Mathematical Economics and Finance

Michael Harrison Patrick Waldron

December 2, 1998

Contents

List of Tables

List of Figures

PREFACE

This book is based on courses MA381 and EC3080, taught at Trinity College Dublin since 1992.

Comments on content and presentation in the present draft are welcome for the benefit of future generations of students.

An electronic version of this book (in LaTeX) is available on the World Wide Web at `http://pwaldron.bess.tcd.ie/teaching/ma381/notes/` although it may not always be the current version.

The book is not intended as a substitute for students' own lecture notes. In particular, many examples and diagrams are omitted and some material may be presented in a different sequence from year to year.

In recent years, mathematics graduates have been increasingly expected to have additional skills in practical subjects such as economics and finance, while economics graduates have been expected to have an increasingly strong grounding in mathematics. The increasing need for those working in economics and finance to have a strong grounding in mathematics has been highlighted by such layman's guides as **?**, **?**, **?** (adapted from **?**) and **?**. In the light of these trends, the present book is aimed at advanced undergraduate students of either mathematics or economics who wish to branch out into the other subject.

The present version lacks supporting materials in *Mathematica* or *Maple*, such as are provided with competing works like **?**.

Before starting to work through this book, mathematics students should think about the nature, subject matter and scientific methodology of economics while economics students should think about the nature, subject matter and scientific methodology of mathematics. The following sections briefly address these questions from the perspective of the outsider.

What Is Economics?

This section will consist of a brief verbal introduction to economics for mathematicians and an outline of the course.

What is economics?

1. Basic microeconomics is about the allocation of wealth or expenditure among different physical goods. This gives us relative prices.

2. Basic finance is about the allocation of expenditure across two or more time periods. This gives us the term structure of interest rates.

3. The next step is the allocation of expenditure across (a finite number or a continuum of) states of nature. This gives us rates of return on risky assets, which are random variables.

Then we can try to combine 2 and 3.
Finally we can try to combine 1 and 2 and 3.
Thus finance is just a subset of micoreconomics.
What do consumers do?
They maximise ‘utility’ given a budget constraint, based on prices and income.
What do firms do?
They maximise profits, given technological constraints (and input and output prices).
Microeconomics is ultimately the theory of the determination of prices by the interaction of all these decisions: all agents simultaneously maximise their objective functions subject to market clearing conditions.

What is Mathematics?

This section will have all the stuff about logic and proof and so on moved into it.

NOTATION

Throughout the book, $\mathbf{x}$ *etc.* will denote points of $\Re^n$ for $n > 1$ and x *etc.* will denote points of $\Re$ or of an arbitrary vector or metric space X. **X** will generally denote a matrix.

Readers should be familiar with the symbols $\forall$ and $\exists$ and with the expressions 'such that' and 'subject to' and also with their meaning and use, in particular with the importance of presenting the parts of a definition in the correct order and with the process of proving a theorem by arguing from the assumptions to the conclusions. Proof by contradiction and proof by contrapositive are also assumed. There is a book on proofs by Solow which should be referred to here.[1]

$\Re^N_+ \equiv \left\{\mathbf{x} \in \Re^N : x_i \geq 0, i = 1, \ldots, N\right\}$ is used to denote the *non-negative orthant* of $\Re^N$, and $\Re^N_{++} \equiv \left\{\mathbf{x} \in \Re^N : x_i > 0, i = 1, \ldots, N\right\}$ used to denote the *positive orthant*.

$^\top$ is the symbol which will be used to denote the transpose of a vector or a matrix.

[1] Insert appropriate discussion of all these topics here.

Part I

MATHEMATICS

Chapter 1

LINEAR ALGEBRA

1.1 Introduction

[To be written.]

1.2 Systems of Linear Equations and Matrices

Why are we interested in solving simultaneous equations?
We often have to find a point which satisfies more than one equation simultaneously, for example when finding equilibrium price and quantity given supply and demand functions.

