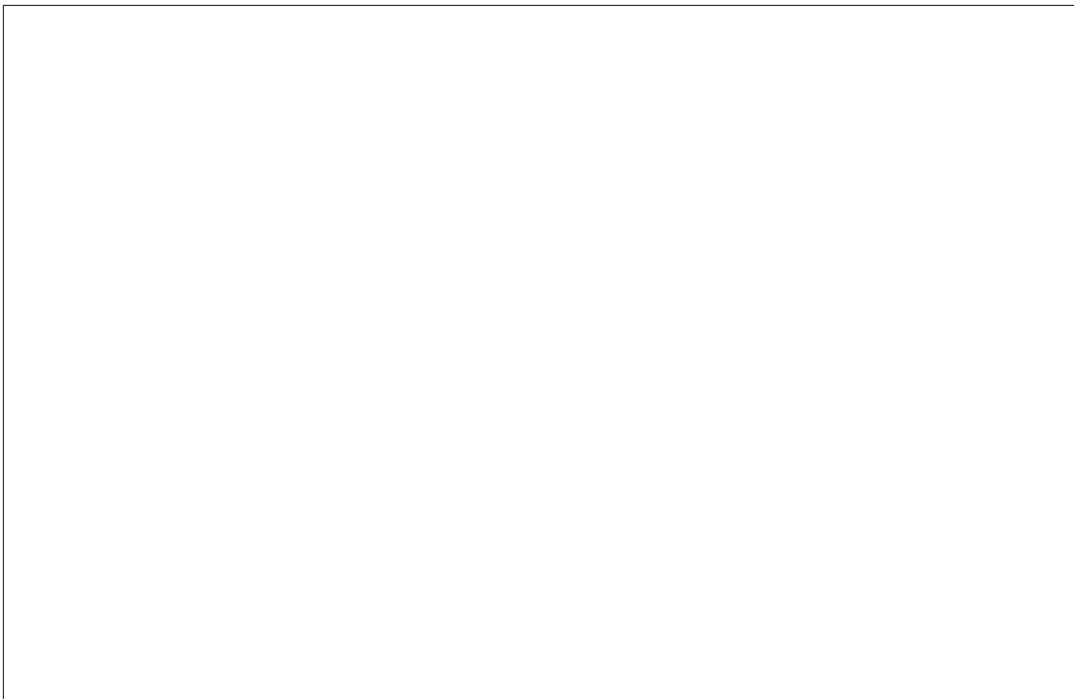
4.6.1 Newton's law of cooling

Newton's law of cooling states that the rate at which a "body" cools is proportional to the temperature difference between the body and its surrounding medium.

If T is the temperature of the body and T_m the temperature of the surrounding medium then, according to Newton,

$$\frac{dT}{dt} = -k(T - T_m), \quad k > 0.$$

Here the constant k is chosen such that if $T > T_m$, $T'(t)$ is negative, describing cooling.

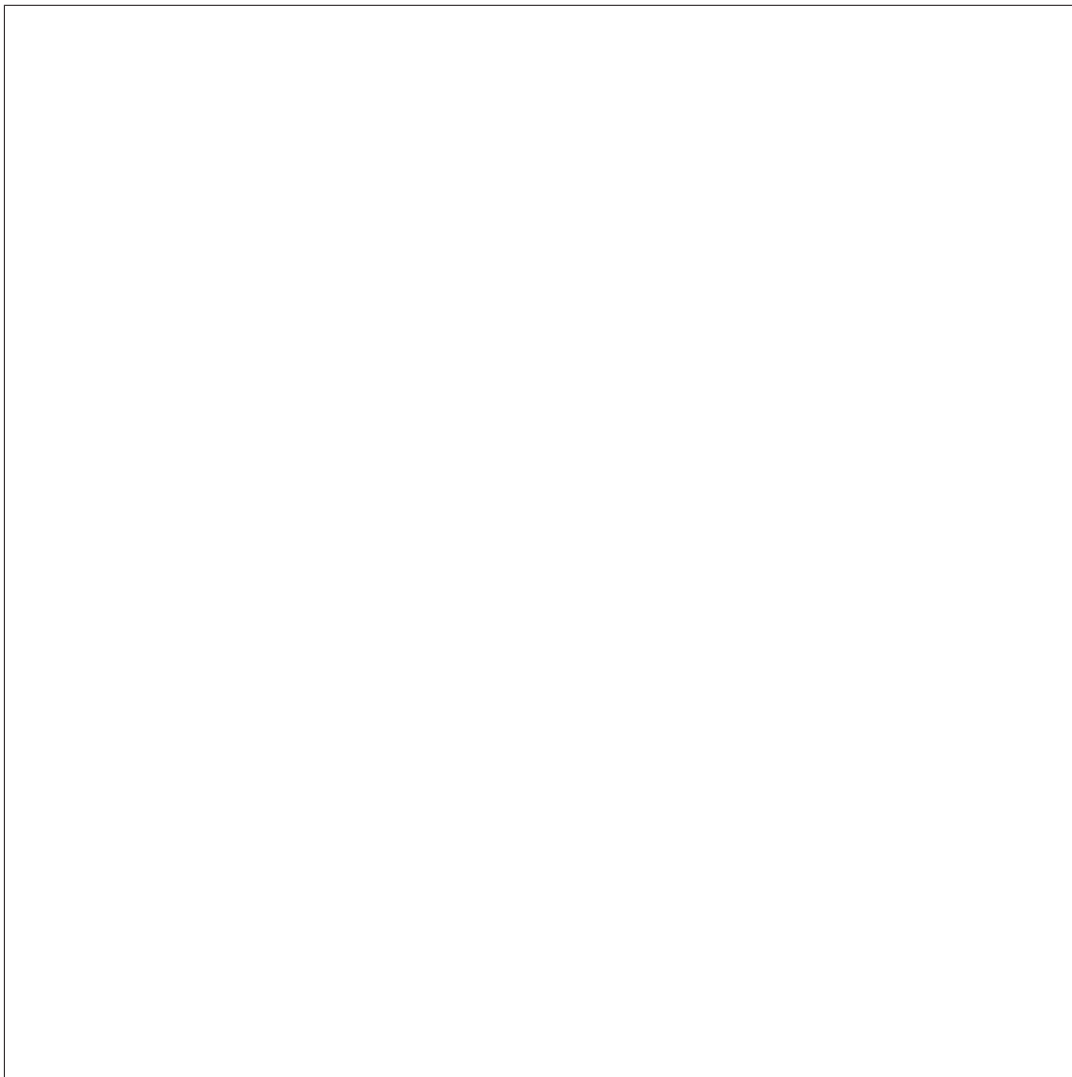


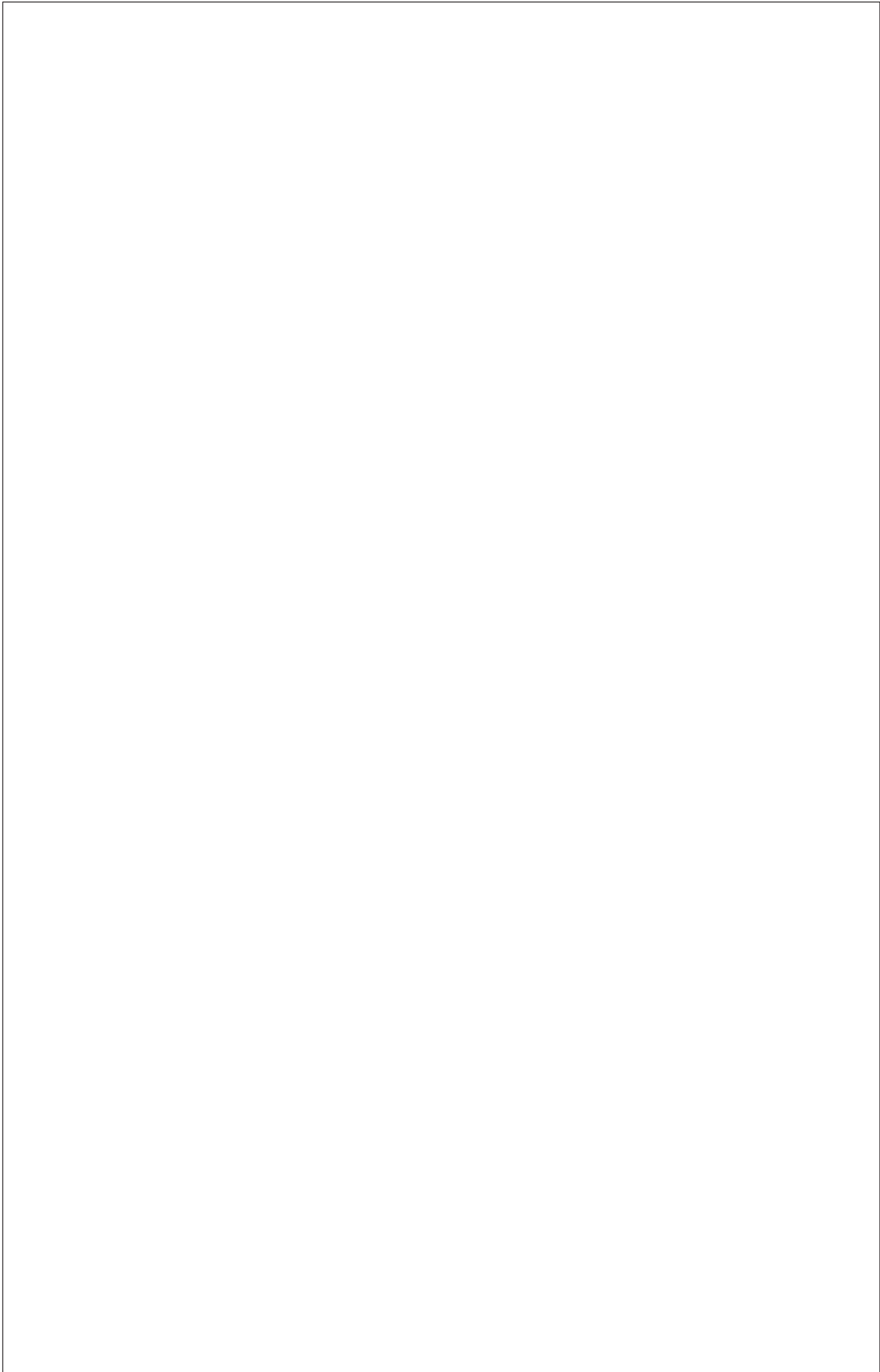
In real-life problems the cooling constant k is usually not known and you will have to infer it from measurements.

Example: CSI Victoria Carl W.'s body was discovered in his cell at Barwon Prison, Victoria on the 19th of April 2010. Dr. Murrance received a leaked copy of the coroner's report stating that:

- C.W.'s body discovered at 11am on 19 April 2010.
- Body temperature at 11am: 34.8°C .
- Body temperature at 12.30pm: 34.1°C .
- Temperature in C.W.'s cell: 21.1°C .

Use the above information to determine the time of Carl's death, assuming his body temperature at the time of death was 37°C .





4.6.2 Malthus: "...doomed to misery and vice"

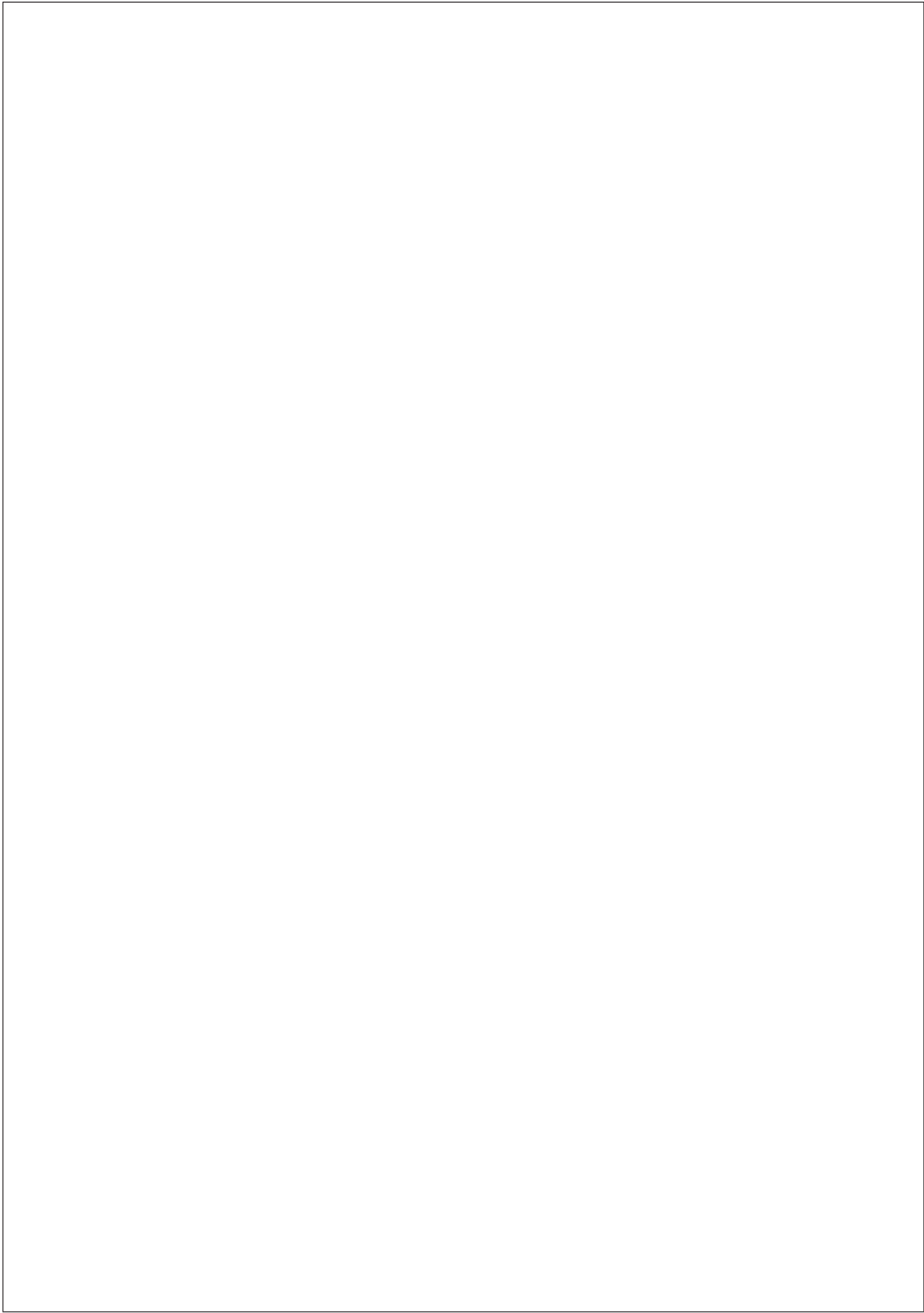
Without inhibiting factors the rate at which the human population grows is proportional to its existing population; if P is the population at time t then

$$\frac{dP}{dt} = rP,$$

where r is the growth rate (per human per year, say).

The world population is currently growing at a rate of approximately $r = 1.25$. If we take as initial condition the year 2000 with roughly 6 billion people on the planet, and ignore global warming so that we may take the total area of land on Earth to be constant at 150 million square kilometres, find in which year each human will have, on average, one square metre of land left to stand on.





4.6.3 Verhulst to the rescue; down with the Malthusian law

The problem with the Malthusian population model is that it does not contain a “damping” factor reflecting issues such as over-population, limited food or water supply, etc. The **logistic model** of Verhulst is the simplest model that has such built-in damping:

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{\theta}\right),$$

where θ is some very large constant, known as the **carrying capacity**.

Note that when P is small(ish) then P/θ is small and we approximately have

$$\frac{dP}{dt} \approx rP$$

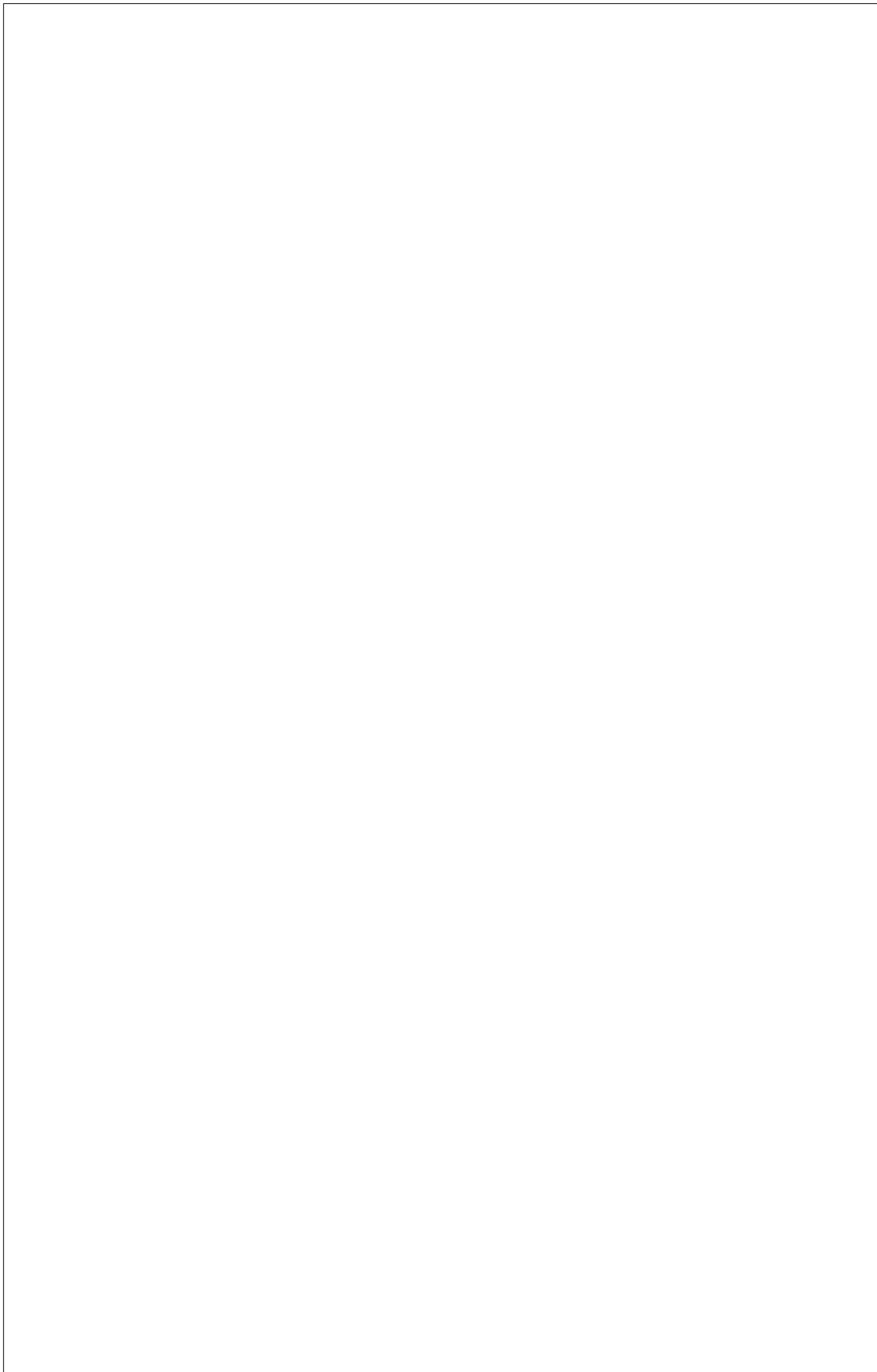
as before. When the population becomes very large (i.e., P/θ very close to 1 so that not just $1 - P/\theta \approx 0$ but also $P(1 - P/\theta) \approx 0$) then growths comes almost to a standstill.

Example: Identify the singular (or equilibrium) solutions to the logistic model, and interpret these solutions in terms of population growth.



Example: Introduce new variables $x = P/\theta$ and $\tau = rt$ in order to solve the logistic equation. Interpret the solutions corresponding to $x < 0$ and $x > 1$.





Example: The Pacific Halibut Fishery uses a logistic model plus an extra term to take into account their harvesting, which is at a rate proportional to the existing population.

If E is the constant of proportionality for harvesting, their model is

$$\frac{dP}{dt} = rP\left(1 - \frac{P}{\theta}\right) - EP.$$

What restriction must be put on E so that the population will not die out?

4.6.4 Extra reading: numerically solving ODEs

Matlab has some more sophisticated numerical methods than Euler's. Matlab's `ode45` uses a Runge–Kutta fourth-order integration technique. To use it you need to write two little programs. The first one, say `logistic.m`, defines the differential equation

```
function dy=logistic(t,y)
dy=y*(1-y)
```

The second program will numerically solve the ODE and plot your solution.

```
y0=0.1;
tfinal =6;
[t,y]= ode45 ('logistic', tfinal, y0);
plot (t,y)
```


4.7 Solving Linear First Order Equations

At the end of this section you should understand:

- How to identify a first order linear ODE.
- How to solve a first order linear ODE using an integrating factor.

If you cannot solve a differential equation by separation, your next plan of attack should be to test if it is **linear** and use the **integration factor** method described below. Further exposition and examples are given in Stewart Section 10.5.

4.7.1 Definition

A first-order ODE is **linear** if it can be put in the form

$$\frac{dy}{dt} + p(t)y = q(t)$$

for some functions $p(t)$ and $q(t)$.

Exercise: Which of the following are linear?

- $y' = x$
- $y' = \log x$
- $y' = y^2$
- $y' = y$
- $y'y = 1$
- $y' = y + 3 \sin x$
- $y' = t^5 - t^2y$
- $y' = \sin y + e^x$
- $(y')^2 = y$

Recall the product rule for differentiation (written here in reverse order):

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

This formula is key to solving first-order linear ODEs.

Example: Solve $\frac{dy}{dt} + \left(\frac{2t}{t^2 + 1}\right)y = t^2 + 1$.



The lucky thing (it wasn't luck really) about the previous example was that after multiplication by $t^2 + 1$, the equation exactly matched the product rule. This raises the following natural question.

Question: Can we always multiply a first-order linear ODE by some function $I(t)$, so that its left-hand side matches the left-hand side of the product rule?

This question can be answered in the affirmative and any function $I(t)$ that does the trick is known as an **integrating factor**.

Given the ODE

$$\frac{dy}{dt} + p(t)y = q(t)$$

we multiply by the yet-to-be-found integrating factor $I(t)$:

$$I(t)\frac{dy}{dt} + I(t)p(t)y = I(t)q(t). \quad (*)$$

We think of I as g and y as f . Then the first term on the left is $f'g$ and consequently we want the second term on the left to be $fg' = yI'$. Given that this second term is in fact Ipy , we infer that I must satisfy the ODE

$$I'(t) = I(t)p(t).$$

This is good news since this is a separable ODE which we already know how to solve:

$$I(t) = \exp\left(\int p(t) dt\right).$$

Important remark: Any integrating factor will do. Hence when you try to solve $I' = Ip$, i.e., when you try to compute the above integral (assuming you can explicitly do so), you do not need to worry about keeping track of constants. For example, if $p(t) = 2t$, you would get

$$I(t) = \exp\left(2 \int t dt\right) = e^{t^2+C} = e^C e^{t^2} = B e^{t^2},$$

where $B = \exp(C) \neq 0$. When you substitute this in $(*)$ the constant B would simply be an overall factor which may thus be divided out. It is therefore perfectly okay to write on your assignment: **Because constants of integration do not matter,**

$$I(t) = \exp\left(2 \int t dt\right) = e^{t^2}.$$

To confuse your tutor, you might even write: **Because constants of integration do not matter,**

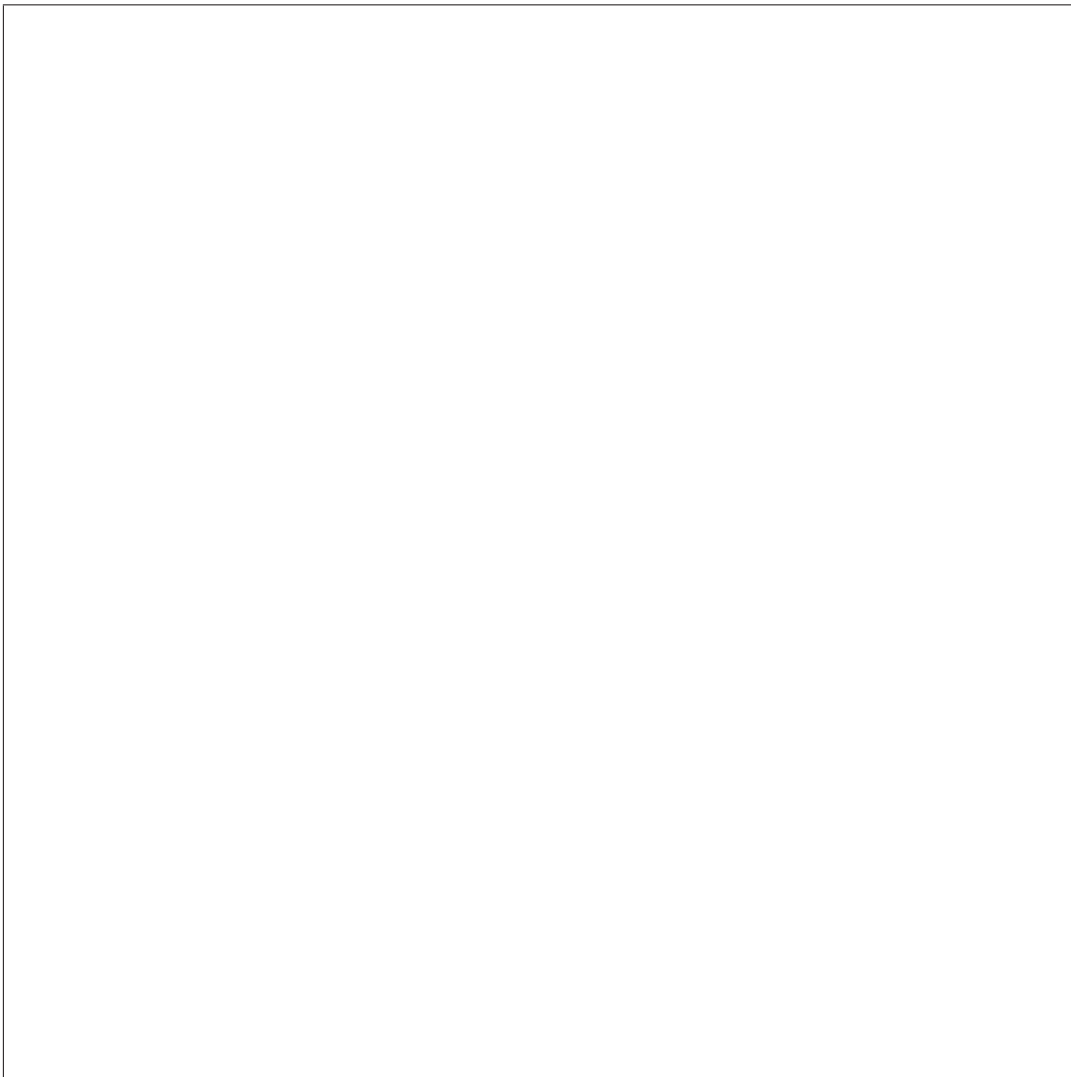
$$I(t) = \exp\left(2 \int t dt\right) = e^{t^2+65\pi^{123}}.$$

Summary

To solve a first-order linear ODE:

1. Write the equation in the form $\frac{dy}{dt} + py = q$.
2. Find an integrating factor $I = \exp\left(\int p dt\right)$.
3. Multiply the ODE by I and apply the product rule to get $\frac{d}{dt}(Iy) = Iq$.
4. Integrate both sides with respect to t . (This time, don't forget about constants of integration!).

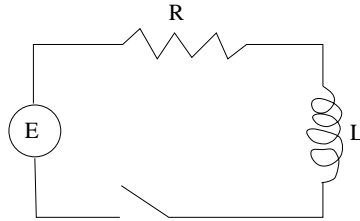
Example: Solve the ODE $\frac{dy}{dt} = y + t$.



Example: Solve the IVP $x^2 \frac{dy}{dx} + xy = 1$, $y(1) = 2$.

4.7.2 Partially worked-out example: RL circuit

The ODE describing the current in a circuit with a resistor and an inductor (known as an RL circuit) turns out to be linear. To obtain this equation consider the change in voltage around the circuit. The voltage drop due to a resistor is RJ where R is the resistance (in Ohms) and J is the current (in Amperes).



The voltage drop due to the inductor is $L \frac{dJ}{dt}$ where L is the inductance (in Henries).

According to **Kirchhoff's Law** the sum of the voltage drops is equal to the supplied voltage $E(t)$ (in Volts). Hence

$$L \frac{dJ}{dt} + RJ = E(t). \quad (*)$$

Example: Suppose in an RL circuit a battery supplies a constant voltage of 80V, the inductance is 2H and the resistance is 10Ω . (i) Find an expression for $J(t)$, and (ii) determine the the current after 1 second if $J(0) = 0$.

Solution: We first put (*) into standard form:

$$\frac{dJ}{dt} + \frac{RJ}{L} = \frac{E(t)}{L}.$$

An integrating factor for this ODE is

$$I(t) = \exp(R/L \int dt) = \exp(Rt/L).$$

Multiplying the ODE by this factor and using the product rule gives

$$e^{Rt/L} \frac{dJ}{dt} + e^{Rt/L} \frac{RJ}{L} = \frac{E(t)}{L} e^{Rt/L} \quad \Rightarrow \quad \frac{d}{dt} (e^{Rt/L} J) = \frac{E(t)}{L} e^{Rt/L}$$

Integrating the previous equation leads to

$$e^{Rt/L} J = \frac{1}{L} \int E(t) e^{Rt/L} dt \quad \Rightarrow \quad J(t) = \frac{1}{L} e^{-Rt/L} \int E(t) e^{Rt/L} dt.$$

So far we have treated everything completely generally. Now we use that the battery supplies constant voltage, i.e., E does not depend on t . This yields

$$J(t) = \frac{E}{L} e^{-Rt/L} \int e^{Rt/L} dt \Rightarrow J(t) = \frac{E}{L} e^{-Rt/L} \left(\frac{L}{R} e^{Rt/L} + C \right),$$

so that we finally obtain

$$J(t) = \frac{E}{R} + D e^{-Rt/L}.$$

Since $J(0) = 0$, the constant $D = -E/R$:

$$J(t) = \frac{E}{R} \left(1 - e^{-Rt/L} \right).$$

Since $E = 80$, $R = 10$ and $L = 2$:

$$J(t) = 8 \left(1 - e^{-5t} \right).$$

We can finally determine $J(1)$ as $J(1) = 8(1 - \exp(-5))$ Amperes.

Example: Now suppose that $E(t) = E_0 \sin(\omega t)$. Find an expression for $J(t)$.

4.8 Linear Second-Order Differential Equations with Constant Coefficients

On completion of this section, you should understand the following:

- How to identify a linear second order ODE.
- The superposition principle for homogeneous linear second-order ODEs.
- The general solution theorem.
- How to solve homogeneous linear second-order ODEs with constant coefficients.

Second-order ODEs are covered in Stewart in Sections 18.1 and 18.2. They often correspond with unforced (homogeneous) oscillators in our applications, but they arise in many other problems.

4.8.1 Definition

If the unknown function and its derivatives appear linearly in an ODE, the ODE is said to be **linear**. Otherwise it is **nonlinear**.

Which of the following ODEs are linear?

- $y'' + y' = y \log x$
- $y' = y^2 + y''$
- $y''y' = y$
- $y'y = y''$
- $y' + x^3y'' = xy + 3 \sin x$
- $y' = y''t^5 - t^2y$
- $y'' = \sin y + e^x$
- $y'' + (y')^2 = y$

A second-order linear ODE is an ODE that can be written in the form

$$\frac{d^2y}{dt^2} + p(t)\frac{dy}{dt} + q(t)y = r(t).$$

The functions $p(t)$ and $q(t)$ are called the **coefficients** of the ODE and if $r(t) = 0$ the ODE is called **homogeneous**.

4.8.2 The superposition principle

If $y_1(t)$ and $y_2(t)$ are solutions of

$$\frac{d^2y}{dt^2} + p(t)\frac{dy}{dt} + q(t)y = 0,$$

then so is

$$y(t) = c_1 y_1(t) + c_2 y_2(t).$$

Proof



Theorem: If $y_1(t)$ and $y_2(t)$ are two linearly independent solutions (i.e., $y_2(t) \neq Cy_1(t)$) of the homogeneous second-order linear ODE

$$\frac{d^2y}{dt^2} + p(t)\frac{dy}{dt} + q(t)y = 0$$

then the **general solution** is given by

$$y(t) = c_1y_1(t) + c_2y_2(t).$$

4.8.3 Reduction of order

The previous theorem states that all we need to do in order to solve a homogeneous linear second-order ODE is to find two linearly independent solutions. Unfortunately, that is where the good news ends: no general method (such as the integrating factor method for linear first-order ODE) is known to find two such solutions.

There is however a method to find a second solution once a first solution has been found. So, if you are clever enough to guess one solution (and a guess it will have to be, given that we have no general method) you can obtain a second, linearly independent solution. This goes by the name of **reduction of order**.

Assume your smart little sister has guessed that $y_1(x)$ is a solution to

$$\frac{d^2y}{dx^2} + p(x)\frac{dy}{dx} + q(x)y = 0.$$

To find a second solution $y_2(x)$ you set

$$y_2(x) = u(x)y_1(x).$$

with $u(x)$ a yet-to-be-determined function. Substituting this in the ODE gives

$$u''y_1 + 2u'y_1' + uy_1'' + p(u'y_1 + uy_1') + quy_1 = 0.$$

This may also be written as

$$u''y_1 + u'(y_1p + 2y_1') + u(y_1'' + py_1' + qy_1) = 0.$$

But y_1 is a solution to the ODE so the third term actually vanishes! Thus

$$u'' + u'\left(p + 2\frac{y_1'}{y_1}\right) = 0.$$

This is a first-order linear ODE in $v = u'$ which we know how to solve using an integrating factor.

Example: Dr. Viv has cleverly guessed that $y(x) = x^2$ solves the ODE $x^2y'' + 2xy' - 6y = 0$. Use reduction of order to find the general solution.



4.8.4 Homogeneous linear second-order ODEs with constant coefficients

We next discuss how to solve a homogeneous linear second-order ODE with constant coefficients:

$$ay'' + by' + cy = 0, \quad a \neq 0.$$

The trick is to try to find a solution of the form $y = e^{\lambda t}$. Substituting this into the ODE gives

$$a\lambda^2 e^{\lambda t} + b\lambda e^{\lambda t} + ce^{\lambda t} = 0.$$

Since $e^{\lambda t} \neq 0$ we may divide both sides of the above equation by this term to be left with the quadratic equation

$$a\lambda^2 + b\lambda + c = 0.$$

This equation is called the **characteristic equation** of the ODE.

Various scenarios can now arise, which we discuss in more detail below.

Case 1: Two distinct real roots

If $b^2 - 4ac > 0$ the characteristic equation $a\lambda^2 + b\lambda + c = 0$ has two distinct real roots, say λ_1 and λ_2 .

In this case the general solution to the ODE is

$$y = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}.$$

Question: What will this solution look like for very large t ?

Example: Solve the ODE $y'' - 3y' + 2y = 0$.

Case 2: A single root of multiplicity 2

If $b^2 - 4ac = 0$ the characteristic equation $a\lambda^2 + b\lambda + c = 0$ has one root, say μ , of **multiplicity 2**. (Of course $\mu = -b/(2a)$ but it is more convenient to work with μ).

This is somewhat of a problem because now we only have single solution to the ODE: $y = e^{\mu t}$.

Fortunately, we know that **reduction of order** will give us a second linearly independent solution, so that we try $y = u(t)e^{\mu t}$. By the product rule

$$\begin{aligned}y' &= u'e^{\mu t} + \mu u e^{\mu t} \\y'' &= u''e^{\mu t} + 2\mu u'e^{\mu t} + \mu^2 u e^{\mu t}.\end{aligned}$$

Substituting this into $ay'' + by' + cy = 0$ gives

$$\left[(a\mu^2 + b\mu + c)u + (2a\mu + b)u' + au'' \right] e^{\mu t} = 0.$$

Since μ is a solution to the characteristic equation, $a\mu^2 + b\mu + c = 0$. Moreover, since $\mu = -b/2a$, $2a\mu + b = 0$. Hence we are left with just

$$u'' = 0.$$

This is easily solved as

$$u(t) = (a + bt).$$

Using this with $a = 0$ and $b = 1$ we now have two linearly independent solutions:

$$y_1(t) = e^{\mu t} \quad \text{and} \quad y_2(t) = te^{\mu t},$$

where $\mu = -b/2a$.

Important remark: In actual problems you do not have to repeat the above derivation. It is enough for you to claim that if the characteristic equation has a single root of multiplicity 2 then the above are a pair of linearly independent solutions and

$$y(t) = (c_1 + c_2 t)e^{\mu t}$$

is the general solution.

Question: What will this solution look like for very large t ?

Example: Solve the IVP $y'' + 6y' + 9y = 0$, $y(0) = 2$, $y'(0) = 0$.

Case 3: Complex roots

If $b^2 - 4ac < 0$ the characteristic equation $a\lambda^2 + b\lambda + c = 0$ has a pair of complex conjugate roots, say $\alpha \pm i\beta$.

This leads to the general solution

$$y(t) = \left(c_1 e^{i\beta t} + c_2 e^{-i\beta t} \right) e^{\alpha t}.$$

For those who want the solution to be manifestly real we note that

$$e^{\pm i\beta t} = \cos(\beta t) \pm i \sin(\beta t),$$

so that

$$y(t) = \left((c_1 + c_2) \cos(\beta t) + (c_1 - c_2)i \sin(\beta t) \right) e^{\alpha t}.$$

Introducing new constants b_1 and b_2 by $b_1 = c_1 + c_2$ and $b_2 = (c_1 - c_2)i$ we thus get

$$y(t) = \left(b_1 \cos(\beta t) + b_2 \sin(\beta t) \right) e^{\alpha t}.$$

Example: Solve $y'' + 9y = 0$.



Example: Solve the IVP $y'' - 4y' + 5y = 0$, $y(0) = 1$, $y'(0) = 0$.

Challenge problem: Solve the ODE $y'''' + y''' - y'' + y' - 2y = 0$.

4.8.5 Extra reading: numerical solutions

Numerical solutions of second order ODEs usually require one second order equation to be split into two first order equations. Then the two coupled ODEs are solved together. For example, the equation

$$z'' + z' - z = \cos(x), \quad z(0) = 1, \quad z'(0) = 2$$

can be converted into:

$$\begin{aligned} z' &= y \\ y' &= -y + z + \cos(x) \end{aligned}$$

with initial conditions $z(0) = 1$, $y(0) = 2$. Then an Euler method can be used on these equations. For instance, the code to implement Euler's method would look like:

```
deltax=0.1;
x(1)=0;
z(1)=1;
y(1)=2;
for i = 1:100
    x(i+1) = x(i) + deltax;
    z(i+1) = z(i) + deltax*y(i);
    y(i+1) = y(i) + deltax*(-y(i) + z(i) + cos(x(i)));
end
plot(x,z)
```

4.9 Inhomogeneous Linear Second-Order ODEs

4.9.1 Inhomogeneous linear second-order ODEs with constant coefficients

After this section you should understand the following concepts:

- The difference between a particular solution and the general solution.
- How to find a particular solution.

Inhomogeneous second-order ODEs, also known as ODEs with forcing, are covered in Stewart in Section 18.2. They are very common in applications and often they require a bit of ingenuity to solve.

We will only consider inhomogeneous linear second-order ODEs, i.e., equations of the form

$$ay'' + by' + cy = r(x),$$

where r is a continuous function.

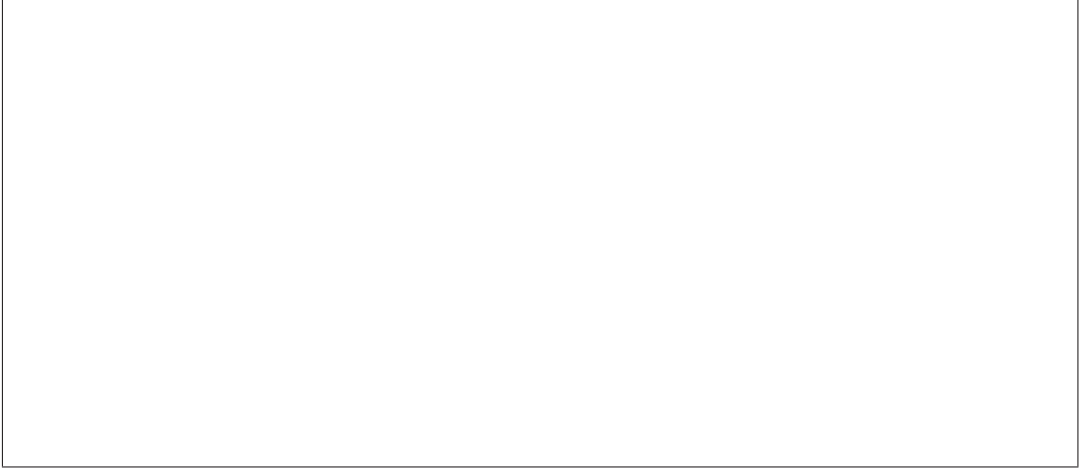
First we outline the general strategy of solving such equations.

1. First set $r(x) = 0$ and solve the corresponding homogeneous ODE. Its solution, $y_H(x) = c_1y_1(x) + c_2y_2(x)$, is known as the **complementary function**.
2. Find a solution, $y_P(x)$ to the inhomogeneous equation. Any solution will do and goes by the name of **particular solution**.
3. The general solution is $y(x) = y_H(x) + y_P(x)$.

Clearly the remaining question is: How does one find a particular solution? Below we will describe two methods. The first is again **reduction of order** and is the more cumbersome, albeit most general method. We will illustrate it in the following example.

Partially worked-out example: Solve the ODE $y'' + y' - 2y = x^2 - 2x + 3$.

- Show that the complementary function is given by $y_H(x) = c_1 e^x + c_2 e^{-2x}$.



- Set either $y(x) = e^x u(x)$ or $y(x) = e^{-2x} u(x)$. We will take the first option and leave the second option as homework.

Since $y = ue^x$,

$$y' = e^x(u' + u) \quad \text{and} \quad y'' = e^x(u'' + 2u' + u).$$

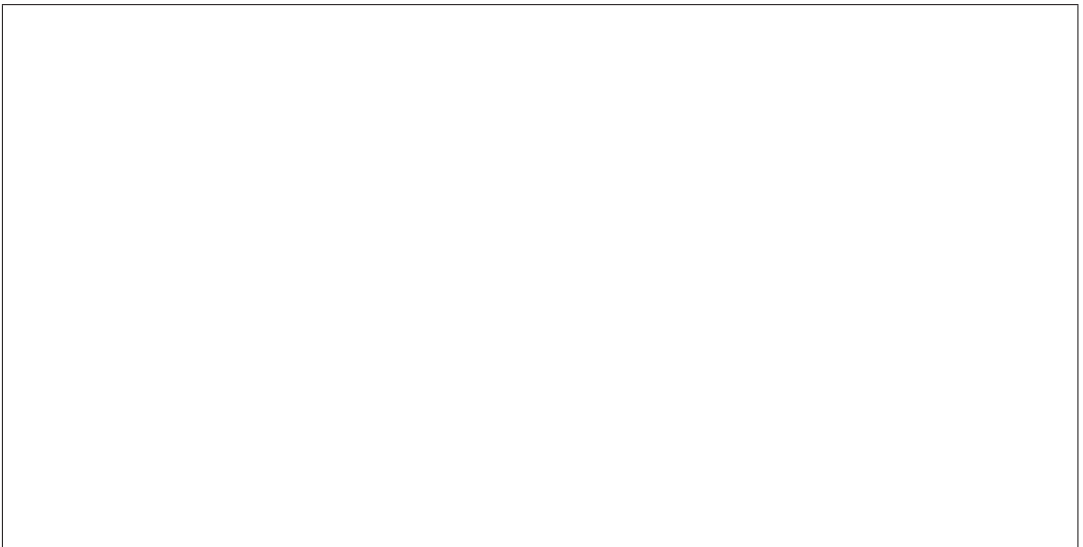
Substituting this into the ODE and dividing by e^x gives

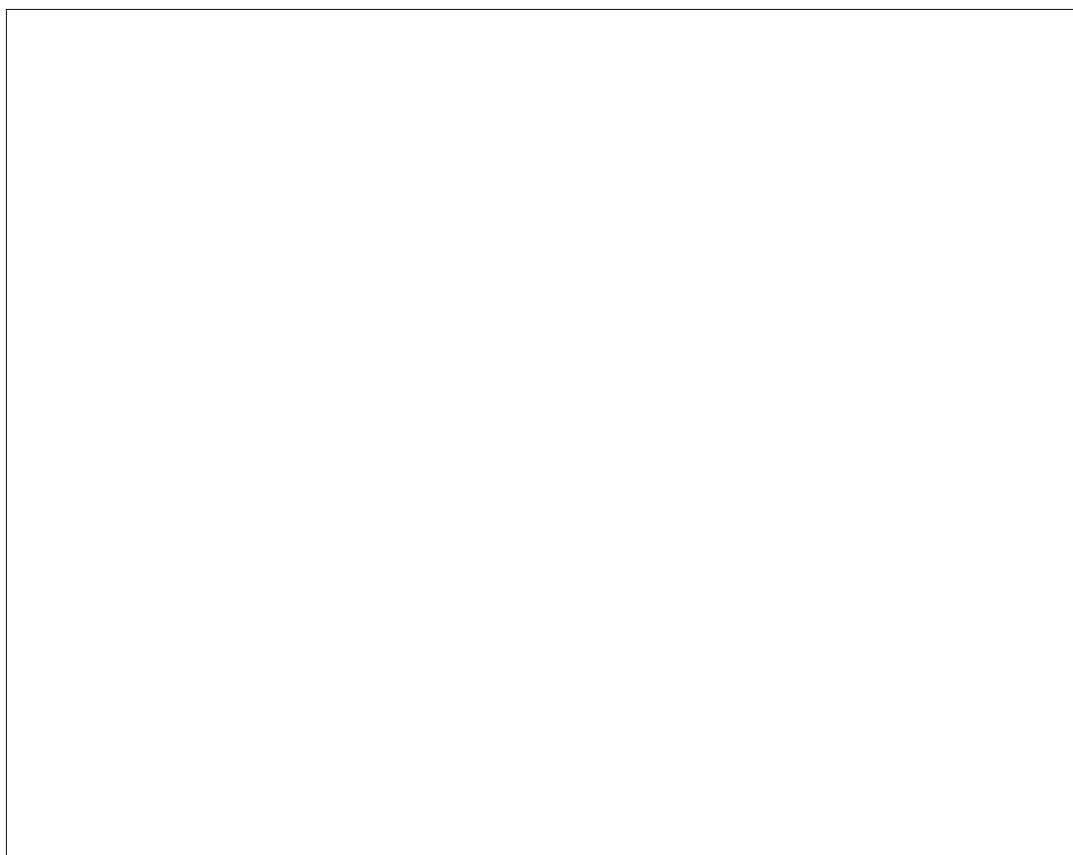
$$u'' + 3u' = e^x(x^2 - 2x + 3).$$

This is a linear first-order ODE for u' .