- To be an equilibrium, the point (Q, P) must lie on both the supply and demand curves.
- Now both supply and demand curves can be plotted on the same diagram and the point(s) of intersection will be the equilibrium (equilibria):

- solving for equilibrium price and quantity is just one of many examples of the simultaneous equations problem
- The ISLM model is another example which we will soon consider at length.
- We will usually have many relationships between many economic variables defining equilibrium.

The first approach to simultaneous equations is the equation counting approach:

- a rough rule of thumb is that we need the same number of equations as unknowns
- this is neither necessary nor sufficient for existence of a unique solution, *e.g.*
  - fewer equations than unknowns, unique solution:

$$x^2 + y^2 = 0 \;\Rightarrow\; x = 0, y = 0$$

  - same number of equations and unknowns but no solution (dependent equations):

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

  - more equations than unknowns, unique solution:

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

$$\Rightarrow\; x = 1, \qquad y = 1$$

Now consider the geometric representation of the simultaneous equation problem, in both the generic and linear cases:

- two curves in the coordinate plane can intersect in 0, 1 or more points
- two surfaces in 3D coordinate space typically intersect in a curve
- three surfaces in 3D coordinate space can intersect in 0, 1 or more points
- a more precise theory is needed

There are three types of *elementary row operations* which can be performed on a system of simultaneous equations without changing the solution(s):

1. Add or subtract a multiple of one equation to or from another equation
2. Multiply a particular equation by a non-zero constant
3. Interchange two equations

Note that each of these operations is reversible (invertible).
Our strategy, roughly equating to *Gaussian elimination* involves using elementary row operations to perform the following steps:

1. (a) Eliminate the first variable from all except the first equation
   (b) Eliminate the second variable from all except the first two equations
   (c) Eliminate the third variable from all except the first three equations
   (d) *&c.*
2. We end up with only one variable in the last equation, which is easily solved.
3. Then we can substitute this solution in the second last equation and solve for the second last variable, and so on.
4. Check your solution!!

Now, let us concentrate on simultaneous linear equations:
(2×2 EXAMPLE)

$$x + y = 2 \quad (1.2.1)$$
$$2y - x = 7 \quad (1.2.2)$$

- Draw a picture
- Use the Gaussian elimination method instead of the following
- Solve for x in terms of y

$$\begin{aligned} x &= 2 - y \\ x &= 2y - 7 \end{aligned}$$

- Eliminate x

$$2 - y = 2y - 7$$

- Find y

$$\begin{aligned} 3y &= 9 \\ y &= 3 \end{aligned}$$

- Find x from either equation:

$$\begin{aligned} x &= 2 - y = 2 - 3 = -1 \\ x &= 2y - 7 = 6 - 7 = -1 \end{aligned}$$

SIMULTANEOUS LINEAR EQUATIONS (3×3 EXAMPLE)

- Consider the general 3D picture ...

- Example:

$$\begin{aligned} x + 2y + 3z &= 6 \quad (1.2.3) \\ 4x + 5y + 6z &= 15 \quad (1.2.4) \\ 7x + 8y + 10z &= 25 \quad (1.2.5) \end{aligned}$$

- Solve one equation (1.2.3) for x in terms of y and z:

$$x = 6 - 2y - 3z$$

- Eliminate x from the other two equations:

$$\begin{aligned} 4\,(6 - 2y - 3z) + 5y + 6z &= 15 \\ 7\,(6 - 2y - 3z) + 8y + 10z &= 25 \end{aligned}$$

- What remains is a 2×2 system:

$$\begin{aligned} -3y - 6z &= -9 \\ -6y - 11z &= -17 \end{aligned}$$

- Solve each equation for y:

$$\begin{aligned} y &= 3 - 2z \\ y &= \frac{17}{6} - \frac{11}{6}z \end{aligned}$$

- Eliminate y:

$$3 - 2z = \frac{17}{6} - \frac{11}{6}z$$

- Find z:

$$\begin{aligned} \frac{1}{6} &= \frac{1}{6}z \\ z &= 1 \end{aligned}$$

- Hence $y = 1$ and $x = 1$.