- Solve this linear first-order to find

$$u(x) = \frac{1}{4}(-2x^2 + 2x - 7)e^{-x} + c_1 + c_2 e^{-3x}.$$



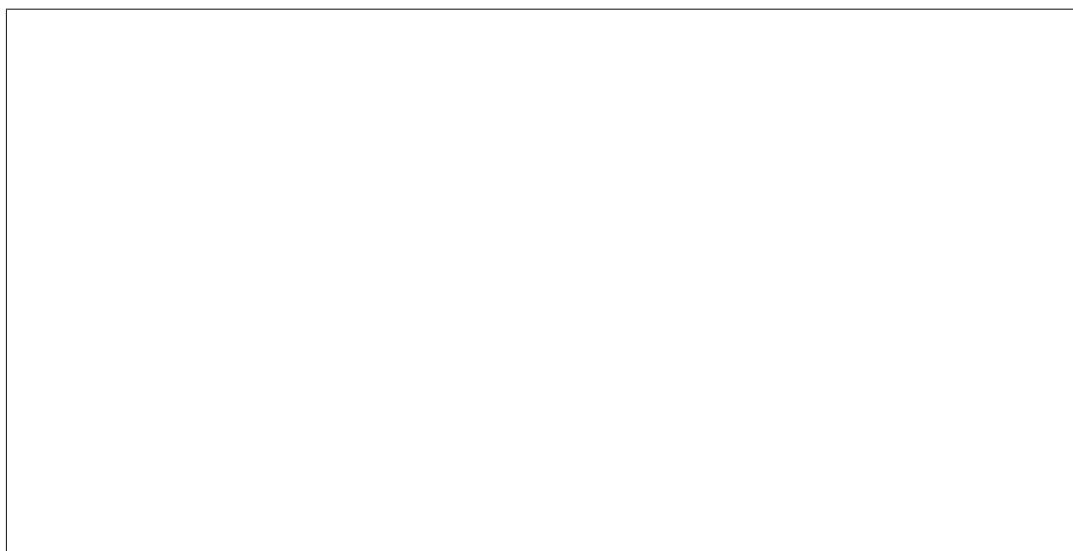


- Use that $y(x) = e^x u(x)$ to finally find

$$y(x) = \frac{1}{4}(-2x^2 + 2x - 7) + c_1 e^x + c_2 e^{-3x}$$

so that $y_P(x) = \frac{1}{4}(-2x^2 + 2x - 7)$.

- By direct substitution check that the above $y_P(x)$ is indeed a particular solution of the ODE.



Important remark: The previous method did produce the general solution given the complementary solution. It did not produce *just* a particular solution, which is all that is needed once the complementary solution is known. Below we will show that in many simple examples, you can save yourself a lot of work by *only going for a particular solution*.

We will consider the following cases for the inhomogeneous term $r(x)$:

- $r(x)$ is a polynomial.
- $r(x)$ is a polynomial times an exponential.
- $r(x)$ is a polynomial times a simple trigonometric function.
- $r(x)$ is a combination of the above.

Case 1: $r(x)$ is a polynomial.

The method is best illustrated through an example, and again we take

$$y'' + y' - 2y = x^2 - 2x + 3.$$

Because the polynomial on the right is of **degree 2** we differentiate the ODE 2 times to get

$$\begin{aligned} y''' + y'' - 2y' &= 2x - 2 \\ y'''' + y''' - 2y'' &= 2 \end{aligned}$$

Remembering that we are only after a particular solution, not the general solution, we can take $y''_P = -1$ as solution of the last equation since it correctly gives

$$0 + 0 - 2(-1) = 2.$$

If $y''_P = -1$ then $y'_P = -x + C$. Substituting this into

$$y''' + y'' - 2y' = 2x - 2$$

gives

$$0 + (-1) - 2(-x + C) = 2x - 2$$

so that $C = 1/2$, i.e., $y'_P = -x + 1/2$, or $y_P = (-x^2 + x + D)/2$. Substituting this in the original ODE gives

$$(-1) + (-x + 1/2) - (-x^2 + x + D) = x^2 - 2x + 3,$$

so that $D = -7/2$ and

$$y_P = \frac{1}{4}(-2x^2 + 2x - 7).$$

Important remark: In the above computations we twice had to fix a constant. Because “constants *are* constants” we could have *ignored all x -stuff* and streamlined the computations as follows:

... then $y'_P = -x + C$. Substituting this into

$$y''' + y'' - 2y' = 2x - 2$$

and ignoring any x 's gives

$$0 + (-1) - 2(C) = -2$$

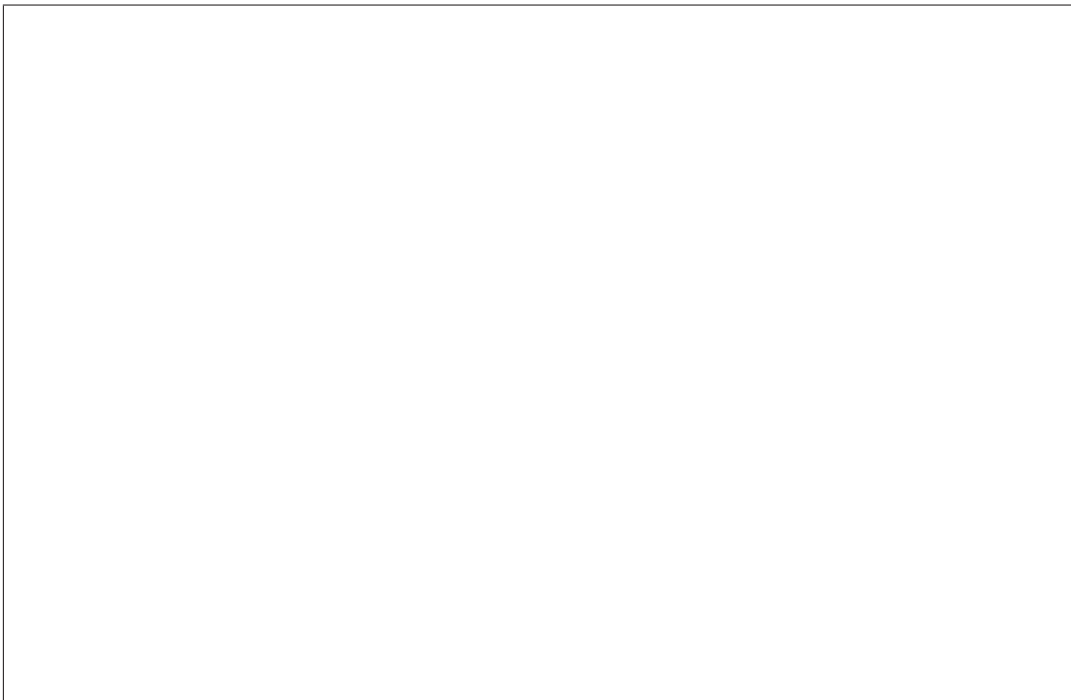
so that $C = 1/2$, i.e., $y'_P = -x + 1/2$, or $y_P = (-x^2 + x + D)/2$. Substituting this in the original ODE again ignoring any x 's gives

$$(-1) + (1/2) - (D) = 3,$$

so that $D = -7/2$ and

$$y_P = \frac{1}{4}(-2x^2 + 2x - 7).$$

Example: Find a particular solution of $3y'' - 2y' + 4y = 2x^3 + 1$.



Case 2: $r(x)$ is a polynomial times an exponential

Example: Find a particular solution to $y'' - y = 3xe^{2x}$.

Partially worked-out solution: In this case $r(x) = 3xe^{2x}$ is not a polynomial; no amount of differentiating will reduce it to a constant.

We thus have to get rid of the exponential first.

The trick is to put $y(x) = u(x)e^{2x}$ with $u(x)$ an unknown function.

Exercise: Show that u must satisfy the ODE $u'' + 4u' + 3u = 3x$.

We do not need to find the general solution of the ODE satisfied by u , but again only need a particular solution.

Exercise: Find a particular solution u_P .

Exercise: Use u_P to find a particular solution y_P .

Case 3: $r(x)$ is a polynomial times a simple trigonometric function

Example: Find a particular solution to $y'' - y = 2x \cos x$.

Partially worked-out solution: In this case $r(x) = 2x \cos x$.

The trick now is *to go complex*.

That is, we first determine a particular solution of $z'' - z = 2xe^{ix}$ and then take the real part of this to get a particular solution of $y'' - y = 2x \cos x$.





Case 4: $r(x)$ is a sum of the previous three cases.

The above three cases are enough to also deal with linear combinations such as

$$r(x) = x^2 + 2x + 1 + e^{2x}(x + 1)$$

or

$$r(x) = x^3 + \cos(2x) + xe^x.$$

Theorem: If y_{P_1} is a particular solution of $ay'' + by' + cy = r_1(x)$ and y_{P_2} is a particular solution of $ay'' + by' + cy = r_2(x)$, then $y_P = y_{P_1} + y_{P_2}$ is a particular solution of $ay'' + by' + cy = r_1(x) + r_2(x)$.

Example: Find a particular solution of $y'' - 2y' + y = x^2 + 1 + e^{-2x} + \cos x$.



4.10 Simple Harmonic Motion and Damped Oscillations

4.10.1 Main points

- You should know how to derive and solve the equations of motion for a damped oscillator given just the mass, damping and spring constants.
- You should also be able to classify the motion into under, over or critically damped, and know what these notions mean.

This material is covered in Stewart Section 18.3 and deals with the modelling of oscillators by second order ODEs.

4.10.2 The undamped spring

A spring of (natural) length L is stretched a distance s by a weight with mass m .

In equilibrium we take the position of the weight (viewed as a point-particle) to be $x = 0$, with the positive x -axis pointing down.

If we pull down the weight and then release it, it will start to oscillate.

If we assume the weight-spring-system moves free of any resistance (no air-resistance and no internal friction in the spring) we say the spring is **undamped**.

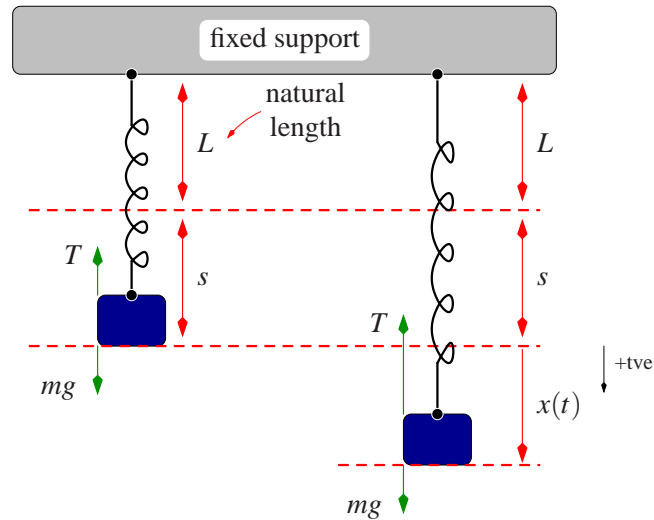
To derive an ODE for the undamped spring we must combine **Newton's second law** and **Hooke's law**.

According to Hooke's linear spring law the restoring force T exerted by the spring ("trying to unstretch") is proportional to the distance d it is stretched:

$$T = -kd.$$

Here $k > 0$ is called the **spring constant**; the larger the spring constant, the more tightly-coiled the spring.

Remark: As we all know from experience, Hooke's law only works if we do not stretch a spring too much.



According to Hooke's law the total force F is given by

$$F = \underbrace{-k(s+x)}_{\text{restoring force}} + \underbrace{mg}_{\text{gravitational force}},$$

with $x = x(t)$ the displacement out of equilibrium and g ($\approx 9.8 \text{ m/s}^2$) the acceleration on the Earth's surface.

Before we pulled down the weight the system was in equilibrium, so that

$$ks = mg.$$

We are thus left with

$$F = -kx. \quad (*)$$

According to **Newton's second law of motion** the total force equals mass times acceleration:

$$F = mx''. \quad (**)$$

Equating (*) and (**) leads to the **equation of motion**

$$mx'' + kx = 0$$

(which holds throughout the Universe, not just on Earth).

The equation of motion only depends on the ratio of k and m , and it is customary to introduce

$$\omega^2 = \frac{k}{m},$$

where $\omega > 0$ is known as the **angular frequency**.

Then the equation of motion for the undamped spring is thus given by

$$x'' + \omega^2 x = 0.$$

This is a homogeneous second-order linear ODE with constant coefficients.

The corresponding characteristic equation is given by

$$\lambda^2 + \omega^2 = 0$$

with purely imaginary roots

$$\lambda_1 = i\omega \quad \text{and} \quad \lambda_2 = -i\omega.$$

Two linearly independent solutions are thus

$$e^{i\omega t} \quad \text{and} \quad e^{-i\omega t}$$

or

$$\cos(\omega t) \quad \text{and} \quad \sin(\omega t).$$

and the general solution is

$$x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t).$$

Although we could stop here it is standard to rewrite the above solution in the alternative form

$$x(t) = A \cos(\omega t - \phi),$$

where the two constants c_1 and c_2 have been traded for two new constants $A > 0$ and $\phi \in (-\pi, \pi]$, known as the **amplitude** and **phase shift**.

Example: Find the relation between c_1, c_2 and A, ϕ .

4.10.3 Example

A spring with mass 9kg and spring constant 4N/m is pulled down 1m and given an initial upward kick of -0.5 m/s . Solve for position as a function of time and explicitly determine the amplitude of the oscillations.



4.10.4 The damped spring

If you try the “spring-weight” system at home it is unlikely to display true **simple harmonic motion**. If that were possible you could build a perpetuum mobile.

To make the model more realistic we need to take into account the effect of **damping**. This is caused by air-resistance and mechanical friction of the spring.

The most common model of damping assumes that it corresponds to a force F_d , proportional to the velocity of the moving object, and opposite to the direction of movement. Hence

$$F_d = -\beta x',$$

where $\beta > 0$ is the **damping constant**.

The total force F for the damped spring is thus given by

$$F = \underbrace{-\beta x'}_{\text{damping force}} + \underbrace{-k(s+x)}_{\text{restoring force}} + \underbrace{mg}_{\text{gravitational force}}.$$

As before,

$$ks = mg,$$

so that

$$F = -\beta x' - kx.$$

Thanks to Newton’s second law this yields the equation of motion

$$mx'' + \beta x' + kx = 0.$$

This equation only depends on the ratios of k/m and β/m , and it is customary to not only set

$$\omega^2 = \frac{k}{m} > 0$$

but also

$$2p = \frac{\beta}{m} > 0.$$

Then the equation of motion for the damped spring is given by

$$x'' + 2p x' + \omega^2 x = 0.$$

This is a homogeneous second-order linear ODE with constant coefficients.

The corresponding characteristic equation is given by

$$\lambda^2 + 2p\lambda + \omega^2 = 0,$$

which has **discriminant**

$$D = 4(p^2 - \omega^2).$$

There are three different cases to consider.

- $D < 0$ (i.e., $p < \omega$).

This is the **underdamped** or **weakly-damped** spring.

- $D = 0$ (i.e., $p = \omega$).

This is the **critically-damped** spring.

- $D > 0$ (i.e., $p > \omega$).

This is the **overdamped** or **strongly-damped** spring.

4.10.5 Underdamping

For $p < \omega$ the characteristic equation

$$\lambda^2 + 2p\lambda + \omega^2 = 0$$

has two complex conjugate roots

$$\lambda_1 = -p + i\sqrt{\omega^2 - p^2} \quad \text{and} \quad \lambda_2 = -p - i\sqrt{\omega^2 - p^2}.$$

Two linearly independent solutions are thus

$$e^{-pt} \cos(\omega_p t) \quad \text{and} \quad e^{-pt} \sin(\omega_p t),$$

where

$$\omega_p = \sqrt{\omega^2 - p^2} = \omega \sqrt{1 - (p/\omega)^2}.$$

The general solution is

$$x(t) = e^{-pt} \left(c_1 \cos(\omega_p t) + c_2 \sin(\omega_p t) \right).$$

Again it is best to write this in terms of a single trigonometric function:

$$x(t) = A e^{-pt} \cos(\omega_p t - \phi) = A(t) \cos(\omega_p t - \phi).$$

Here $A(t)$ is a time-dependent amplitude:

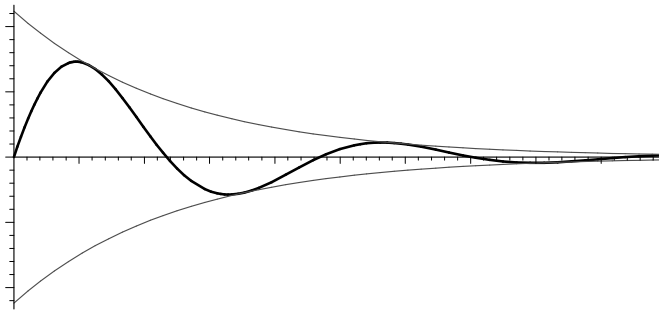
$$A(t) = Ae^{-pt}$$

and ω_p is the frequency of the underdamped spring.

Since

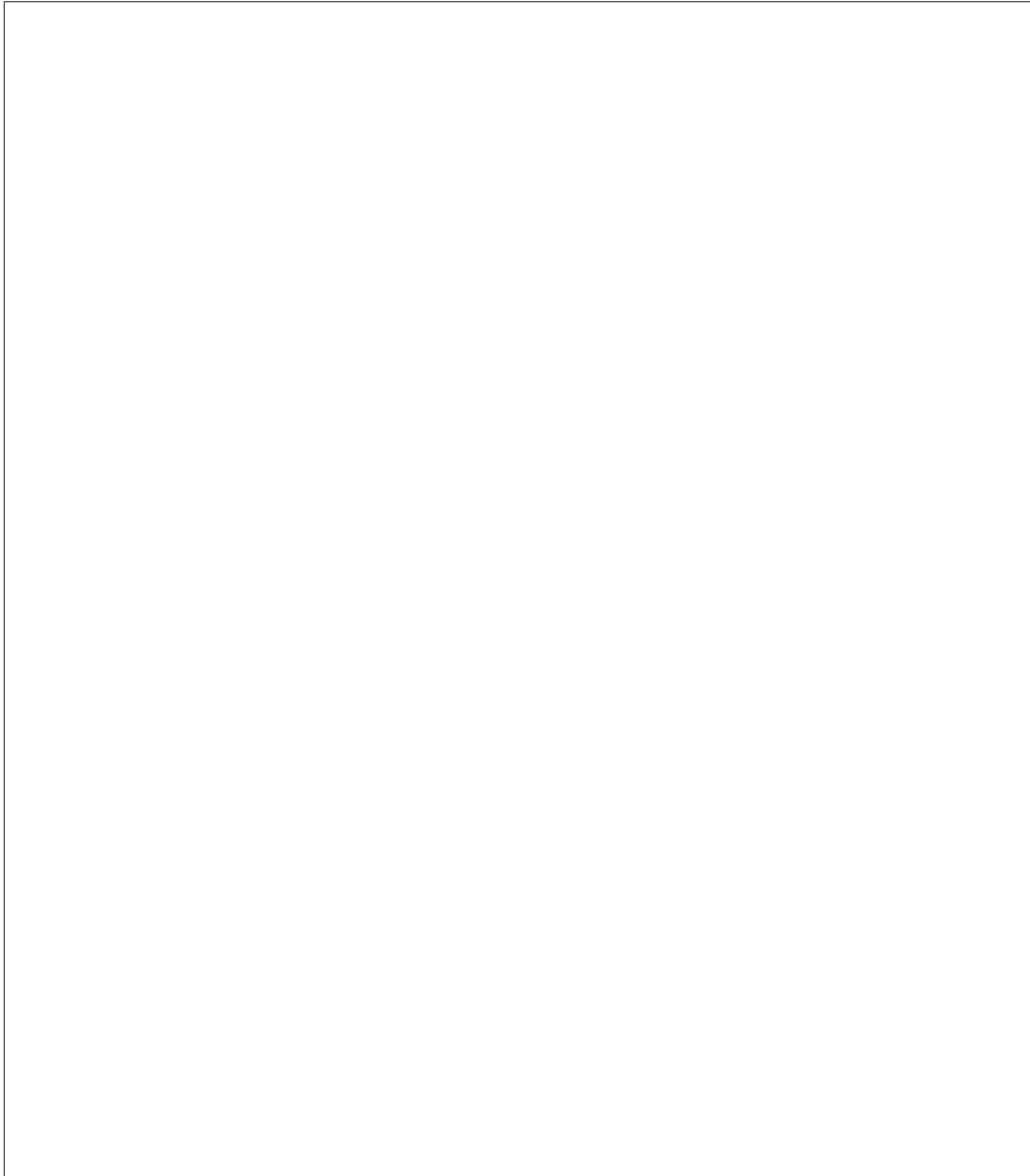
$$\frac{\omega_p}{\omega} = \sqrt{1 - (p/\omega)^2} < 1$$

the damping of the spring leads to a **frequency red-shift** and an increase in the period of the spring.



When p approaches ω the period tends to infinity so that we may expect that the critically-damped spring does not display oscillatory motion. This is the case considered in the next section.

Example: A mass of 1kg is attached to a spring hanging under gravity with damping constant 0.2 and spring constant 4. Find the position of the mass after time t if it is pulled down 1m from the equilibrium position and released without kick.



4.10.6 Critical damping

For $p = \omega$ the characteristic equation

$$\lambda^2 + 2p\lambda + \omega^2 = 0$$

has one root of multiplicity two: $\lambda = -p$. Two linearly independent solutions are thus

$$e^{-pt} \quad \text{and} \quad te^{-pt}$$

and the general solution is

$$x(t) = e^{-pt}(c_1 + c_2t).$$

Unlike the underdamped spring it is straightforward to express c_1 and c_2 in terms of the initial conditions $x(0)$ and $x'(0)$:

$$c_1 = x(0) \quad \text{and} \quad c_2 = x(0)p + x'(0).$$

Therefore

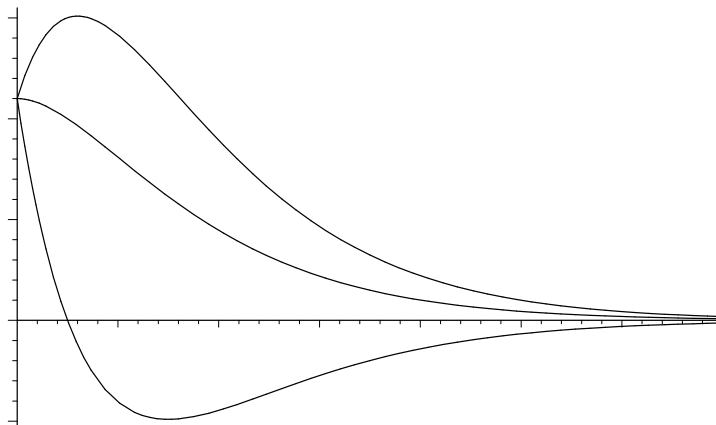
$$x(t) = e^{-pt}(x(0) + x(0)pt + x'(0)t).$$

The linear polynomial $x(0) + x(0)pt + x'(0)t$ vanishes for

$$t = -\frac{x(0)}{x(0)p + x'(0)}.$$

This corresponds to a positive time t if $x'(0) < -x(0)p$.

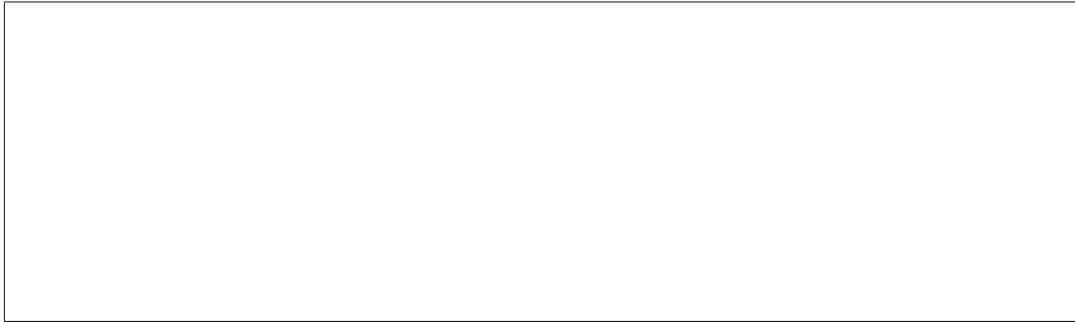
Depending on the initial conditions, the critically-damped spring behaves like



- The top-most curve corresponds to $x(0)$ and $x'(0)$ both positive, so that the weight is not just pulled down a distance $x(0)$ but is also given a downward kick on release.

- The middle curve corresponds to $x'(0) = 0$ so that we only pull down the weight and then release it.
- The third curve corresponds to $x'(0) < 0$ so that upon release the weight is kicked upward. Since the spring *overshoots the equilibrium* we must also have that $x'(0) < -x(0)p$.

Example: A mass of 1kg is attached to a spring hanging under gravity, with spring constant 4. What value of the damping constant implies critical damping?



4.10.7 Overdamping

For $p > \omega$ the characteristic equation

$$\lambda^2 + 2p\lambda + \omega^2 = 0$$

has two distinct real roots

$$\lambda_1 = -p - \sqrt{p^2 - \omega^2} =: -\alpha$$

$$\lambda_2 = -p + \sqrt{p^2 - \omega^2} =: -\beta$$

with $\alpha > \beta > 0$.

Two linearly independent solutions are thus

$$e^{-\alpha t} \quad \text{and} \quad e^{-\beta t}$$

and the general solution is

$$x(t) = c_1 e^{-\alpha t} + c_2 e^{-\beta t}.$$

For generic initial conditions $x(0)$ and $x'(0)$, the constants c_1 and c_2 are both nonzero.

Assuming this, we have that for sufficiently large t

$$x(t)_{\text{critical damping}} \sim c_2 t e^{-\beta t}$$

and

$$x(t)_{\text{overdamped}} \sim c_2 e^{-\beta t}.$$

Since $\beta = p - \sqrt{\dots} < p$ we see that the overdamped spring takes longer to return to equilibrium than the critically damped spring.

The overdamped spring describes a door-closer that is too tight. It may take a long time for the door to actually close because the strong damping restricts (nearly) all movement.

A tradeswoman will try to adjust a door-closer such that the damping will be as close as possible to critical damping. As always, **Maths Rules!**

Example: A mass of 1kg is attached to a spring hanging under gravity with damping constant 5 and spring constant 4. Find the position of the mass after time t if it is pulled down 1m from the equilibrium position and released without kick.



4.10.8 Summary of motion of a spring:

The equation of motion of a mass m attached to a damped spring is described by the ODE

$$x'' + 2px' + \omega^2 x = 0,$$

where

$$\omega^2 = \frac{k}{m} > 0, \quad 2p = \frac{\beta}{m} > 0$$

and k is the spring constant and β the damping constant.

The roots of the characteristic equation $\lambda^2 + 2p\lambda + \omega^2 = 0$ determine the type of motion:

- $p = 0$; No damping — pure oscillatory or **harmonic** motion.

The roots are $\pm i\omega$ and the general solution is given by

$$x(t) = c_1 \cos(\omega t) + c_2 \sin(\omega t) = A \cos(\omega t - \phi).$$

- $p < \omega$; Underdamped motion.

The roots are $-p \pm i\omega_p$, $\omega_p = \omega \sqrt{1 - (p/\omega)^2}$ and the general solution is given by

$$x(t) = e^{-pt} (c_1 \cos(\omega_p t) + c_2 \sin(\omega_p t)) = A e^{-pt} \cos(\omega_p t - \phi).$$

- $p = \omega$; Critically damped motion.

The single root (of multiplicity 2) is given by $-p$ and the general solution is given by

$$x(t) = e^{-pt} (c_1 + c_2 t).$$

- $p > \omega$; Overdamped motion.

The two roots (denoted $-\alpha$ and $-\beta$) are $-p \pm \sqrt{p^2 - \omega^2}$ and the general solution is given by

$$x(t) = c_1 e^{-\alpha t} + c_2 e^{-\beta t}.$$

4.10.9 The pendulum

People as old as Dr. VO, with even older grandparents, will remember those beautiful old pendulum clocks. In such a clock a pendulum of mass m is suspended from a pivot by a long, thin metal rod.

Now suppose that the pendulum moves in an arc of a circle with radius r . If the pendulum makes an angle (say in the anti-clockwise direction) θ with its equilibrium position, then it has travelled a distance $r\theta$ from this position. According to Newton's second law of motion

$$m \frac{d^2}{dt^2}(r\theta) = F,$$

where F is the force in the direction of motion. Hence

$$rm \frac{d^2\theta}{dt^2} = -mg \sin \theta.$$

This is a nonlinear ODE which displays very interesting nonlinear behaviour. If we however assume that θ is very small (the pendulum does not swing very far out, unlike your little sister on the swing perhaps), then we have $\sin \theta \simeq \theta$. Making this approximation we get

$$\frac{d^2\theta}{dt^2} + \frac{g}{r} \theta = 0.$$

Setting $g/r = \omega^2$ as before we obtain the same ODE describing simple harmonic motion.

Question: What is the period of the undamped pendulum?

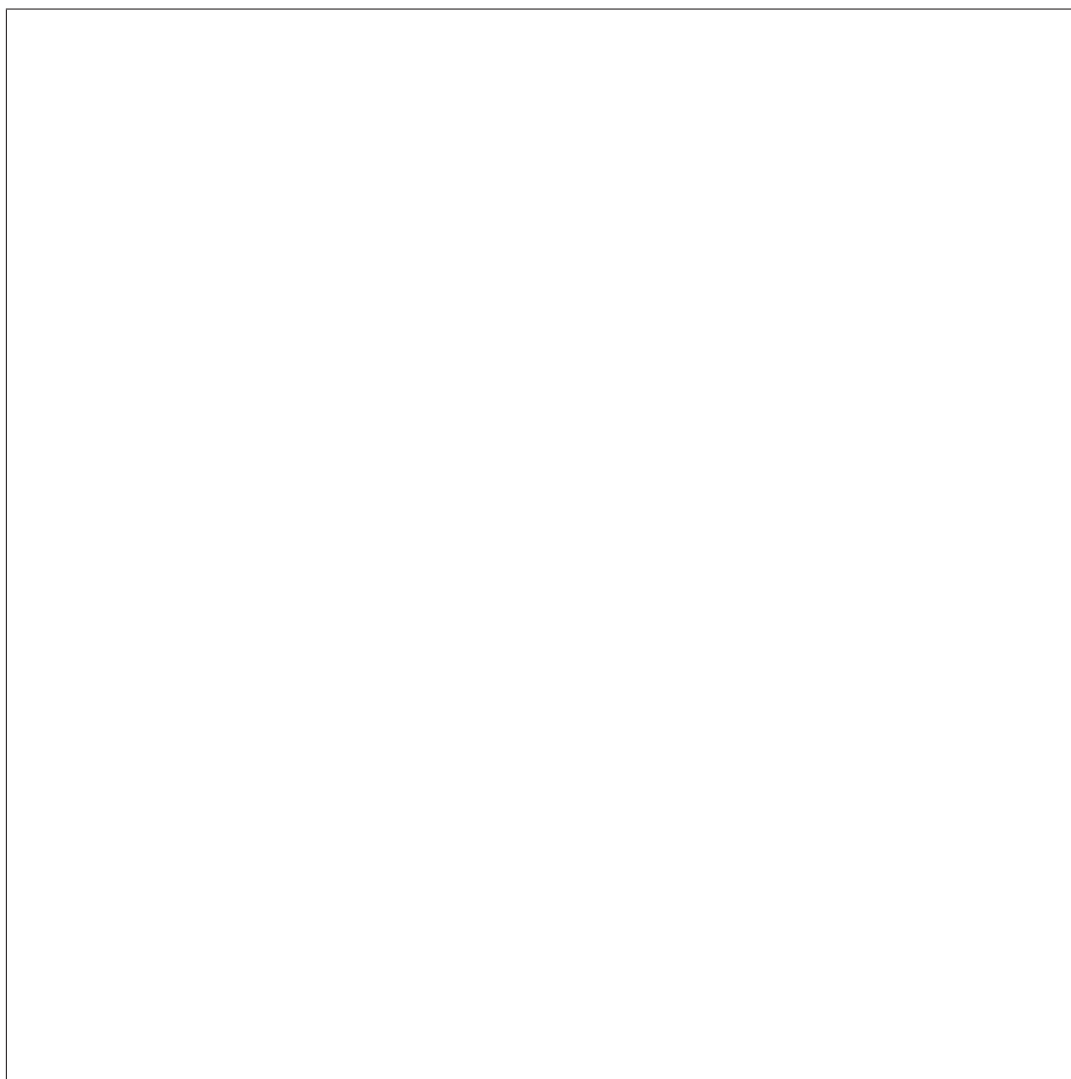
As before one can make the model more realistic by assuming damping proportional to velocity.

Example: If the damping constant is γ show that this leads to the ODE

$$\theta'' + 2p\theta' + \omega^2\theta = 0,$$

where

$$\omega^2 = \frac{g}{r} > 0, \quad 2p = \frac{\gamma}{r} > 0.$$



Remark: It is interesting to note that the ODE describing the (damped) pendulum does not depend on the mass of the pendulum.

5 Parametrisation of Curves and Line Integrals

Stewart covers **parametrisation** in chapters 11 and (some of) 13. There are some features of Matlab, especially plotting which will work with **parametric** forms.

5.1 Parametrisation of Curves

By the end of this section, you should understand what a parametric representation is, and you should be able to parametrise ellipses, circles, spirals, helices and piecewise-linear paths (although we will cover piecewise-linear paths in more detail in the next section). You should be able to plot parametric forms in Matlab and roughly sketch them on paper.

Imagine some curve in the xy -plane, for example, $x^2 + 2y^2 = 1$ or $3x + y = 2$. One way to think of the curve is as the trajectory of a moving particle, such that at time t the particle is at the point $(x(t), y(t))$ on the curve. In other words, we describe the curve by

$$\begin{cases} x = f(t) \\ y = g(t). \end{cases}$$

Such equations for the curve are called **parametric equations**, and the auxiliary variable t , is called a **parameter**. The functions f and g are said to provide a **parametrisation** of the curve. Before we look at some examples, it should be clear from our mental picture of a moving particle that a parametrisation of a curve is not unique. For example,

$$\begin{cases} x = F(t) \\ y = G(t), \end{cases}$$

where, say $F(t) = f(3 - t^2)$ and $G(t) = g(3 - t^2)$, is another parametrisation of the same curve; we simply have changed the velocity and initial (i.e., $t = 0$) position of the imaginary particle.

Example: Sketch the curve parametrised as

$$\begin{cases} x = t^2 - 2t \\ y = t + 1 \end{cases}$$

and give an equation for the curve in terms of just the Cartesian coordinates x and y . Give two alternative parametrisations of the curve.

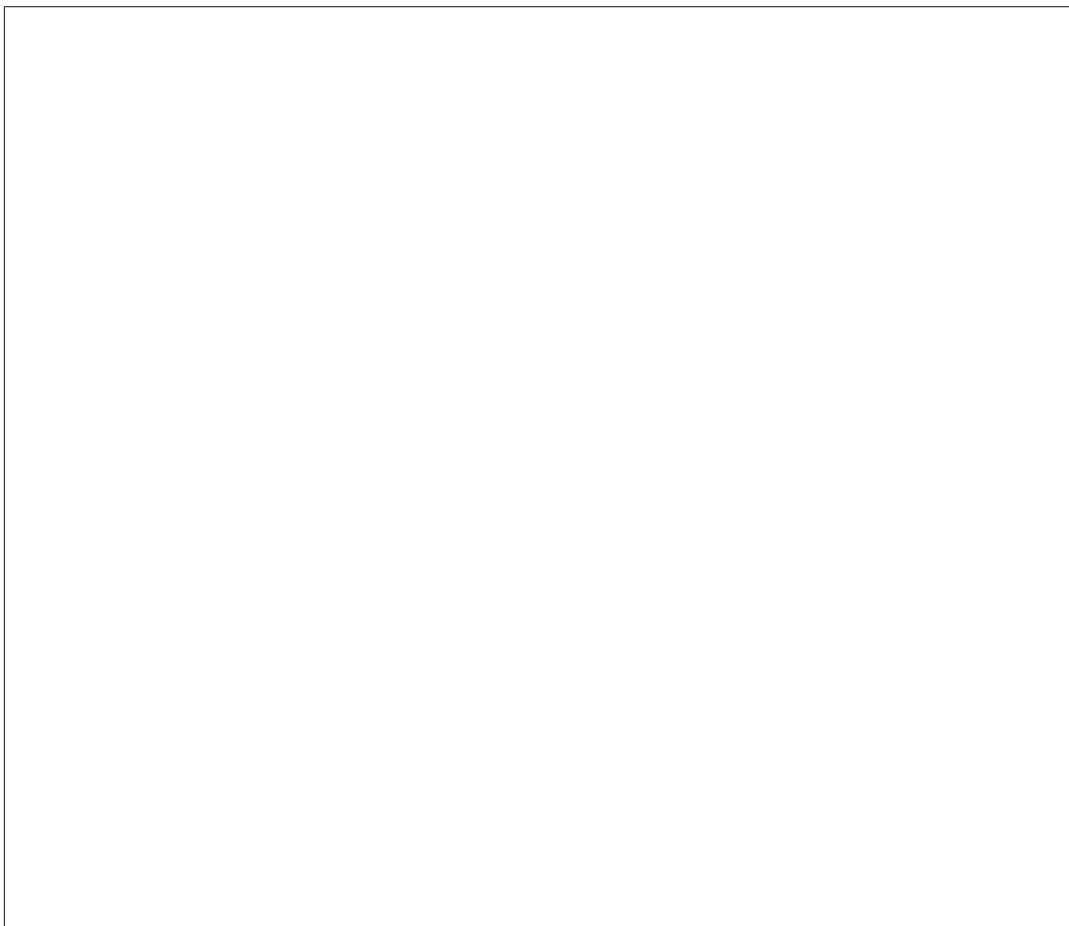


Sometimes one has reasons to restrict t to a finite interval. In general, the curve with parametric equations

$$\begin{cases} x = f(t) \\ y = g(t) \end{cases} \quad a \leq t \leq b$$

starts at the point $(f(a), g(a))$ and finishes at the point $(f(b), g(b))$. One has to be careful though, since even a finite time-interval can still give rise to an infinitely-long curve.

Example: Sketch the curve $\begin{cases} x = t^2 - 2t \\ y = t + 1 \end{cases} \quad 0 \leq t \leq 4.$



Example: Sketch the Archimedian spiral given by $\begin{cases} x = t \cos t \\ y = t \sin t \end{cases} \quad t \geq 0$
and verify that it may also be described by $x^2 + y^2 = \arctan^2(y/x)$.



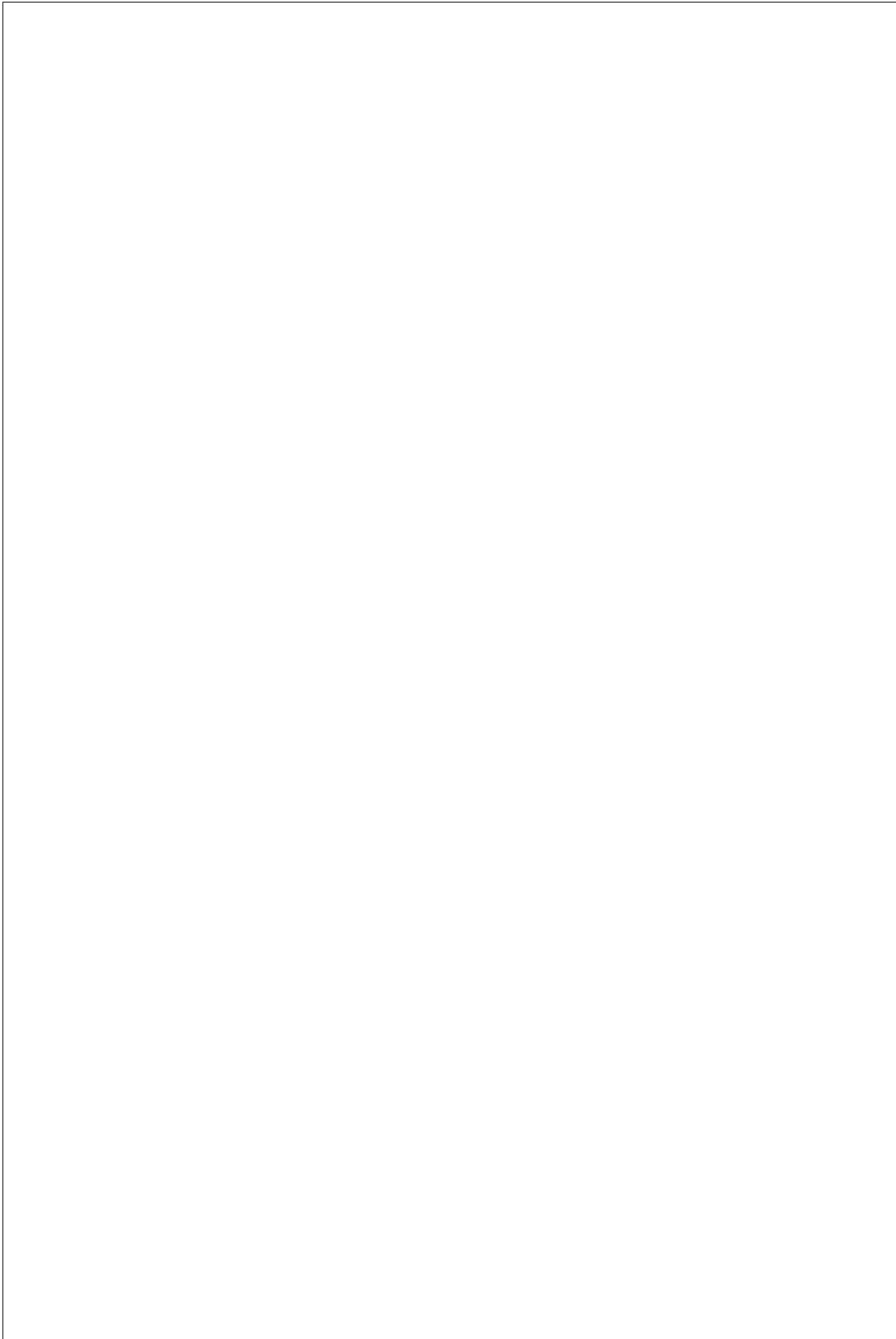
If you like spirals you could also try to sketch

$$\begin{cases} x = \frac{\cos t}{t} \\ y = \frac{\sin t}{t} \end{cases} \quad t > 0; \quad \begin{cases} x = \frac{\cos t}{\sqrt{t}} \\ y = \frac{\sin t}{\sqrt{t}} \end{cases} \quad t > 0; \quad \begin{cases} x = e^{-t} \cos t \\ y = e^{-t} \sin t \end{cases} \quad t \in \mathbb{R}.$$