1.3 Matrix Operations

We motivate the need for matrix algebra by using it as a shorthand for writing systems of linear equations, such as those considered above.

- The steps taken to solve simultaneous linear equations involve only the coefficients so we can use the following shorthand to represent the system of equations used in our example:

  This is called a *matrix*, *i.e.*— a rectangular array of numbers.

- We use the concept of the *elementary matrix* to summarise the elementary row operations carried out in solving the original equations:

  (Go through the whole solution step by step again.)

- Now the rules are

  – Working column by column from left to right, change all the below diagonal elements of the matrix to zeroes

  – Working row by row from bottom to top, change the right of diagonal elements to 0 and the diagonal elements to 1

  – Read off the solution from the last column.

- Or we can reorder the steps to give the *Gaussian elimination* method:

  column by column everywhere.

1.4 Matrix Arithmetic

- Two $n \times m$ matrices can be added and subtracted element by element.

- There are three notations for the general 3×3 system of simultaneous linear equations:

  1. 'Scalar' notation:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3
\end{aligned}
$$

2. 'Vector' notation without factorisation:

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

3. 'Vector' notation with factorisation:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

It follows that:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{pmatrix}$$

- From this we can deduce the general multiplication rules:

  The ijth element of the matrix product **AB** is the product of the ith row of **A** and the jth column of **B**.

  A row and column can only be multiplied if they are the same 'length.'

  In that case, their product is the sum of the products of corresponding elements.

  Two matrices can only be multiplied if the number of columns (*i.e.* the row lengths) in the first equals the number of rows (*i.e.* the column lengths) in the second.

- The *scalar product* of two vectors in $\Re^n$ is the matrix product of one written as a row vector ($1 \times n$ matrix) and the other written as a column vector ($n \times 1$ matrix).

- This is independent of which is written as a row and which is written as a column.

So we have $\mathbf{C} = \mathbf{AB}$ if and only if $c_{ij} = \sum k = 1^n a_{ik}b_{kj}$.
Note that multiplication is associative but not commutative.
Other binary matrix operations are addition and subtraction.
Addition is associative and commutative. Subtraction is neither.
Matrices can also be multiplied by scalars.
Both multiplications are distributive over addition.

We now move on to unary operations.
The additive and multiplicative identity matrices are respectively $\mathbf{0}$ and $\mathbf{I}_n \equiv \left(\delta^i_j\right)$.
$-\mathbf{A}$ and $\mathbf{A}^{-1}$ are the corresponding inverse. Only non-singular matrices have multiplicative inverses.
Finally, we can interpret matrices in terms of linear transformations.

- The product of an $m \times n$ matrix and an $n \times p$ matrix is an $m \times p$ matrix.
- The product of an $m \times n$ matrix and an $n \times 1$ matrix (vector) is an $m \times 1$ matrix (vector).
- So every $m \times n$ matrix, $\mathbf{A}$, defines a function, known as a *linear transformation*,
$$T_{\mathbf{A}} : \Re^n \to \Re^m : \mathbf{x} \mapsto \mathbf{Ax},$$
which maps $n-$dimensional vectors to $m-$dimensional vectors.
- In particular, an $n \times n$ square matrix defines a linear transformation mapping $n-$dimensional vectors to $n-$dimensional vectors.
- The system of n simultaneous linear equations in n unknowns
$$\mathbf{Ax} = \mathbf{b}$$
has a unique solution $\forall \mathbf{b}$ if and only if the corresponding linear transformation $T_{\mathbf{A}}$ is an invertible or bijective function: $\mathbf{A}$ is then said to be an *invertible matrix*.