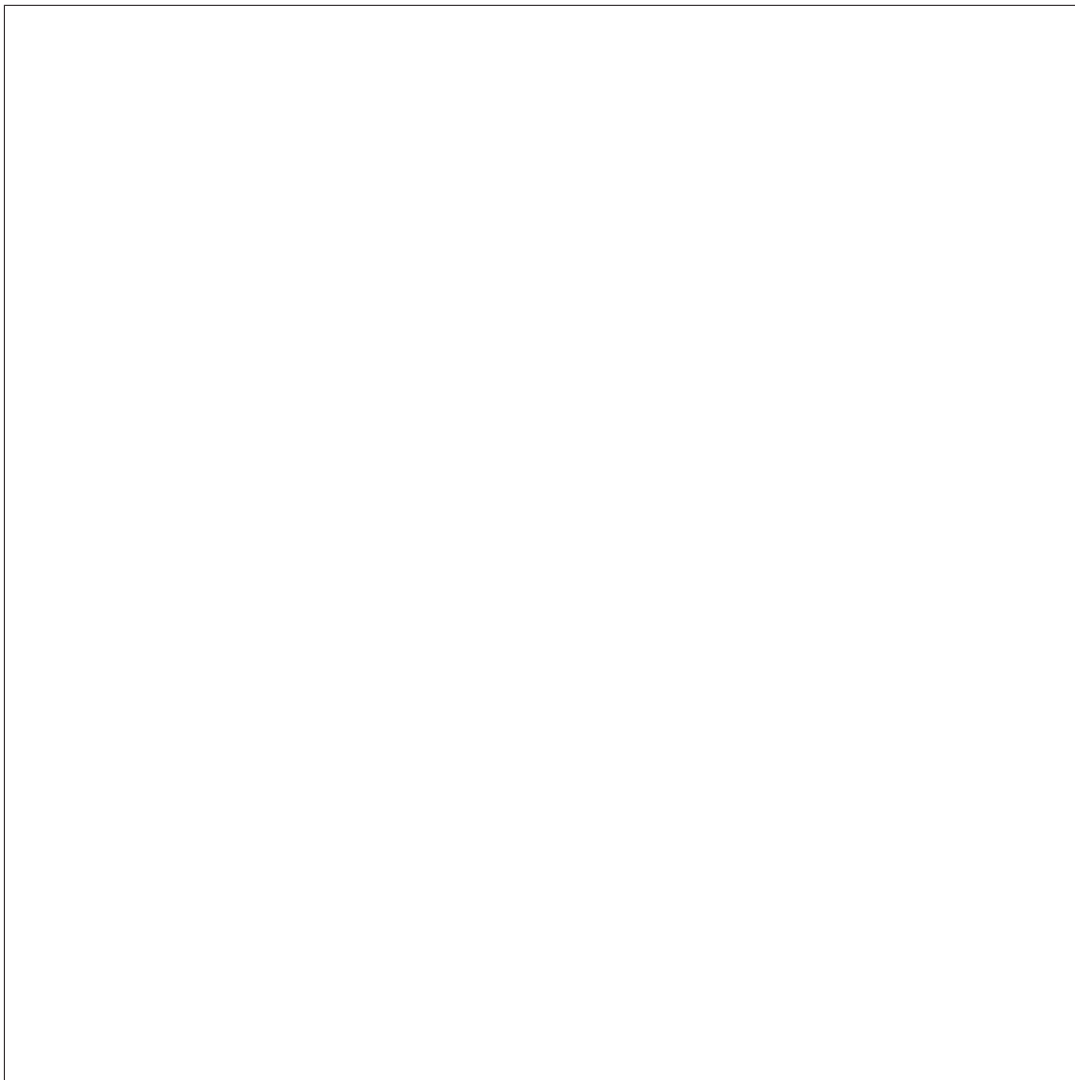
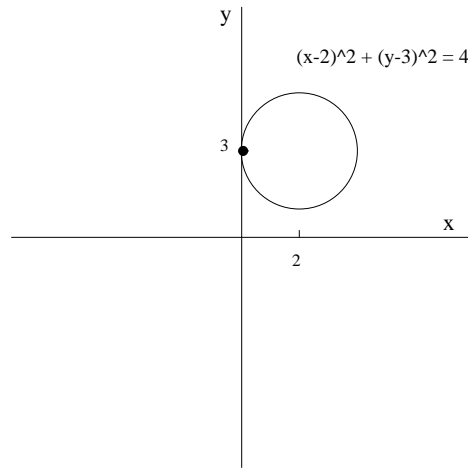
For the true connoisseurs there even is the amazing Cornu spiral; who said maths couldn't be fun?

$$\begin{cases} x = \int_0^t \cos(s^2) ds \\ y = -\int_0^t \sin(s^2) ds, \end{cases} \quad t > 0.$$

Example: Parametrise the circle of radius r centered at (a, b) .



Example: Parametrise the circle with radius 2 centered at $(2, 3)$, such that the imaginary particle tracing this circle is travelling in clockwise direction and at time $t = 0$ is at $(0, 3)$.



Example: Parametrise the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$



A much more fancy parametrisation may be given using Jacobi's elliptic functions sn , cn and dn . You will hopefully learn about these functions in a more advanced mathematics course.

5.1.1 Curves in \mathbb{R}^3

In general, a curve in three dimensions can be parametrised in the form

$$\begin{cases} x = f(t) \\ y = g(t) \\ z = h(t). \end{cases}$$

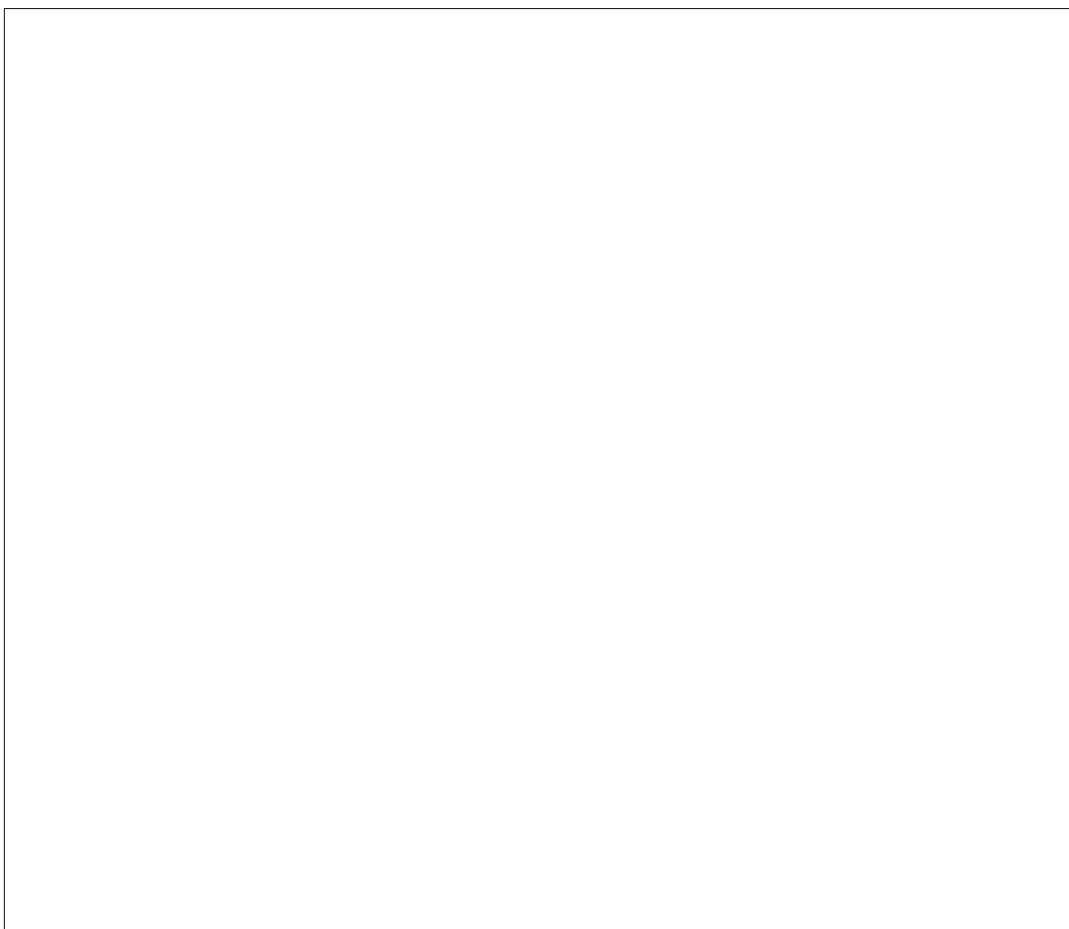
A **helix** is a three-dimensional curve that has shape of a wound spring, or a piece of string wrapped around a cylinder.

One possible parametrisation of the helix is

$$\begin{cases} x = r \cos(\pi t) \\ y = r \sin(\pi t) \\ z = t \end{cases}$$

for $t \in \mathbb{R}$.

Example: Sketch the helix for $t \in [-2, 2]$.



5.1.2 Piecewise linear paths

Example: Sketch the curve parametrised by

$$\begin{array}{lll} x(t) = 1, & y(t) = t, & 0 \leq t \leq 1 \\ x(t) = 2 - t, & y(t) = t, & 1 < t \leq 2 \\ x(t) = 2 - t, & y(t) = 2, & 2 < t \leq 3. \end{array}$$



5.1.3 Extra reading: Matlab plotting parametrised equations

Early in the semester, you performed many types of plots, including parametric plotting. The `ezplot` function handles parametric plots, so to plot the ellipse in the earlier example, the command is simply

```
ezplot('2*cos(t)', '1/3*sin(t)')
```

To plot the spiral, simply use

```
ezplot('exp(-t/100)*cos(t)', 'exp(-t/100)*sin(t)')
```

Executing this command indicates the default range for the t values. To get a longer spiral, $0 \leq t \leq 100$, run

```
ezplot('exp(-t/100)*cos(t)', 'exp(-t/100)*sin(t)', [0,100])
```

5.2 Position Vectors, Velocity and Acceleration

By the end of this section, you should understand the following concepts:

- The relationship between a position vector and the parametric equations of a curve.
- The relationship between position, velocity and acceleration vectors.

Stewart covers the following material in Section 14.4. Parametric equations are used to describe **position**, **velocity** and **acceleration**.

5.2.1 Position vector

The **position vector**

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$$

traces out the path given by the parametrisation $(x(t), y(t))$. Similarly in three dimensions, the position vector

$$\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$$

traces out the path given by the parametrisation $(x(t), y(t), z(t))$.

5.2.2 Parametrising straight lines

Say we want to parametrise the straight line between the points (a, b) and (c, d) . We also require that (a, b) corresponds to parameter value $t = \alpha$, and (c, d) corresponds to $t = \alpha + \Delta t$ (where $\Delta t > 0$).

Let \mathbf{r}_0 be the position vector for (a, b) and \mathbf{r}_1 the position vector for (c, d) . The vector

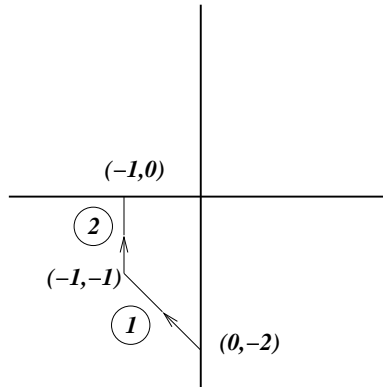
$$\mathbf{v} = \mathbf{r}_1 - \mathbf{r}_0$$

is parallel to the line between (a, b) and (c, d) .

A parametrisation for the line is then

$$\mathbf{r}(t) = \mathbf{r}_0 + \frac{t - \alpha}{\Delta t} \mathbf{v} = \left(\frac{\alpha + \Delta t - t}{\Delta t} \right) \mathbf{r}_0 + \left(\frac{t - \alpha}{\Delta t} \right) \mathbf{r}_1.$$

Example: Parametrise the path given in the following diagram, so that path ① is given for $0 \leq t \leq 1$ and the path ② is given for $1 \leq t \leq 3$.



Example: Find the position vector of any point on the line passing through the point $(0, 1, 3)$ and parallel to the vector $2\mathbf{i} - 3\mathbf{j} + \frac{1}{2}\mathbf{k}$.



5.2.3 Velocity vector

The vector tangent to the path of motion, with magnitude equal to the speed is the **velocity vector**. If $\mathbf{r}(t)$ is the position vector, then

$$\Delta \mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$$

is approximately tangent to the curve traced out by $\mathbf{r}(t)$. The approximation gets better as $\Delta t \rightarrow 0$. So the velocity vector \mathbf{v} is given by

$$\begin{aligned} \mathbf{v}(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{r}}{\Delta t} \\ &= \frac{d\mathbf{r}}{dt}. \end{aligned}$$

In component-form, if $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$, then the velocity is

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k}.$$

Speed is the magnitude of velocity:

$$\text{speed} = \|\mathbf{v}(t)\| = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}.$$

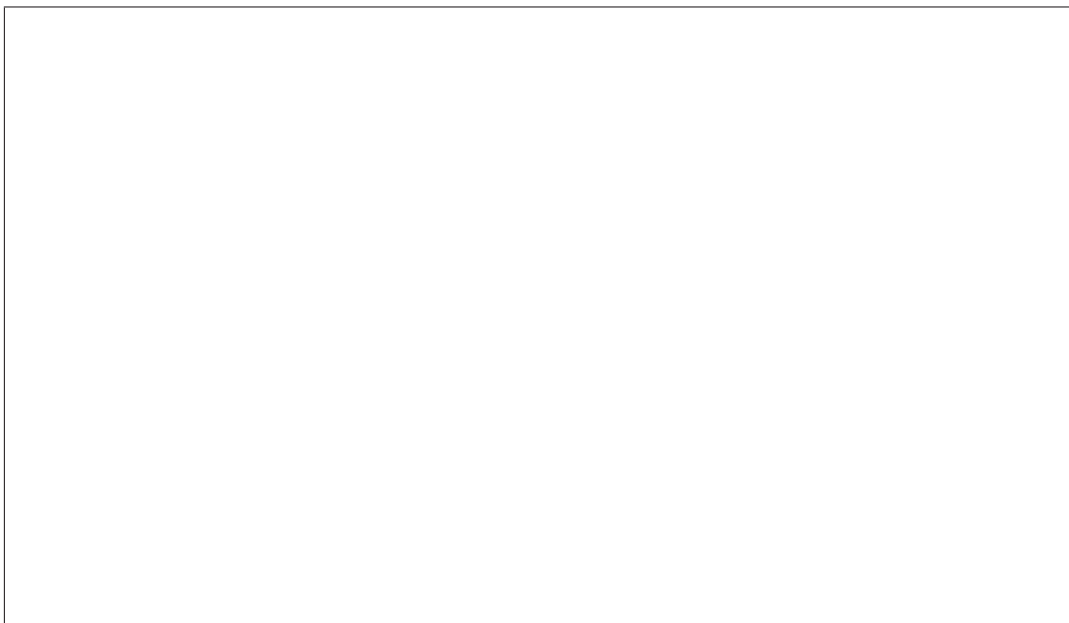
Note that if an object moves with **constant velocity**, this means that it has constant speed and direction, i.e., it travels in a straight line.

5.2.4 Acceleration

The **acceleration vector** is defined as

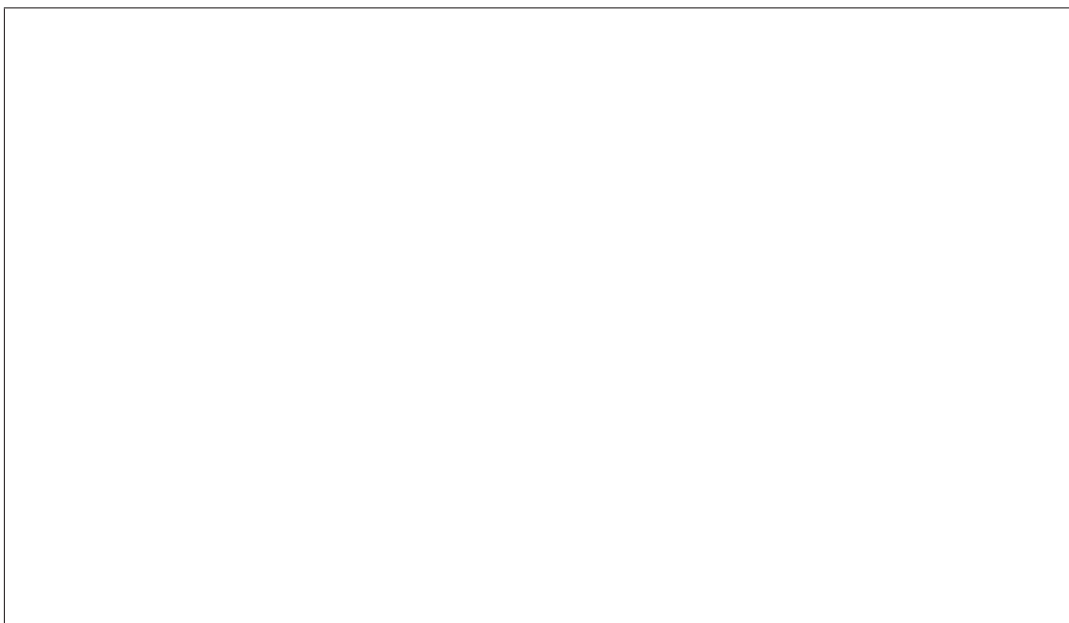
$$\mathbf{a}(t) = \frac{d\mathbf{v}}{dt} = \frac{d^2x}{dt^2}\mathbf{i} + \frac{d^2y}{dt^2}\mathbf{j} + \frac{d^2z}{dt^2}\mathbf{k}.$$

Example: The velocity of an object is $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$ and at $t = 0$ the object passes through $(-2, 1, 0)$. Find the position vector.



Example: Find the position and velocity vectors of a car moving along the helical path (e.g., the Indooroopilly Shopping Centre carpark)

$$x = \cos 5t, \quad y = \sin 5t, \quad z = t.$$



5.2.5 Who wants to be a millionaire?

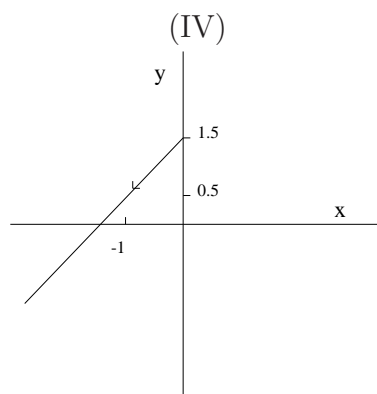
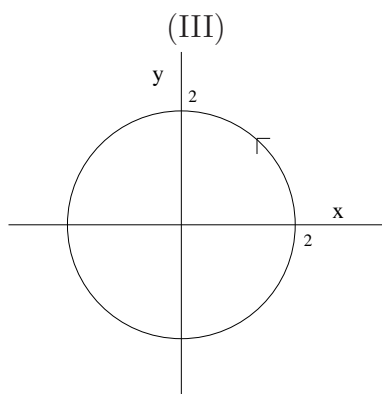
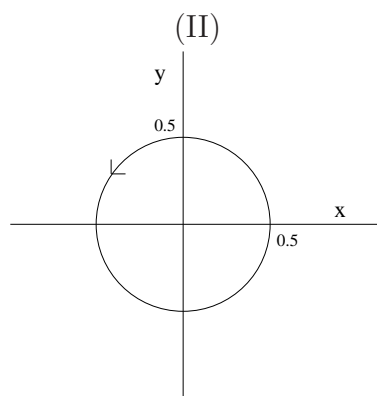
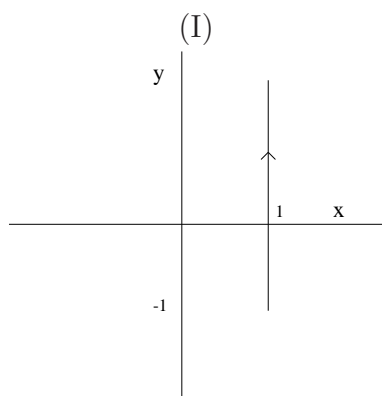
At a recent mathematics trivia night hosted by Dr. Viv, she asked the assembled mathematicians which position vector fits which curve? Unsurprisingly, none of the professors gave the correct answer. Can you do better? And can you locate the points corresponding to $t = 0$?

(a) $\mathbf{r}(t) = (-t, 3/2 - t)$

(b) $\mathbf{r}(t) = (1, t - 1)$

(c) $\mathbf{r}(t) = \frac{1}{2}(-\sin t, \cos t)$

(d) $\mathbf{r}(t) = 2(\cos 3t, \sin 3t)$.



5.2.6 Extra reading: stone on a string

Imagine you spin a stone on a string above your head so that it moves in a circle of radius 1m at a height of 2m above the ground at a constant speed with period π seconds. If the string breaks, find the position, velocity and acceleration vectors of the stone after the string breaks.

Before the string breaks

$$\mathbf{r}(t) = \cos 2t\mathbf{i} + \sin 2t\mathbf{j} + 2\mathbf{k}$$

$$\mathbf{v}(t) = -2\sin 2t\mathbf{i} + 2\cos 2t\mathbf{j},$$

so that $\mathbf{r}(0) = \mathbf{i} + 2\mathbf{k}$ and $\mathbf{v}(0) = 2\mathbf{j}$.

After the string breaks the stone moves under gravity

$$m\mathbf{a}(t) = -mg\mathbf{k} \quad \Rightarrow \quad \mathbf{a}(t) = -g\mathbf{k} \quad \Rightarrow \quad \frac{d\mathbf{v}}{dt} = -g\mathbf{k}.$$

Hence

$$\mathbf{v} = -gt\mathbf{k} + \mathbf{v}(0) = 2\mathbf{j} - gt\mathbf{k}.$$

Since $\frac{d\mathbf{r}}{dt} = \mathbf{v}$ this in turn implies that

$$\mathbf{r}(t) = 2t\mathbf{j} - \frac{1}{2}gt^2\mathbf{k} + \mathbf{r}(0)$$

so that

$$\mathbf{r}(t) = \mathbf{i} + 2t\mathbf{j} + \left(2 - \frac{1}{2}gt^2\right)\mathbf{k}.$$

5.3 Line Integrals and Work Done

By the end of this section, you should understand the following concepts:

- How to use a line integral to calculate the arc length of a curve.
- How to use a line integral to calculate the work done by a field of force.

The content of this chapter is covered in Stewart over a few sections: arc length in Sections 9.1 and 11.4 and general line integrals in 17.2.