A matrix has an inverse if and only the corresponding linear transformation is an invertible function:

- Suppose $\mathbf{Ax} = \mathbf{b}_0$ does not have a unique solution. Say it has two distinct solutions, $\mathbf{x}_1$ and $\mathbf{x}_2$ ($\mathbf{x}_1 \neq \mathbf{x}_2$):
$$\begin{aligned} \mathbf{Ax}_1 &= \mathbf{b}_0 \\ \mathbf{Ax}_2 &= \mathbf{b}_0 \end{aligned}$$
This is the same thing as saying that the linear transformation $T_{\mathbf{A}}$ is not injective, as it maps both $\mathbf{x}_1$ and $\mathbf{x}_2$ to the same image.
- Then whenever $\mathbf{x}$ is a solution of $\mathbf{Ax} = \mathbf{b}$:
$$\begin{aligned} \mathbf{A}\left(\mathbf{x} + \mathbf{x}_1 - \mathbf{x}_2\right) &= \mathbf{Ax} + \mathbf{Ax}_1 - \mathbf{Ax}_2 \\ &= \mathbf{b} + \mathbf{b}_0 - \mathbf{b}_0 \\ &= \mathbf{b}, \end{aligned}$$
so $\mathbf{x} + \mathbf{x}_1 - \mathbf{x}_2$ is another, different, solution to $\mathbf{Ax} = \mathbf{b}$.

- So uniqueness of solution is determined by invertibility of the coefficient matrix **A** independent of the right hand side vector **b**.
- If **A** is not invertible, then there will be multiple solutions for some values of **b** and no solutions for other values of **b**.

So far, we have seen two notations for solving a system of simultaneous linear equations, both using elementary row operations.

1. We applied the method to scalar equations (in x, y and z).
2. We then applied it to the augmented matrix $(\mathbf{A}\ \mathbf{b})$ which was reduced to the augmented matrix $(\mathbf{I}\ \mathbf{x})$.

   Now we introduce a third notation.
3. Each step above (about six of them depending on how things simplify) amounted to premultiplying the augmented matrix by an elementary matrix, say
   $$\mathbf{E}_6\mathbf{E}_5\mathbf{E}_4\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\left(\mathbf{A}\ \mathbf{b}\right) = \left(\mathbf{I}\ \mathbf{x}\right). \tag{1.4.1}$$
   Picking out the first 3 columns on each side:
   $$\mathbf{E}_6\mathbf{E}_5\mathbf{E}_4\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}. \tag{1.4.2}$$
   We define
   $$\mathbf{A}^{-1} \equiv \mathbf{E}_6\mathbf{E}_5\mathbf{E}_4\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1. \tag{1.4.3}$$
   And we can use Gaussian elimination in turn to solve for each of the columns of the inverse, or to solve for the whole thing at once.

   Lots of properties of inverses are listed in MJH's notes (p.A7?).

The transpose is $\mathbf{A}^\top$, sometimes denoted $\mathbf{A}'$ or $\mathbf{A}^t$.
A matrix is symmetric if it is its own transpose; skewsymmetric if $\mathbf{A}^\top = -\mathbf{A}$.
Note that $\left(\mathbf{A}^\top\right)^{-1} = \left(\mathbf{A}^{-1}\right)^\top$.
Lots of strange things can happen in matrix arithmetic.
We can have $\mathbf{AB} = \mathbf{0}$ even if $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$.

Definition 1.4.1 *orthogonal rows/columns*

Definition 1.4.2 *idempotent matrix* $\mathbf{A}^2 = \mathbf{A}$

Definition 1.4.3 *orthogonal*[1] *matrix* $\mathbf{A}^\top = \mathbf{A}^{-1}$.

Definition 1.4.4 *partitioned matrices*

Definition 1.4.5 *determinants*

Definition 1.4.6 *diagonal, triangular and scalar matrices*

[1] This is what **?** calls something that it seems more natural to call an orthonormal matrix.