5.3.1 Length of a curve

Say we have a curve in \mathbb{R}^3 parametrised by

$$x = x(t), \quad y = y(t), \quad z = z(t), \quad a \leq t \leq b.$$

We aim to calculate the length of the curve. This is commonly referred to as the **arc length**. We first approximate the curve as a series of straight lines:



In this case, with Δs_i representing the length of a small segment of the curve corresponding to a small change Δt in the parameter, and given $\mathbf{r}(t) = (x(t), y(t), z(t))$ we have the approximation

$$\Delta s_i \approx \left| \frac{d\mathbf{r}}{dt} \right| \Delta t.$$

The length of the curve is approximately the sum of the lengths of each line segment:



This approximation improves for smaller (and hence more) line segments. In the limit when the number of line segments tend to infinity, we obtain an exact result, giving the arc length formula

$$\text{arc length} = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2} dt.$$

This formula is often written more succinctly using the notion of a **line integral**. If the function $s(t)$ gives the length of the curve C between $\mathbf{r}(a)$ and $\mathbf{r}(t)$ — so that the arclength we are after is nothing but $s(b)$ — then

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}.$$

Hence the length of the arc between $\mathbf{r}(a)$ and $\mathbf{r}(b)$ is given by the line integral

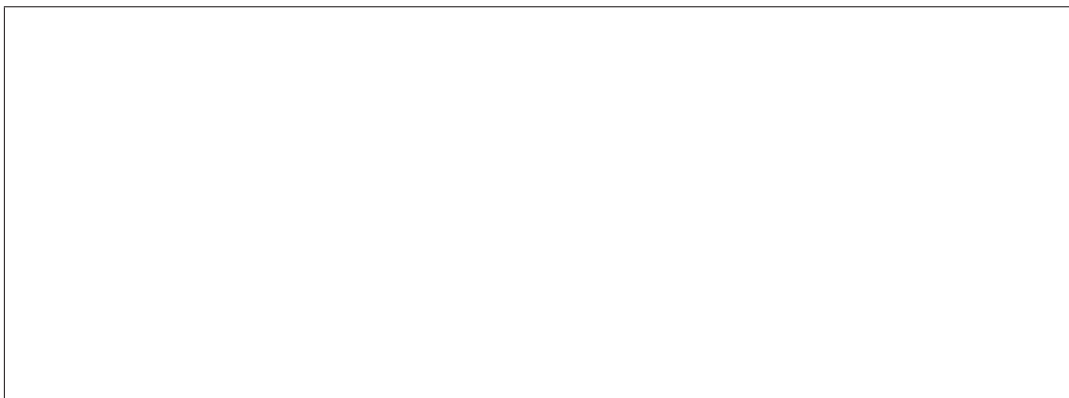
$$\text{arc length} = \int_C ds.$$

Example: Find the length of the helix
$$\begin{cases} x = \cos 10t \\ y = \sin 10t \\ z = t \end{cases} \quad 0 \leq t \leq \pi.$$

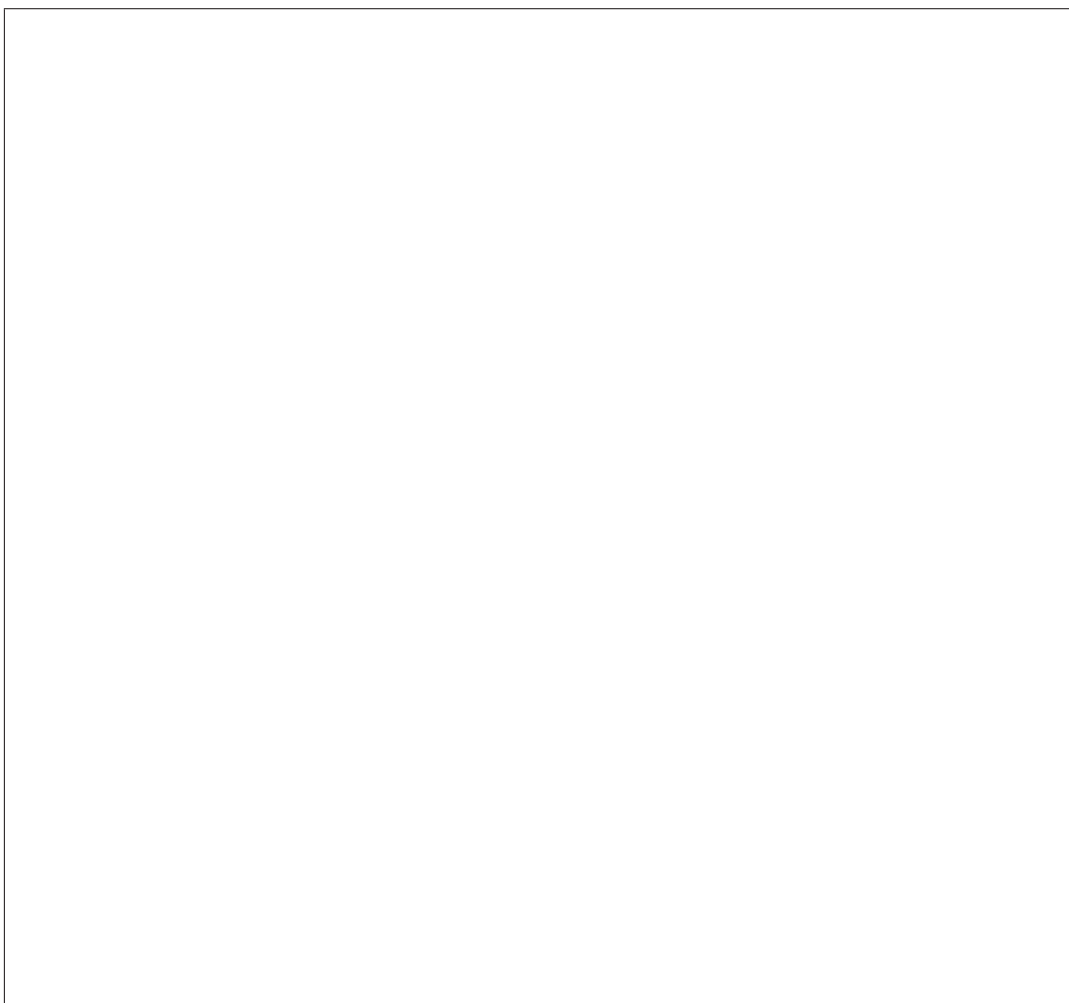


5.3.2 Work done by a constant force

In one dimension, the work done by a constant force F in moving an object along a straight line of length d is $W = Fd$.



In two or three dimensions, the work done by a constant force in moving a particle along a straight line from P to Q is $W = F \cdot \overrightarrow{PQ}$.



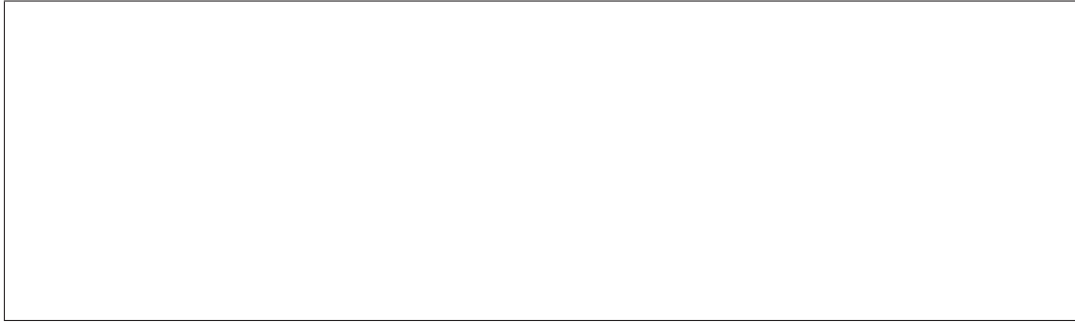
5.3.3 Work done over a curve

We now consider the more general case of the work done by a force field

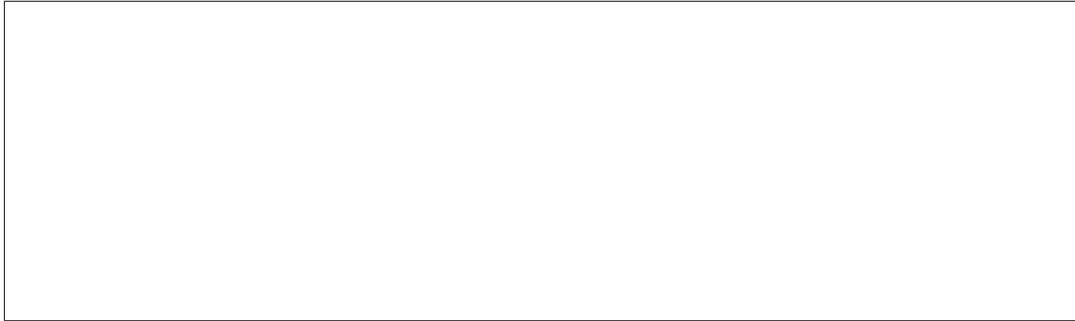
$$\mathbf{F}(x, y, z) = F_1(x, y, z)\mathbf{i} + F_2(x, y, z)\mathbf{j} + F_3(x, y, z)\mathbf{k}$$

which moves an object along a curve C .

First we give an approximation: divide C into n arcs, so that the i th arc has length Δs_i .



We approximate Δs_i as a straight line (for Δs_i small enough), and \mathbf{F} as constant over Δs_i by evaluating \mathbf{F} at a specified point P_i on the arc. Let $\mathbf{T}(P_i)$ be a unit tangent vector to the curve C at the point P_i . We then approximate the direction over the arc as the vector $\mathbf{T}(P_i)\Delta s_i$.



Thus the work done over the i th arc is approximately

$$W_i \approx \mathbf{F}(P_i) \cdot \mathbf{T}(P_i)\Delta s_i.$$

Summing up over all arcs gives an approximation to the total work done:

$$W \approx \sum W_i \approx \sum \mathbf{F}(P_i) \cdot \mathbf{T}(P_i)\Delta s_i.$$

Taking the limit $\Delta s_i \rightarrow 0$ gives

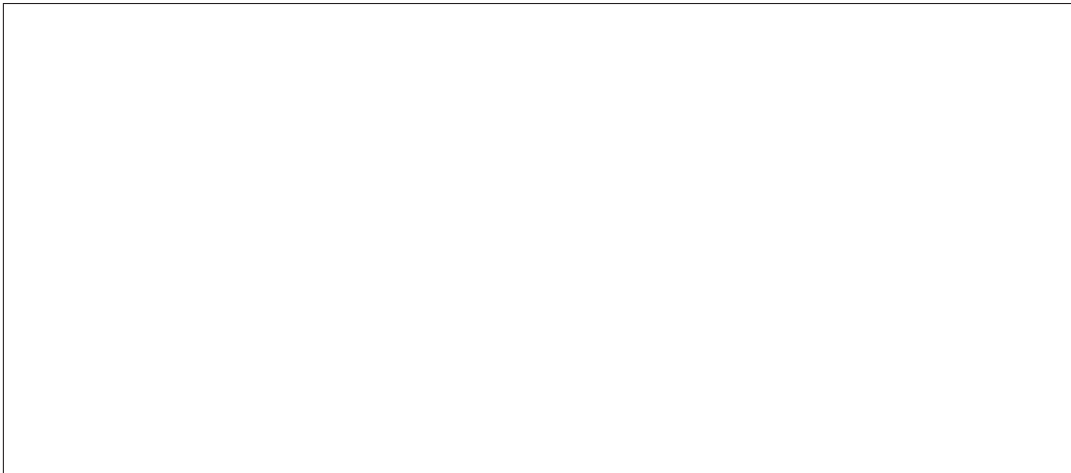
$$W = \int_C \mathbf{F} \cdot \mathbf{T} ds.$$

This last integral, which is an integral over the **curve** C , is again known as a **line integral**.

To evaluate the line integral, we use a parametrisation of the curve C . Let C be parametrised by

$$\mathbf{r}(t) = (x(t), y(t), z(t)), \quad a \leq t \leq b.$$

We let Δt represent a small change in t corresponding to Δs_i . Let P_i correspond to parameter value t_i .

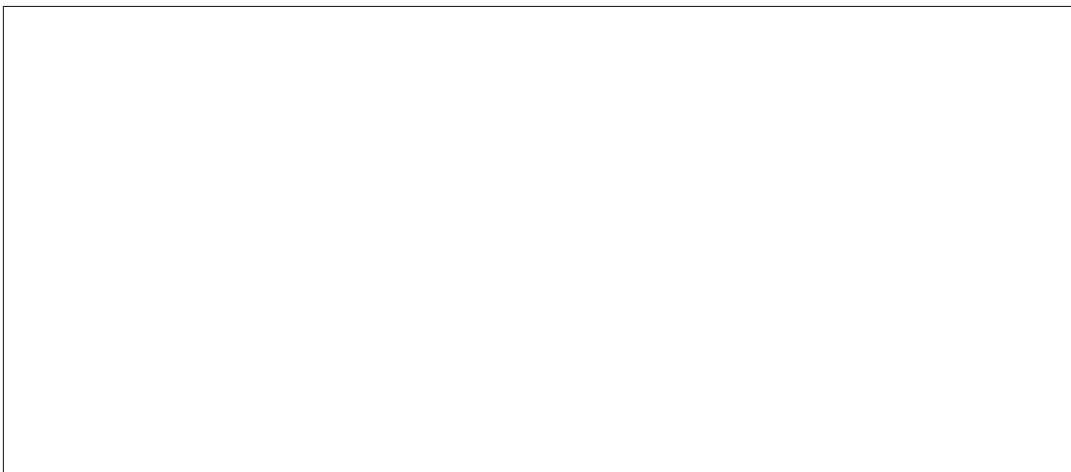


As we saw previously, as $\Delta s_i \rightarrow 0$,

$$\Delta s_i \approx |\mathbf{r}'(t_i)|\Delta t,$$

and we can write the unit tangent vector to the curve at the point P_i in terms of the parametrisation as

$$\mathbf{T}(P_i) = \frac{\mathbf{r}'(t_i)}{|\mathbf{r}'(t_i)|}.$$



Hence the work done over the i th arc is approximately

$$W_i \approx \mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i)\Delta t.$$

Summing up over all arcs gives an approximation to the total work done:

$$W \approx \sum W_i \approx \sum \mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i) \Delta t.$$

Finally, taking the limit as $\Delta t \rightarrow 0$, we have

$$W = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt.$$

It is common to express these integrals as

$$\int_C \mathbf{F} \cdot d\mathbf{r}.$$

This is merely a notational convenience. This notation means

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt,$$

where C is parametrised by $\mathbf{r}(t)$, $a \leq t \leq b$.

Example: Find the work done by moving a mass of 2 kg along an arc of the circle $x^2 + z^2 = h^2$ from $(-h, 0, 0)$ to $(0, 0, h)$. Does the direction we choose matter?



5.3.4 Extra reading: length of a spiral

Example: Find the length of the spiral

$$\begin{cases} x(t) = e^{-t/10} \cos t \\ y(t) = e^{-t/10} \sin t \end{cases} \quad \text{for } t \geq 0.$$

Solution:

We compute

$$\begin{aligned} \mathbf{r}(t) &= e^{-t/10}(\cos t \mathbf{i} + \sin t \mathbf{j}) \\ \mathbf{v}(t) &= e^{-t/10} \left(-\frac{1}{10} \cos t - \sin t \right) \mathbf{i} + e^{-t/10} \left(-\frac{1}{10} \sin t + \cos t \right) \mathbf{j} \\ \|\mathbf{v}(t)\| &= \sqrt{\left(\frac{1}{100} + 1 \right) e^{-t/5}} = \frac{1}{10} \sqrt{101} e^{-t/10}. \end{aligned}$$

Hence

$$\text{arc length} = \int_0^{\infty} \|\mathbf{v}(t)\| dt = \frac{1}{10} \sqrt{101} \int_0^{\infty} e^{-t/10} dt.$$

The **improper integral** can be computed as

$$\begin{aligned} \int_0^{\infty} e^{-t/10} dt &= \lim_{A \rightarrow \infty} \int_0^A e^{-t/10} dt \\ &= -10 \lim_{A \rightarrow \infty} [e^{-t/10}]_0^A \\ &= -10 \lim_{A \rightarrow \infty} (e^{-A/10} - 1) \\ &= 10. \end{aligned}$$

Therefore the arc length of the spiral is $\sqrt{101}$.

Example: Find the length of the parabola $y = x^2$ for x between 0 and 2.

Solution: First we parametrise the parabola as $x = t$, $y = t^2$ for $0 \leq t \leq 2$. Hence

$$\begin{aligned} \mathbf{r}(t) &= (t, t^2) \quad \Rightarrow \quad \mathbf{v}(t) = \mathbf{i} + 2t\mathbf{j} \\ \Rightarrow \|\mathbf{v}(t)\| &= \sqrt{1 + (2t)^2} \quad \Rightarrow \quad \text{arc length} = \int_0^2 \sqrt{1 + (2t)^2} dt. \end{aligned}$$

(If you are really clever you can perhaps show that this gives $\sqrt{17} + \operatorname{arcsinh}(4)/4$.)

5.4 Gradient Fields

5.4.1 Main points

By the end of this section, you should understand the following:

- The relation between the gradient field and the potential function.
- That conservative fields and gradient fields are equivalent.
- The potential function will give you an easy method to calculate line integrals for gradient fields (the fundamental theorem for line integrals).

Gradient fields are a special type of force field which are conservative, and they give a relatively easy way to determine if the field is conservative. Stewart covers this in Sections 17.1 and 17.3.

5.4.2 Conservative fields

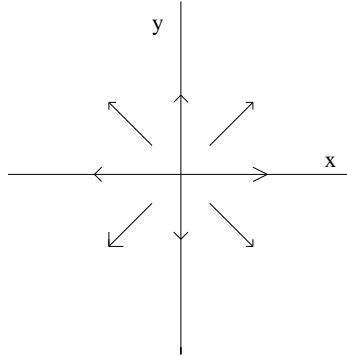
If $\int_A^B \mathbf{F} \cdot d\mathbf{r}$ is independent of the path taken, then \mathbf{F} is called a **conservative field**.

Gravity is a conservative field because

$$\int_A^B \mathbf{F} \cdot d\mathbf{r} = -mg \int_A^B dz = -mg(z(B) - z(A)),$$

and it does not matter how you get from A to B . Any constant field must be conservative by the same reasoning.

Take the field $\mathbf{F} = x\mathbf{i} + y\mathbf{j}$, which you can picture as a set of vectors $\mathbf{F}(x, y)$ at each point (x, y) .



Since

$$\mathbf{F} \cdot d\mathbf{r} = (x\mathbf{i} + y\mathbf{j}) \cdot (dx\mathbf{i} + dy\mathbf{j}) = x dx + y dy,$$

we get

$$\begin{aligned} \int_A^B \mathbf{F} \cdot d\mathbf{r} &= \int_A^B x dx + y dy \\ &= \frac{1}{2} \int_A^B d(x^2 + y^2) = \frac{1}{2} [x^2 + y^2]_A^B. \end{aligned}$$

Once again \mathbf{F} is a conservative field. But it was chosen in rather a special way. \mathbf{F} is a **gradient field**, i.e., $\mathbf{F} = \nabla f$ for some f . In this case $f(x, y) = (x^2 + y^2)/2$ since

$$\mathbf{F} = \nabla \left(\frac{x^2}{2} + \frac{y^2}{2} \right) = (x, y).$$

Theorem: Any gradient field \mathbf{F} is a conservative and hence $\int_A^B \mathbf{F} \cdot d\mathbf{r}$ is path-independent.

To show this, let $\mathbf{F} = \nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$. Then $\mathbf{F} \cdot d\mathbf{r} = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = df$. So

$$\int_A^B \mathbf{F} \cdot d\mathbf{r} = \int_A^B df = f(B) - f(A).$$

Finding the work done against a conservative field really comes down to finding a function f such that $\mathbf{F} = \nabla f$.

Example: Find the work done by the electric field

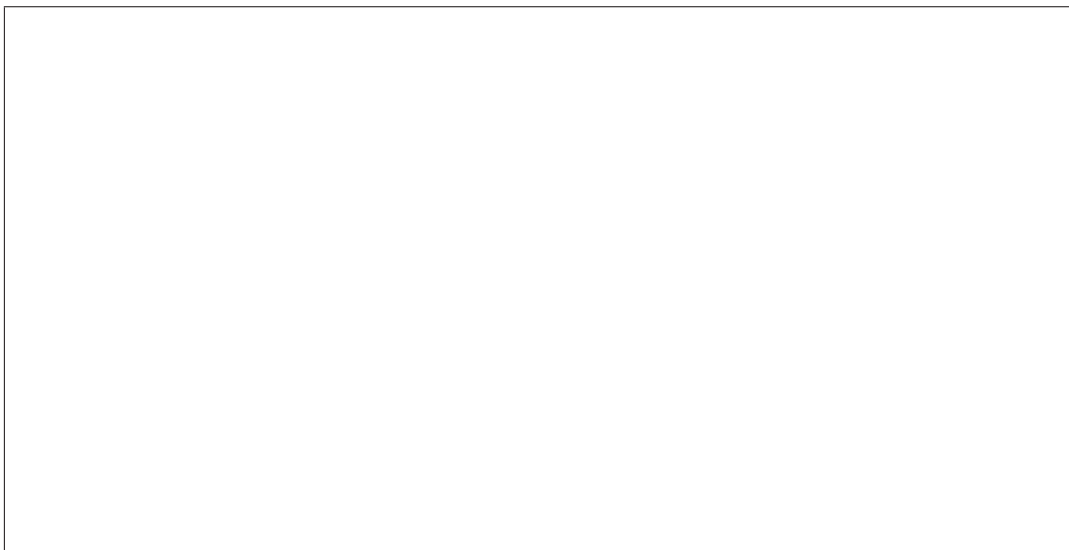
$$\mathbf{E} = \frac{\alpha(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})}{x^2 + y^2 + z^2}$$

along any path from $(1, 0, 0)$ to $(1, 1, 1)$.

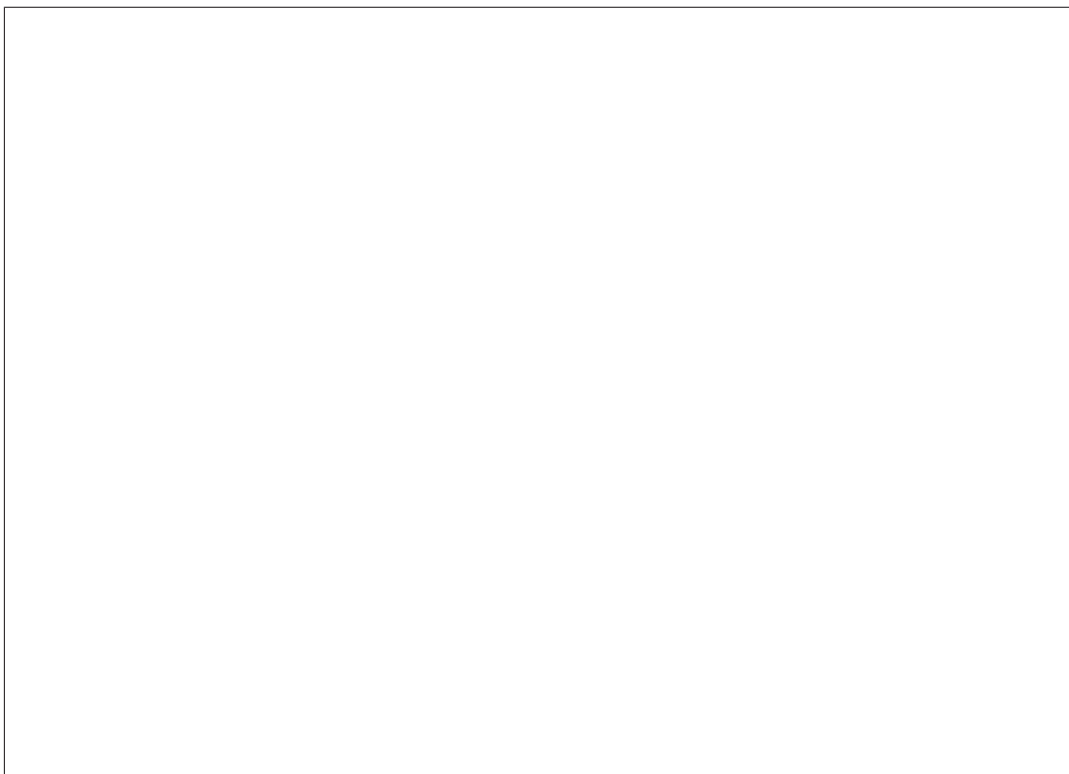


Example: Evaluate the work done by the force \mathbf{F} from $(0, 0)$ to $(1, 2)$, where

$$\mathbf{F} = \nabla f \quad \text{and} \quad f(x, y) = xy.$$



Example: Show that $\mathbf{F}(x, y) = (x + y, x)$ is a gradient field.



Example: Show that $\mathbf{F}(x, y) = \frac{(x + y)}{2} \mathbf{i} + \frac{y}{2} \mathbf{j}$ is not a gradient field.



5.5 Work Done or Path Integrals for Non-conservative Fields

After this section, you should know how to recognise a non-conservative field, and you should be able to perform parametrisations and calculate path integrals.

Line integrals for non-conservative fields are harder to evaluate because you must perform a parametrisation of the path. Section 17.2 in Stewart covers some useful theory and examples.

Path integrals for forces which are not conservative depend on the chosen path. To work them out we need to parametrise the path, by say t , and convert the integral to one over t .

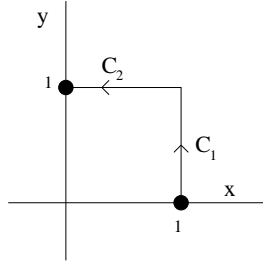
Let $\mathbf{F}(x, y) = F_1(x, y)\mathbf{i} + F_2(x, y)\mathbf{j}$ be a non-conservative field and let $x(t)$ and $y(t)$ provide a parametrisation of the path along which we wish to integrate. Then

$$\mathbf{F} \cdot d\mathbf{r} = F_1(x, y)dx + F_2(x, y)dy$$

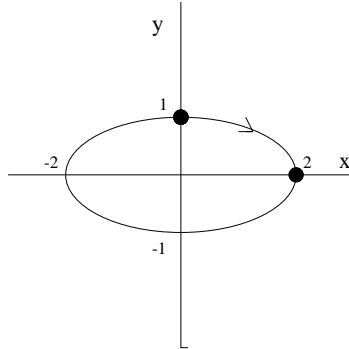
and

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \left(F_1(x(t), y(t)) \frac{dx}{dt} + F_2(x(t), y(t)) \frac{dy}{dt} \right) dt.$$

Example: Evaluate $\int_C \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F}(x, y) = \left(\frac{x+y}{2}\right)\mathbf{i} + \frac{y}{2}\mathbf{j}$ and the path is as shown in the figure below.



Example: Evaluate $\int_C \mathbf{F} \cdot d\mathbf{r}$, where $\mathbf{F} = (x + y)\mathbf{i} + y\mathbf{j}$ and C is the arc of the ellipse $\frac{x^2}{4} + y^2 = 1$ from $(0, 1)$ to $(2, 0)$ as shown in the figure below.



6 MATLAB Guide

6.1 Introduction

6.1.1 Getting started

- Log in using your library username and password. Ask your tutor if you have a problem.
- Open MATLAB just like any other Windows program.
Start→**Programs**→**MATLAB**→**MATLAB R2008b**
- Save all files to your student directory on **H:/** or your own USB flash drive. Files saved onto individual computers or **S:/** will be deleted at the end of semester.

As MATH1051 is a companion for MATH1052, it is generally assumed that you will still be in the process of developing a basic knowledge of MATLAB. Such basic knowledge includes navigating to the correct directory, executing functions and defining variables. Many of these fundamentals will be re-introduced in this document, but some simple operations, etc. may be used without explanation. If you need further guidance then consult MATLAB Help or the MATH1051 notes

6.1.2 Getting help

There are a number of resources available should you need assistance when using MATLAB. The most easily accessible is MATLAB's own built-in Help system.

- Clicking the blue question mark button in the toolbar will open the Help window, as will typing `doc` in the command window. You can search for the topic in question or navigate using the panel on the left.
- If you know the name of the function or operation you need help with, typing `help <function name>` in the command window will display a short description of the function and how to use it. Typing `doc <function name>` will open the Help window and navigate to the page dedicated to that function or operation.

If you have trouble finding what you need with MATLAB's Help system, there is a great deal of documentation and useful articles available online. The Mathworks

website (<http://www.mathworks.com/products/matlab/>) has a lot of helpful descriptions and solutions which go beyond what is available in the MATLAB Help, as well as m-files written and uploaded by users. A Google search often turns up results as well.

6.2 The Basics

The MATLAB Command Window can be used like an advanced calculator. Mathematical operations like addition, subtraction, multiplication, division and exponentiation are executed just as with a calculator, using $+$, $-$, $*$, $/$ and $^$ respectively. MATLAB also has many more complicated functions built in, like `sin`, `cos`, `tan`, `log` and `exp`, as well as constants like π and i . Consult the MATH1051 notes or MATLAB Help for more details on these or other functions.

```
>> 1+6*2
ans =
    13
>> log(10)
ans =
    2.3026
>> exp(cos(pi))
ans =
    0.3679
```

Variables are defined using an equals sign, and can then be used just like regular numbers.

```
>> x = 7
x =
    7
>> 2^x
ans =
   128
```

Vectors and matrices are defined in MATLAB using square brackets. Spaces or commas separate columns and semicolons separate rows. MATLAB has many inbuilt operations for use with matrices, such as addition and multiplication, and more complex functions like `eig` and `rref`. Note that matrix indices start at 1, not 0 (so the top-left cell in a matrix is in the (1,1) position).

```
>> x = [1 2; 5 6]
x =
```



```
    1     2
    5     6
>> x^2
ans =
    11    14
    35    46
>> [2 3]*x
ans =
    17    22
>> x(1,2)
ans =
     2
```

Use the command `clear <variable name>` to remove a variable definition from MATLAB's memory, and the command `clear all` to remove all variables at once. Variables should usually be cleared in between MATLAB sessions to avoid using incorrect values.

6.3 M-files

M-files are programs written in MATLAB's own programming language. They are essentially a series of commands which are executed successively when the M-file is run. There are two main types of M-files in MATLAB:

- *M-file scripts* are sequences of commands which could normally be entered into the command window. They are usually intended to be executed a number of times, so time is saved by putting the commands into the M-file. The operations inside a M-file script have access to all the variables, etc. which have been saved in the command window.
- *M-file functions* begin with

```
function [output variables] = function_name(input variables)
```

and operate like ordinary functions such as `sin` or `log`. They can take input variables, perform operations on them, and produce output variables. The operations inside an M-file function *do not* have access to the variables saved in the command window – they can only access the input variables.

Anything followed by a % in a M-file is a *comment*, and will not affect the execution of the file. You should comment all your M-files to help yourself or others understand how the program works.

For example, put the following commands into a new M-file, and save it as **example_script.m**.

```
x = [1 2 3 4 5];
y = 2*x';
z = x*y
```

Now run this script by typing the command `example_script` in the command window:

```
>> example_script
z =
    110
```

Note that only the value of z was output, because the definitions of x and y were followed by semicolons (suppressing their output).

Put the following into an M-file and save it as **example_function.m**. It will evaluate the polynomial $4x^3 + 2x^2 - 10x - 7$.

```
function output = example_function(x)
% Evaluates the 3rd order polynomial at x
output = 4*x^3 + 2*x^2 - 10*x - 7;
```

Execute the function as you would any regular function in MATLAB:

```
>> example_function(4)
ans =
    241
```

Note that the name of the function (the word following the = sign at the top of the M-file) *must* match the name of the M-file itself.

6.3.1 If statements

An `if` statement in MATLAB is a special type of command which is executed only if a certain condition is satisfied. The general structure of an `if` statement is always

```
if 1st condition
    1st commands
elseif 2nd condition
```

```

    2nd commands
else
    3rd commands
end

```

The conditions above should be boolean statements; i.e. they should take the value 0 or 1. MATLAB has many of these built in; for example, `x == y` returns 1 if $x = y$ and 0 otherwise, and `x <= y` returns 1 if $x \leq y$ and 0 otherwise. If *1st condition* takes the value 1, then *1st commands* will be executed, and the `if` statement will be exited. Otherwise, the next condition will be checked, and so on. If none of the conditions take the value 1, then the commands following `else` will be executed. You can have as many `elseif` lines as you want (or none at all), and you do not need to have an `else` line (in which case MATLAB will do nothing if none of the conditions hold).

For example, the following function will return the absolute value of a real number, as well as stating whether it is positive, negative or zero.

```

function output = abs_val_check(x)
% Outputs the absolute value of x and states whether x
% is positive, negative or zero
if x > 0
    output = x;
    display('positive')
elseif x < 0
    output = -x;
    display('negative')
else
    output = 0;
    display('zero')
end

```

The function can now be executed in the command window in the usual way.

```

>> abs_val_check(5)
positive
ans =
    5

```

6.3.2 For and while loops

Loops allow MATLAB to perform repetitive operations automatically. There are two main types of loops:

- For loops instruct MATLAB to perform a fixed number of operations.
- While loops instruct MATLAB to repeatedly perform operations for as long as a certain condition holds.

The basic structure of a for loop is

```
for index = vector
    commands
end
```

For each value in the vector, the variable `index` is assigned that value, and `commands` are executed.

For example, the following function will calculate the sum of the first 10 integer powers of x .

```
function output = sum_powers(x)
% Evaluates the sum of the first 10 powers of x
temp = 0;
for j = 1:10
    temp = temp + x^j;
end
output = temp;
```

Evaluate the function in the usual way:

```
>> sum_powers(2)
ans =
    2046
```

While loops look a bit like if statements – their basic structure is

```
while condition
    commands
end
```

On the first iteration of the loop, MATLAB checks whether `condition` holds. If it does, then `commands` are executed, and the process is repeated. If on any iteration `condition` does not hold, then MATLAB exits the loop.

For example, the following function performs the same operation as the one above.

```
function output = sum_powers2(x)
% Evaluates the sum of the first 10 powers of x
temp = 0;
```

```

j = 1;
while j <= 10
    temp = temp + x^j;
    j = j + 1;
end
output = temp;

```

Here we are increasing the value of j by 1 in each iteration, and we finish the loop after $j = 10$.

There is a danger with `while` loops: if the condition *always* holds, then the loop will keep iterating forever. When writing `while` loops, make sure that there will always be some point at which the condition will fail. If you happen to find yourself running an infinite loop, use the command `Ctrl+c` to terminate it.

6.4 Visualising Functions

There are many situations, especially when dealing with multivariate calculus, when a function or a set of data is difficult to visualise. There are a number of ways to plot and view functions and data in MATLAB.

6.4.1 Ezplot

The `ezplot` command allows you to plot functions of one variable. You can control the domain and range if you wish, or let MATLAB assign the default domain of $[-2\pi, 2\pi]$. For example, to plot $y = x \sin(x)$ over the domain $[-\pi, \pi]$ and range $[-10, 10]$, use the command

```
>> ezplot('x*sin(x)', [-pi, pi, -10, 10])
```

Remember to insert the quotation marks ' ' around the function. To plot multiple functions at once, use the `hold on` command before you start plotting. You can change the domain and range of the figure without re-plotting by using the `axis` command; for example,

```
>> axis([0, pi, -10, 10])
```

You can edit the plot or provide a title and axis labels by using the **Property Editor** in the **View** menu.

The `ezplot` command can also be used to plot points of the form $f(x, y) = 0$. For example, to plot the points which satisfy $\sin(x) + \cos(y) = 0$ in the domain and range $[-2\pi, 2\pi]$, use the command

```
>> ezplot('sin(x)+cos(y)', [-2*pi, 2*pi, -2*pi, 2*pi])
```

If you wish to open a new figure window to plot functions on a new axis, use the `figure` command.

The `ezplot` command, despite only being able to plot functions of one variable, can help in visualising functions of two or more variables. This is because it can be used to plot *cross sections*. For example, to plot cross sections of a function $z = f(x, y)$, use `hold on` and simultaneously plot the functions $z = f(x, -2)$, $z = f(x, -1)$, \dots , $z = f(x, 2)$. Using different colours, line types and/or labels for each plot may make it easier to visualise the cross sections. Consider also plotting cross sections where x instead of y is held constant (but not on the same axis!).

For example, to plot cross sections of the function $z = f(x, y) = \sin(x) \cos(y)$ for $-2\pi \leq x \leq 2\pi$ and $-2 \leq y \leq 2$, you could use the commands

```
>> hold on
>> ezplot('sin(x)*cos(-2)')
>> ezplot('sin(x)*cos(-1)')
>> ezplot('sin(x)*cos(0)')
>> ezplot('sin(x)*cos(1)')
>> ezplot('sin(x)*cos(2)')
```

6.4.2 Ezcontour

Cross sections may help in visualising a multivariate function, and can often be done easily by hand. However there are many cases where they only produce very confusing or uninformative pictures. *Contour diagrams* are sometimes a better way of visualising multivariate functions. The `ezcontour` command is the easiest way to produce contour diagrams for functions. It is used in much the same way as `ezplot`. The default x and y domains are both $[-2\pi, 2\pi]$, but can be controlled in the same way as for `ezplot`.

```
>> ezcontour('sin(x)*cos(y)')
>> ezcontour('sin(x)*cos(y)', [-10 10 -10 10])
```

The lines in the contour diagram represent points where z takes the same value. The colours of the lines represent the value of z on that line. Use the command `colorbar` or go to **Insert** → **Colorbar** to insert a colourbar, which will show the values of z corresponding to the colours.

The `ezcontourf` command operates in the same way as `ezcontour`, but colours in the gaps between the contours.

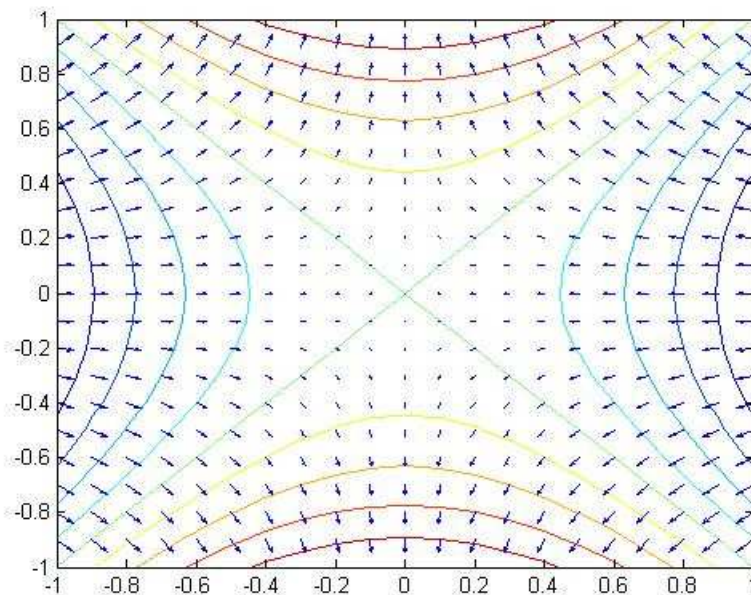
On the MATH1052 webpage is a M-file called **MATH1052quiver.m**. This script looks for a M-file function called **quivfunc.m** (which you must define yourself), and produces a contour diagram for the function, as well as an *arrow plot*. The arrow plot consists of many small arrows which point in the direction of the gradient at their location (ie. *up* the slope of the function at that point). The **quivfunc** function must accept inputs x and y .

For example, you could define **quivfunc** as

```
function output = quivfunc(x,y)
output = 2*x^2 - y^2;
```

Running **MATH1052quiver** will then produce a contour diagram and arrow plot:

```
>> MATH1052quiver
```



You can alter the x and y axis ranges in the first few lines of the **MATH1052quiver** M-file:

```
xmin = -1;
xmax = 1;
ymin = -1;
ymax = 1;
```

6.4.3 Ezsurf and ezmesh

In addition to cross sections and contour diagrams, MATLAB can display multivariate functions using surface plots. There are two main MATLAB functions for plotting surfaces: `ezsurf` and `ezmesh`. Their use is virtually the same as for `ezcontour`: the function is specified in exactly the same way, as is the domain(s). The `ezsurf` command plots a coloured surface, while `ezmesh` plots a coloured mesh. In both cases the colour at any point represents the value of the function at that point; a colourbar can be inserted to aid in viewing the plot.

For example, try using

```
>> ezsurf('sin(x)*cos(y)')
>> ezmesh('sin(x)*cos(y), [-10 10 -10 10]')
```

Note that Figure Windows have a button called **Rotate 3D**. Clicking this button allows you to use the cursor to rotate a plot in three dimensions. Doing so may help you visualise the function more clearly.

6.4.4 Functions of three variables

Since functions of three variables require four dimensions in order to be plotted properly, they are often hard to visualise effectively. One helpful method essentially uses the four-dimensional analog of cross sections. To visualise a function $w = f(x, y, z)$, we can make a number of surface plots where z (or x or y) is held constant.

For example, consider the function $w = f(x, y, z) = \frac{x-y}{x^2+y^2+z^2} \sin(x-z)$ for $-5 \leq x \leq 5$, $-5 \leq y \leq 5$ and $0 \leq z \leq 2$. We can hold z constant at a number of values between 0 and 2, and produce surface plots for each resulting two-variable function: $f(x, y, 0)$, $f(x, y, 0.5)$, $f(x, y, 1)$, $f(x, y, 1.5)$ and $f(x, y, 2)$. Obviously we don't want all these plots to be on the same axis, so use the `figure` command before each plot to create a new Figure Window.

```
>> figure
>> ezsurf('(x-y)/(x^2+y^2+0^2)*sin(x-0)')
>> figure
>> ezsurf('(x-y)/(x^2+y^2+0.5^2)*sin(x-0.5)')
```

and so on.

6.5 Visualising Data

The commands described in the previous section are useful for functions which can be easily described using a number of variables. However, there will be times when

you wish to visualise data which cannot be described using a simple function, or for which such a function is unknown. MATLAB has several commands for plotting and viewing such data.

6.5.1 Plot and plot3

MATLAB's `plot` command is used for plotting two-dimensional data points. To plot the points $(x_1, y_1), \dots, (x_n, y_n)$, use the command `plot(x,y)` where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$.

For example, to plot the points $(1, 2), (4, 2), (3, 1), (7, 4), (2, -3)$, use

```
>> x = [1,4,3,7,2];
>> y = [2,2,1,4,-3];
>> plot(x,y)
```

The `plot` command can take an optional third argument, consisting of a string of letters or symbols which tells MATLAB how to plot the data. For example, `plot(x,y,'ro')` will plot the points as red circles, and `plot(x,y,'g--')` will plot the points connected by a green dotted line. See MATLAB's help files for the full list of plotting options.

Three-dimensional data can also be plotted in MATLAB: simply use `plot3` instead of `plot`, and use vectors x, y and z . The same optional arguments are available. For example, to plot the points $(1, 2, 3), (2, 4, 2), (4, 3, -1), (3, 2, 5)$ as black stars, use the commands

```
>> x = [1,2,4,3];
>> y = [2,4,3,2];
>> z = [3,2,-1,5];
>> plot3(x,y,z,'k*')
```

6.5.2 Contour, surf and mesh

Just as `plot` is the point-wise analog of `ezplot`, the other plotting commands described in the previous section have point-wise versions. They are `contour`, `surf` and `mesh`. These work slightly differently to `plot3`; instead of taking three vectors as arguments, they take two vectors x, y and a matrix Z . These arguments describe a *table of values* for the data: x and y respectively give labels for the columns and rows of Z , producing a set of (x, y, z) points. For example, consider the following table:

	-1	0	1	2
-1	-2	-1	0	1
0	-1	0	1	2
1	0	1	2	3
2	1	2	3	4

Here, x and y are both $[-1, 0, 1, 2]$, and Z is the 4×4 matrix inside the table. Using `surf(x,y,Z)` would instruct MATLAB to create a surface using all the (x, y, z) points, where x and y come from the column and row labels respectively, and z comes from the values inside the table. So the points plotted would be $(-1, -1, -2), (0, -1, -1), (1, -1, 0), \dots, (2, 2, 4)$. If x is of length n and y is of length m , then Z must be a $m \times n$ matrix.

So to produce a surface plot of the above table, one would use the commands

```
>> x = [-1, 0, 1, 2];
>> y = x;
>> Z = [-2, -1, 0, 1; -1, 0, 1, 2; 0, 1, 2, 3; 1, 2, 3, 4];
>> surf(x,y,Z)
```

As with `ezsurf`, a colourbar can be inserted to assist in viewing the plot.

The `surf` command can in fact be used with only the argument Z ; in this case, MATLAB assumes that x and y are the vectors $[1, 2, \dots, n]$ and $[1, 2, \dots, m]$, where Z is a $m \times n$ matrix.

The `mesh` and `contour` commands function in exactly the same way as `surf`. As one would expect, `mesh` produces a mesh plot of the data, and `contour` produces a contour diagram. A coloured-in version of `contour`, called `contourf`, also exists.

6.6 Animations

The final method of presenting multivariate data is with animations. This essentially involves plotting a sequence of cross sections (or the three-dimensional analog thereof), and using each plot as a single frame. Animations are particularly useful if one of the independent variables is time.

Basically the method for producing an animation is:

- Write a loop which will vary t (or another independent variable) over some range;
- For each value of t , produce a plot;

- Use the command `M(j) = getframe` at each plot to store each image into the ‘vector’ M ;
- Use the command `movie(M,n)` to play the movie n times.