1.5 Vectors and Vector Spaces

Definition 1.5.1 *A* vector *is just an* $n \times 1$ *matrix.*

The Cartesian product of n sets is just the set of ordered n-tuples where the ith component of each n-tuple is an element of the ith set.
The ordered n-tuple $(x_1, x_2, \ldots, x_n)$ is identified with the $n \times 1$ column vector
$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$
Look at pictures of points in $\Re^2$ and $\Re^3$ and think about extensions to $\Re^n$.
Another geometric interpretation is to say that a vector is an entity which has both magnitude and direction, while a scalar is a quantity that has magnitude only.

Definition 1.5.2 *A* real *(or* Euclidean*)* vector space *is a set (of vectors) in which addition and scalar multiplication (i.e. by real numbers) are defined and satisfy the following axioms:*

1. *copy axioms from simms 131 notes p.1*

There are vector spaces over other fields, such as the complex numbers.
Other examples are function spaces, matrix spaces.
On some vector spaces, we also have the notion of a dot product or scalar product:

$$\mathbf{u}.\mathbf{v} \equiv \mathbf{u}^\top \mathbf{v}$$

The Euclidean norm of $\mathbf{u}$ is
$$\sqrt{\mathbf{u}.\mathbf{u}} \equiv \| \mathbf{u} \| .$$
A unit vector is defined in the obvious way … unit norm.
The distance between two vectors is just $\| \mathbf{u} - \mathbf{v} \|$.
There are lots of interesting properties of the dot product (MJH's theorem 2).
We can calculate the angle between two vectors using a geometric proof based on the cosine rule.

$$\begin{aligned} \| \mathbf{v} - \mathbf{u} \|^2 &= (\mathbf{v} - \mathbf{u}) . (\mathbf{v} - \mathbf{u}) && (1.5.1) \\ &= \| \mathbf{v} \|^2 + \| \mathbf{u} \|^2 - 2\mathbf{v}.\mathbf{u} && (1.5.2) \\ &= \| \mathbf{v} \|^2 + \| \mathbf{u} \|^2 - 2 \| \mathbf{v} \| \| \mathbf{u} \| \cos\theta && (1.5.3) \end{aligned}$$

Two vectors are orthogonal if and only if the angle between them is zero.

A subspace is a subset of a vector space which is closed under addition and scalar multiplication.
For example, consider row space, column space, solution space, orthogonal complement.

1.6 Linear Independence

Definition 1.6.1 *The vectors* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_r \in \Re^n$ *are linearly independent if and only if*

$$\sum_{i=1}^{r} \alpha_i \mathbf{x}_i = \mathbf{0} \Rightarrow \alpha_i = 0 \forall i.$$

Otherwise, they are linearly dependent.
Give examples of each, plus the standard basis.
If $r > n$, then the vectors must be linearly dependent.
If the vectors are orthonormal, then they must be linearly independent.

1.7 Bases and Dimension

A basis for a vector space is a set of vectors which are linearly independent and which span or generate the entire space.
Consider the standard bases in $\Re^2$ and $\Re^n$.
Any two non-collinear vectors in $\Re^2$ form a basis.
A linearly independent spanning set is a basis for the subspace which it generates.
Proof of the next result requires stuff that has not yet been covered.
If a basis has n elements then any set of more than n elements is linearly dependent and any set of less than n elements doesn't span.
Or something like that.

Definition 1.7.1 *The dimension of a vector space is the (unique) number of vectors in a basis. The dimension of the vector space* $\{\mathbf{0}\}$ *is zero.*

Definition 1.7.2 *Orthogonal complement*

Decomposition into subspace and its orthogonal complement.

1.8 Rank

Definition 1.8.1

The row space *of an* $m \times n$ *matrix* $\mathbf{A}$ *is the vector subspace of* $\Re^n$ *generated by the* m *rows of* $\mathbf{A}$.

The row rank *of a matrix is the dimension of its row space.*

The column space *of an* $m \times n$ *matrix* $\mathbf{A}$ *is the vector subspace of* $\Re^m$ *generated by the* n *columns of* $\mathbf{A}$.

The column rank *of a matrix is the dimension of its column space.*

Theorem 1.8.1 *The row space and the column space of any matrix have the same dimension.*

Proof The idea of the proof is that performing elementary row operations on a matrix does not change either the row rank or the column rank of the matrix.
Using a procedure similar to Gaussian elimination, every matrix can be reduced to a matrix in reduced row echelon form (a partitioned matrix with an identity matrix in the top left corner, anything in the top right corner, and zeroes in the bottom left and bottom right corner).
By inspection, it is clear that the row rank and column rank of such a matrix are equal to each other and to the dimension of the identity matrix in the top left corner.
In fact, elementary row operations do not even change the row space of the matrix.
They clearly do change the column space of a matrix, but not the column rank as we shall now see.
If $\mathbf{A}$ and $\mathbf{B}$ are row equivalent matrices, then the equations $\mathbf{Ax} = \mathbf{0}$ and $\mathbf{Bx} = \mathbf{0}$ have the same solution space.
If a subset of columns of $\mathbf{A}$ are linearly dependent, then the solution space does contain a vector in which the corresponding entries are nonzero and all other entries are zero.
Similarly, if a subset of columns of $\mathbf{A}$ are linearly independent, then the solution space does not contain a vector in which the corresponding entries are nonzero and all other entries are zero.
The first result implies that the corresponding columns or $\mathbf{B}$ are also linearly dependent.
The second result implies that the corresponding columns of $\mathbf{B}$ are also linearly independent.
It follows that the dimension of the column space is the same for both matrices.

Q.E.D.

Definition 1.8.2 *rank*

Definition 1.8.3 *solution space, null space or kernel*

Theorem 1.8.2 *dimension of row space + dimension of null space = number of columns*

The solution space of the system means the solution space of the homogenous equation $\mathbf{Ax} = \mathbf{0}$.
The non-homogenous equation $\mathbf{Ax} = \mathbf{b}$ may or may not have solutions.
System is consistent iff rhs is in column space of A and there is a solution.
Such a solution is called a particular solution.
A general solution is obtained by adding to some particular solution a generic element of the solution space.
Previously, solving a system of linear equations was something we only did with non-singular square systems.
Now, we can solve any system by describing the solution space.

1.9 Eigenvalues and Eigenvectors

Definition 1.9.1 *eigenvalues and eigenvectors and λ-eigenspaces*

Compute eigenvalues using $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. So some matrices with real entries can have complex eigenvalues.
Real symmetric matrix has real eigenvalues. Prove using complex conjugate argument.
Given an eigenvalue, the corresponding eigenvector is the solution to a singular matrix equation, so one free parameter (at least).
Often it is useful to specify unit eigenvectors.
Eigenvectors of a real symmetric matrix corresponding to different eigenvalues are orthogonal (orthonormal if we normalise them).
So we can diagonalize a symmetric matrix in the following sense:
If the columns of $\mathbf{P}$ are orthonormal eigenvectors of $\mathbf{A}$, and $\boldsymbol{\lambda}$ is the matrix with the corresponding eigenvalues along its leading diagonal, then $\mathbf{AP} = \mathbf{P}\boldsymbol{\lambda}$ so $\mathbf{P}^{-1}\mathbf{AP} = \boldsymbol{\lambda} = \mathbf{P}^{\top}\mathbf{AP}$ as $\mathbf{P}$ is an orthogonal matrix.
In fact, all we need to be able to diagonalise in this way is for $\mathbf{A}$ to have n linearly independent eigenvectors.
$\mathbf{P}^{-1}\mathbf{AP}$ and $\mathbf{A}$ are said to be *similar* matrices.
Two similar matrices share lots of properties: determinants and eigenvalues in particular. Easy to show this.
But eigenvectors are different.