For example, the following will produce an animation of the function $f(x, t) = \cos(x)\cos(t)$ for $0 \leq x \leq 6.3$ and $0 \leq t \leq 20$. We cannot use `ezplot` here because MATLAB will be unable to correctly plot the function (it has two variables); we will instead use `plot`, with the points close enough together as to make the curve sufficiently smooth. The outermost loop (with index j) will cycle through the t values, and the inner loop (index k) will cycle through x values.

```
% MATLAB program to produce animation for cos(x)cos(t)
for j=1:101
    t=(j-1)/5;
    for k=1:64
        x(k)=(k-1)/10;
        z(k)=cos(x(k))*cos(t);
    end
    plot(x,z);
    axis([0,6.3,-2,2]);
    M(j)=getframe;
end
movie(M,5)
```

Note that we could not start the j and k loops at 0, because both of those indices were used to index vectors: j was used to index M and k was used to index x and z . MATLAB *always* starts matrix indices at 1, not 0; so we had to start those indices at 1 and then compensate by setting $t = (j - 1)/5$ and $x(k) = (k - 1)/10$.

The process is much the same with three-dimensional ‘cross sections’. The only major difference is that there will now be three nested `for` loops, instead of two: one each to cycle through t , x and y .

For example, the following will produce an animation for $f(x, y, t) = \cos(x)\cos(y)\cos(t)$, with $0 \leq x \leq 6.3$, $0 \leq y \leq 6.3$ and $0 \leq t \leq 20$.

```
% MATLAB program to produce animation for cos(x)cos(y)cos(t)
for j=1:101
    t=(j-1)/5;
    for k=1:64
        x(k)=(k-1)/10;
        for m=1:64
```

```

        y(m)=(m-1)/10;
        Z(m,k)=cos(x(k))*cos(y(m))*cos(t);
    end
end
surf(x,y,Z);
axis([0,6.3,0,6.3,-1,1]);
M(j)=getframe;
end
movie(M,5)

```

6.7 Optimisation

6.7.1 Using `fminsearch`

An important objective of much mathematical analysis is to locate global maxima and minima of multivariate functions. MATLAB contains a powerful inbuilt command called `fminsearch`, which is very effective at finding maxima and minima.

The application of `fminsearch` is as follows:

```
[xval,fval] = fminsearch('functionname',x0)
```

where:

- `functionname` is the name of the M-file function you wish to minimise; it must take a single vector as input.
- `x0` is the initial guess for the location of the minimum; it must be a vector of the same dimensions as the input of `functionname`.
- `xval` is the the location of the minimum.
- `fval` is the value of the function at the minimum.

For example, consider the function $f(x, y) = (x - 1)^2 + y^2 + 3$. We first need to write the M-file function:

```

function output = myfun(input)
% input is a vector: we will separate its elements
% for simplicity
x = input(1);
y = input(2);
output = (x-1)^2 + y^2 + 3;

```

Then if our initial guess for the (x, y) coordinate of the minimum is $(1, 2)$, we would use the command

```
>> [xval,fval] = fminsearch('myfun',[1,2])
xval =
    1.0000    0.0000
fval =
    3.0000
```

So the (x, y) coordinate of the minimum is $(1, 0)$ and the value of the function at that point is 3.

Note that you may have to try a number of different initial guesses before you find the *global* minimum.

To *maximise* a function, you need to find the minimum of the *negative* of the function.

6.7.2 Lines of best fit

When given a set of multi-dimensional data, we often wish to fit it to some kind of function. This can help us understand the relationships between different properties and make predictions for future data. A line of best fit is the simplest way to do so. While MATLAB has its own built-in functions for doing this (for example, look up `polyfit`), we will demonstrate here how to use `fminsearch` to fit a line to two-dimensional data.

Suppose we have a collection of two-dimensional points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ and we wish to find a line of best fit of the form $f(x) = y = mx + c$ (where x and y are the independent and dependent variables respectively). The most accurate such line is the one which minimises the errors between the actual data points and the corresponding points on the line; that is, it must minimise the errors $f(x_i) - y_i = mx_i + c - y_i$. This does not quite make sense: minimising the errors could mean letting them all go to $-\infty$. We actually wish to minimise the *absolute value* of the errors. This can be done, but it turns out to be easier to minimise the *squares* of the errors, which are also always positive.

This is not quite specific enough – we may have many data points, so how can we effectively minimise the square of *each* error? In fact, we must minimise the *sum* of the squares of the errors:

$$\sum_{i=1}^n (mx_i + c - y_i)^2$$

From here, the rest is easy. We need to write a function which takes as input m and c (in a single vector), and outputs the above sum. This function can then be fed into `fminsearch`, which will find the appropriate values for m and c which minimise this sum of squares.

For example, say our experimental data consists of the (x, y) points

$$(20, 3), (17, 5), (21, 2), (19, 4), (16, 5), (23, 2)$$

Then our function may look like the following:

```
function output = best_fit(input)
% Finds the m and c values for a line of best fit
m = input(1);
c = input(2);
x = [20 17 21 19 16 23];
y = [3 5 2 4 5 2];
temp = 0;
for i = 1:length(x)
    temp = temp + (m*x(i) + c - y(i))^2;
end
output = temp;
```

and to find the m and c values for the line of best fit, we apply `fminsearch` in the usual way.

```
>> [coeff,fval] = fminsearch('best_fit',[0,1])
coeffs =
    -0.5100    13.3600
fval =
    0.83
```

So the line of best fit for the above example is $f(x) = y = -0.51x + 13.36$. Plotting the data points and this line on the same axis demonstrates that this is indeed a very good line of best fit.

This method works just as well for higher-dimensional examples. For example, a plane of best fit would have the equation $f(x, y) = z = mx + qy + c$, where x and y are the independent variables and z is the dependent variable. The M-file function would then take as input m, q and c , and output the sum of the squares of the errors:

$$\sum_{i=1}^n (mx_i + qy_i + c - z_i)^2$$

6.8 Differential Equations

6.8.1 Slope fields

A first-order differential equation $dy/dx = f(x, y)$ determines the gradient of any solution curve $y(x)$ at any (x, y) point. A slope field is essentially a visual representation of this differential equation: it uses small lines to indicate the gradient at (x, y) points on a cartesian plane. Given an initial condition, you can use a slope field to visually ‘solve’ a differential equation, by following the direction of the field from the given point.

The M-files **field.m** and **func.m** can be downloaded from the MATH1052 website. The **field.m** file is a MATLAB function that produces a slope field for $dy/dx = f(x, y)$, and **func.m** is used to describe the function $f(x, y)$.

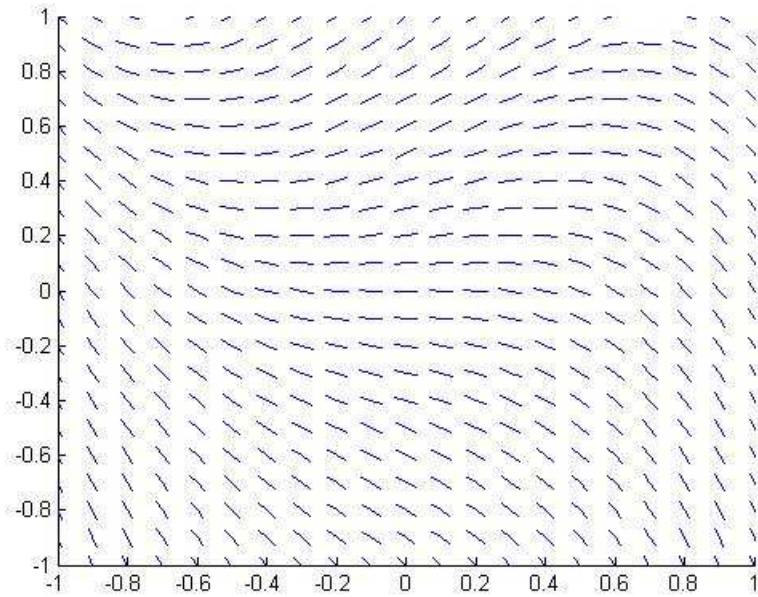
For example, to produce a slope field for the differential equation $dy/dx = y - 2x^2$, change **func.m** to

```
function ydash = func(x,y)
ydash = y - 2*x^2;
```

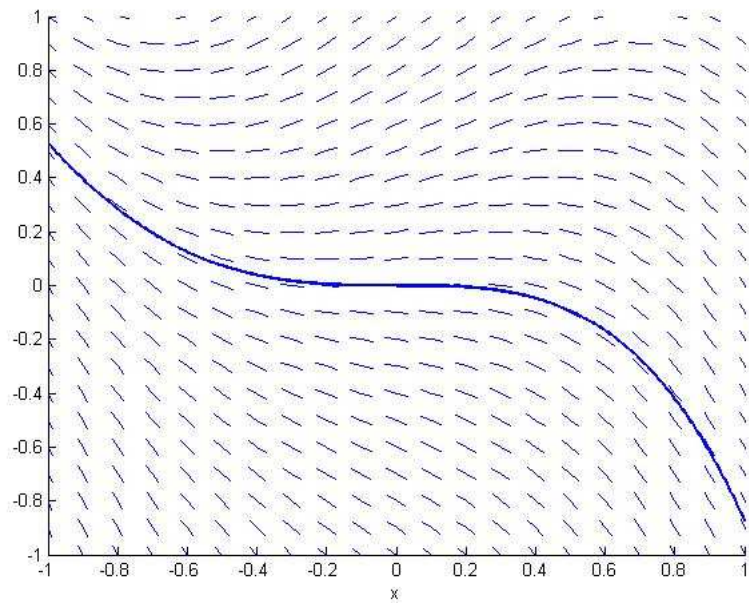
and then run **field.m** with appropriate input arguments:

```
>> field(-1,1,20)
```

The first argument is the lower bound for the horizontal and vertical axes, the second argument is the upper bound, and the third argument is the number of intervals between the field lines. Note that this function will always produce a plot on a square. You should get the following output.



A possible solution to the differential equation is shown below.



Taking the initial condition $y(0) = 0$ and following the field lines to the left and to the right yields a curve like the one shown.

6.8.2 Euler's method

Euler's method is a numerical method for solving differential equations. The algorithm essentially just uses the slope field 'solution' described above:

- Start at a point (x_0, y_0) , described by an initial condition;
- Determine the value of the gradient $f(x_0, y_0)$ at that point;
- Follow in the direction of the gradient for a short distance, to the point (x_1, y_1) ;
- Determine the gradient $f(x_1, y_1)$ at this new point;
- Follow in the direction of the gradient for a short distance, to the point (x_2, y_2) ;

and so on.

This algorithm is very easy to code in MATLAB. For the differential equation $dy/dx = f(x, y)$ with initial condition $y(x_0) = y_0$, you need to define the function $f(x, y)$ (as we did earlier using **func.m**), and then write a **for** loop which performs the calculation

$$y_{i+1} = y_i + f(x_i, y_i)\Delta x$$

where Δx is a small (constant) value which will be the distance along the x -axis between successive points. The y_i points will need to be saved in a vector, as you will need to store them all for the final solution. The x_i points should also be saved in a vector; you can do this before starting the loop with the **linspace** command, or with the ':' (colon) vector notation.

You should produce two sequences of points: the x_i points, which will all be a constant distance Δx apart, and the y_i points, which were calculated in each iteration of the **for** loop. Plotting the y vector against the x vector will illustrate your numerical solution.

You should experiment with changing Δx and/or the number of points used. Euler's method tends to become inaccurate after several thousand steps; you should remember that this is really only an approximate solution to the differential equation. If possible, find the analytic solution and plot it on the same axis to determine if and where the numerical solution becomes inaccurate.

6.8.3 Runge–Kutta method

Euler's method is referred to as a first-order approximation for a differential equation, because the approximation essentially uses a first-order Taylor expansion.

Higher-order methods can yield far more accurate solutions. One such is the Runge–Kutta method, which is technically halfway between a fourth and fifth order method. Consequently, the inbuilt MATLAB program for implementing the method is called ODE45.

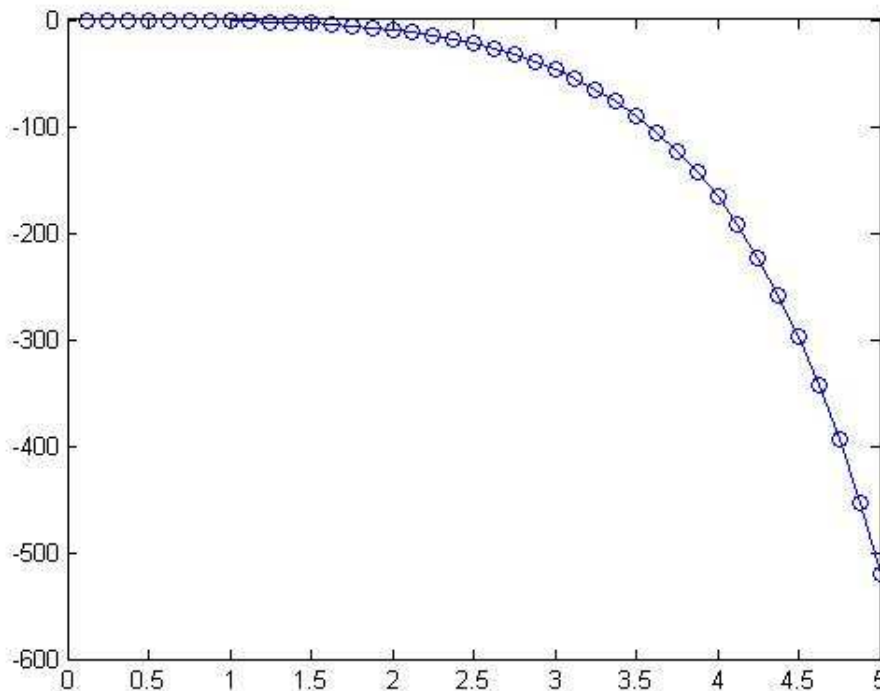
Consider a differential equation problem of the form $dy/dt = f(t, y)$, with $f(t_0) = y_0$ for $t_0 \leq t \leq T$. The MATLAB syntax for solving this problem using ODE45 is

```
ode45('func', [t0,T], y0)
```

where `func` evaluates $f(t, y)$.

For example, consider the differential equation described previously (with x changed to t): $dy/dt = y - 2t^2$, $y(0) = 0$. We will restrict the time interval to $[0, 5]$. We can use the same `func.m` (you can change x to t if you wish). We now only need to execute the ODE45 command:

```
>> ode45('func', [0,5], 0)
```



Alternatively, we could use the command

```
>> [t_out,y_out] = ode45('func', [0,5], 0);
```

In this second case, t_out will be a vector consisting of the t coordinates of the points used to calculate and plot the function, and y_out will be the y coordinates. The plot will not be automatically produced. Using the command

```
>> plot(t_out,y_out)
```

will produce similar output to the plot above.

6.8.4 Systems of differential equations

You have so far seen how to use MATLAB to numerically solve single first-order ODEs. MATLAB's inbuilt ODE solvers can also solve higher-order differential equations, like

$$\frac{d^2y}{dt^2} + e^y = 0, \quad y(0) = 1, \quad \frac{dy}{dt}(0) = 0$$

and coupled systems of ODEs, like

$$\begin{aligned} \frac{dx}{dt} &= f(x, y) = 2x - 3y, & x(0) &= 2 \\ \frac{dy}{dt} &= g(x, y) = x + 2y, & y(0) &= 0. \end{aligned}$$

We will start with the system of equations above. We can use ODE45 in more or less the same way, with only a couple of differences:

- ODE45 stipulates that **func** must take **two** input variables: t and another. However, we have three here: t , x and y . We must write **func** so that it takes inputs (t, u) , where $u = [x, y]$.
- The function **func** will have two outputs: $f(x, y)$ and $g(x, y)$. ODE45 requires that these outputs come in the form of a column vector.

We define our new function as follows:

```
function output = system_func(t,u)
x = u(1);
y = u(2);
f = 2*x - 3*y;
g = x + 2*y;
output = [f;g];
```

We can now use ODE45 in the same way as before (letting the time interval be $[0, 2]$):

```
>> [t_out,u_out] = ode45('system_func', [0,2], [2,0]);
```

The vector t_out has the same meaning as before. The first column of u_out corresponds to the x values, and the second column corresponds to the y values. Plotting t_out against each of these columns will show how x and y change with time. Plotting the columns of u_out against each other will plot x against y (this is called a phase plot).

```
>> plot(t_out,u_out(:,1))
>> plot(t_out,u_out(:,2))
>> plot(u_out(:,1),u_out(:,2))
```

Higher order ODEs can be solved with ODE45 by first converting them into systems of ODEs. Consider the second-order ODE mentioned previously:

$$y'' + e^y = 0, \quad y(0) = 1, y'(0) = 0$$

We introduce a ‘new’ variable $x = y'$ (so $x' = y''$). This leads to the system of (first-order) ODEs:

$$\begin{aligned} \frac{dx}{dt} &= f(x,y) = -e^y & x(0) &= 0 \\ \frac{dy}{dt} &= g(x,y) = x & y(0) &= 1. \end{aligned}$$

We can now write a function to evaluate these:

```
function output = secondorder_func(t,u)
x = u(1);
y = u(2);
f = -exp(y);
g = x;
output = [f;g];
```

The command to execute ODE45 is essentially the same as before (note that our initial value is a vector, because we have two variables):

```
>> [t_out,u_out] = ode45('secondorder_func',[0,5],[0,1]);
```

Since x was an extra variable inserted to assist with solving, we only actually care about y , ie. the second column of u_out . So to plot y against t , use

```
>> plot(t_out,u_out(:,2))
```

6.9 Parametric Functions

One way of presenting a relationship between variables x and y is to write y as a function of x , ie. $y = f(x)$. You have seen that MATLAB is well equipped to plot functions in this form.

Not all relationships between variables can be expressed directly using functions, however. For example, the unit circle on the cartesian plane cannot be described with a simple function $y = f(x)$. It can be plotted in MATLAB by plotting multiple functions with `hold on`:

```
>> figure
>> hold on
>> ezplot('sqrt(1 - x^2)', [-1,1])
>> ezplot('-sqrt(1 - x^2)', [-1,1])
```

However this is needlessly complicated, and may not be possible for more complex relationships. *Parametric functions* are a much easier way to describe relationships such as these, and MATLAB can plot them just as easily.

Parametric functions relate the variables (eg. x and y) through an additional variable, t , called the *parameter*. For example, the unit circle can be described with the parametric functions $x(t) = \cos(t)$ and $y(t) = \sin(t)$. In this particular case, t can be interpreted as the angle around the circle going anticlockwise from the x axis.

The syntax for plotting a two-dimensional parametric relationship is

```
ezplot('xfunc', 'yfunc', [a,b])
```

where `xfunc` and `yfunc` are the parametric representations of x and y respectively, and the parameter t varies over the interval $[a, b]$. So to plot the unit circle, we can use the command

```
>> ezplot('cos(t)', 'sin(t)', [0,2*pi])
```

The default interval for t is $[0, 2\pi]$.

MATLAB has a very similar inbuilt function for plotting three-dimensional parametric relationships, called `ezplot3`. Unsurprisingly, its syntax is

```
ezplot3('xfunc', 'yfunc', 'zfunc', [a,b])
```

For example, to plot a *helix* (a spring-shaped object) of radius 1 and height 20, you could use the command

```
>> ezplot3('cos(t)', 'sin(t)', 't/4', [0,80])
```

The `ezplot3` command has an additional option for showing the progression of the (x, y, z) coordinate as t changes. Use the syntax

```
ezplot3('xfunc','yfunc','zfunc',[a,b],'animate')
```

When using `ezplot3`, remember that you can rotate the figure, as well as change the axis ranges with the `axis` command. Also consider using the commands `axis square` and/or `axis equal`.

6.10 Common commands and operations

6.10.1 Common numerical operations and constants

The following table lists the MATLAB commands for a number of commonly used functions which operate on numbers, as well as some constants. Use MATLAB's `help` command for more information.

Operation	MATLAB Code	Operation	MATLAB Code	Constant	MATLAB Code
$x + y$	<code>x+y</code>	e^x	<code>exp(x)</code>	π	<code>pi</code>
$x - y$	<code>x-y</code>	$ x $	<code>abs(x)</code>	$\sqrt{-1}$	<code>i</code>
xy	<code>x*y</code>	$\sin(x)$	<code>sin(x)</code>	e	<code>exp(1)</code>
x/y	<code>x/y</code>	$\cos(x)$	<code>cos(x)</code>	∞	<code>Inf</code>
x^n	<code>x^n</code>	$\tan(x)$	<code>tan(x)</code>	$-\infty$	<code>-Inf</code>
\sqrt{x}	<code>sqrt(x)</code>	$\arcsin(x)$	<code>asin(x)</code>		
$\log(x)$	<code>log(x)</code>	$\arccos(x)$	<code>acos(x)</code>		
$\log_{10}(x)$	<code>log10(x)</code>	$\arctan(x)$	<code>atan(x)</code>		
$x \times 10^n$	<code>xen</code>	$n!$	<code>factorial(n)</code>		

6.10.2 Vector and matrix operations

The following table lists the MATLAB commands for a number of commonly used operations which can be performed on vectors and/or matrices.

Operation	MATLAB Code
Entering a row vector: $\mathbf{a} = (1 \ 3 \ 4)$	<code>a=[1,3,4]</code>
Entering a column vector: $\mathbf{b} = \begin{pmatrix} 4 \\ 0 \\ 7 \end{pmatrix}$	<code>b=[4;0;7]</code>
Entering a matrix: $A = \begin{pmatrix} 2 & 4 \\ 0 & 1 \end{pmatrix}$	<code>A=[2,4;0,1]</code>
Second component of vector \mathbf{b} :	<code>b(2)</code>
Entry (1,2) of a matrix A:	<code>A(1,2)</code>
Transpose: $\mathbf{c} = \mathbf{a}^T$	<code>c=a'</code>
Adding vectors or matrices: $\mathbf{b} + \mathbf{c}$	<code>b+c</code>
Dot product: $\mathbf{b} \cdot \mathbf{c}$	<code>dot(b,c)</code>
Cross product: $\mathbf{b} \times \mathbf{c}$	<code>cross(b,c)</code>
Norm: $\ \mathbf{a}\ $	<code>norm(a)</code>
Multiplying matrices: AB	<code>A*B</code>
Matrix determinant: $\det(A)$	<code>det(A)</code>
Inverse: A^{-1}	<code>inv(A)</code>
Vector dimension:	<code>length(a)</code>
Matrix dimensions:	<code>size(A)</code>
Minimum value in vector \mathbf{b} :	<code>min(b)</code>
Maximum value in vector \mathbf{b} :	<code>max(b)</code>

6.10.3 Special matrices

MATLAB has a number of special matrices built-in, and these can be produced with certain commands, given in the table below.

Matrix	MATLAB Code
$n \times n$ identity matrix	<code>eye(n)</code>
$m \times n$ matrix of ones	<code>ones(m,n)</code>
$m \times n$ matrix of zeros	<code>zeros(m,n)</code>