1.10 Quadratic Forms

A quadratic form is

1.11 Symmetric Matrices

Symmetric matrices have a number of special properties

1.12 Definite Matrices

Definition 1.12.1 *An $n \times n$ square matrix $\mathbf{A}$ is said to be*

$$\begin{array}{rcl} \text{positive definite} & \iff & \mathbf{x}^\top \mathbf{A}\mathbf{x} > 0 \,\forall \mathbf{x} \in \Re^n,\ \mathbf{x} \neq \mathbf{0} \\ \text{positive semi-definite} & \iff & \mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0 \,\forall \mathbf{x} \in \Re^n \\ \text{negative definite} & \iff & \mathbf{x}^\top \mathbf{A}\mathbf{x} < 0 \,\forall \mathbf{x} \in \Re^n,\ \mathbf{x} \neq \mathbf{0} \\ \text{negative semi-definite} & \iff & \mathbf{x}^\top \mathbf{A}\mathbf{x} \leq 0 \,\forall \mathbf{x} \in \Re^n \end{array}$$

Some texts may require that the matrix also be symmetric, but this is not essential and sometimes looking at the definiteness of non-symmetric matrices is relevant.
If $\mathbf{P}$ is an invertible $n \times n$ square matrix and $\mathbf{A}$ is any $n \times n$ square matrix, then $\mathbf{A}$ is positive/negative (semi-)definite if and only if $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is.
In particular, the definiteness of a symmetric matrix can be determined by checking the signs of its eigenvalues.
Other checks involve looking at the signs of the elements on the leading diagonal.
Definite matrices are non-singular and singular matrices can not be definite.
The commonest use of positive definite matrices is as the variance-covariance matrices of random variables. Since

$$v_{ij} = \mathrm{Cov}\left[\tilde{r}_i, \tilde{r}_j\right] = \mathrm{Cov}\left[\tilde{r}_j, \tilde{r}_i\right] \tag{1.12.1}$$

and

$$\begin{aligned} \mathbf{w}^\top \mathbf{V}\mathbf{w} &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \mathrm{Cov}\left[\tilde{r}_i, \tilde{r}_j\right] && (1.12.2) \\ &= \mathrm{Cov}\left[\sum_{i=1}^{N} w_i \tilde{r}_i, \sum_{j=1}^{N} w_j \tilde{r}_j\right] && (1.12.3) \\ &= \mathrm{Var}[\sum_{i=1}^{N} w_i \tilde{r}_i] \geq 0 && (1.12.4) \end{aligned}$$

a variance-covariance matrix must be real, symmetric and positive semi-definite.

In Theorem 3.2.4, it will be seen that the definiteness of a matrix is also an essential idea in the theory of convex functions.

We will also need later the fact that the inverse of a positive (negative) definite matrix (in particular, of a variance-covariance matrix) is positive (negative) definite.

Semi-definite matrices which are not definite have a zero eigenvalue and therefore are singular.

Chapter 2

VECTOR CALCULUS

2.1 Introduction

[To be written.]

2.2 Basic Topology

The aim of this section is to provide sufficient introduction to topology to motivate the definitions of continuity of functions and correspondences in the next section, but no more.

- A *metric space* is a non-empty set X equipped with a *metric*, *i.e.* a function $d : X \times X \rightarrow [0, \infty)$ such that
  1. $d(x, y) = 0 \iff x = y$.
  2. $d(x, y) = d(y, x)\ \forall x, y \in X$.
  3. The triangular inequality:
  $$d(x, z) + d(z, y) \geq d(x, y)\ \forall x, y, z \in X.$$
- An *open ball* is a subset of a metric space, X, of the form
  $$B\epsilon(x) = \{y \in X : d(y, x) < \epsilon\}.$$
- A subset A of a metric space is *open*
  $\iff$
  $\forall x \in A, \exists \epsilon > 0$ such that $B\epsilon(x) \subseteq A$.

- A is *closed* $\iff X - A$ is open. (Note that many sets are neither open nor closed.)

- A neighbourhood of $x \in X$ is an open set containing x.

Definition 2.2.1 *Let* $X = \Re^n$*.* $A \subseteq X$ *is compact* $\iff$ A *is both closed and bounded (i.e.* $\exists \mathbf{x}, \epsilon$ *such that* $A \subseteq B\epsilon(\mathbf{x})$*).*

We need to formally define the interior of a set before stating the separating theorem:

Definition 2.2.2 *If* Z *is a subset of a metric space* X*, then the* interior *of* Z*, denoted* $\text{int } Z$*, is defined by*

$$z \in \text{int } Z \iff B\epsilon\,(z) \subseteq Z \text{ for some } \epsilon > 0.$$

2.3 Vector-valued Functions and Functions of Several Variables

Definition 2.3.1 *A* function *(or* map*)* $f : X \to Y$ *from a* domain X *to a* co-domain Y *is a rule which assigns to each element of* X *a unique element of* Y*.*

Definition 2.3.2 *A* correspondence $f : X \to Y$ *from a* domain X *to a* co-domain Y *is a rule which assigns to each element of* X *a non-empty subset of* Y*.*

Definition 2.3.3 *The* range *of the function* $f : X \to Y$ *is the set* $f(X) = \{f(x) \in Y : x \in X\}$.

Definition 2.3.4 *The function* $f : X \to Y$ *is* injective *(one-to-one)*
$\iff$
$f(x) = f(x') \Rightarrow x = x'$.

Definition 2.3.5 *The function* $f : X \to Y$ *is* surjective *(onto)*
$\iff$
$f(X) = Y$

Definition 2.3.6 *The function* $f : X \to Y$ *is* bijective *(or* invertible*)*
$\iff$
it is both injective and surjective.

Note that if $f: X \to Y$ and $A \subseteq X$ and $B \subseteq Y$, then

$$f(A) \equiv \{f(x) : x \in A\} \subseteq Y$$

and

$$f^{-1}(B) \equiv \{x \in X : f(x) \in B\} \subseteq X.$$

Definition 2.3.7 *A* vector-valued function *is a function whose co-domain is a subset of a vector space, say* $\Re^N$. *Such a function has* N component functions.

Definition 2.3.8 *A* function of several variables *is a function whose domain is a subset of a vector space.*

Definition 2.3.9 *The function* $f: X \to Y$ *(*$X \subseteq \Re^n, Y \subseteq \Re$*) approaches the* limit y^* *as* $\mathbf{x} \to \mathbf{x}^*$
$\iff$
$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \parallel \mathbf{x} - \mathbf{x}^* \parallel < \delta \Longrightarrow |f(\mathbf{x}) - \mathbf{y}^*)| < \epsilon.$

This is usually denoted

$$\lim_{\mathbf{x} \to \mathbf{x}^*} f(\mathbf{x}) = y^*.$$

Definition 2.3.10 *The function* $f: X \to Y$ *(*$X \subseteq \Re^n, Y \subseteq \Re$*) is* continuous at $\mathbf{x}^*$
$\iff$
$\forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \parallel \mathbf{x} - \mathbf{x}^* \parallel < \delta \Longrightarrow |f(\mathbf{x}) - f(\mathbf{x}^*)| < \epsilon.$

This definition just says that f is continuous provided that

$$\lim_{\mathbf{x} \to \mathbf{x}^*} f(\mathbf{x}) = f(\mathbf{x}^*).$$

? discusses various alternative but equivalent definitions of continuity.

Definition 2.3.11 *The function* $f : X \to Y$ *is* continuous
$\iff$
it is continuous at every point of its domain.

We will say that a vector-valued function is continuous if and only if each of its component functions is continuous.
The notion of continuity of a function described above is probably familiar from earlier courses. Its extension to the notion of continuity of a correspondence, however, while fundamental to consumer theory, general equilibrium theory and much of microeconomics, is probably not. In particular, we will meet it again in Theorem 3.5.4. The interested reader is referred to **?** for further details.

Definition 2.3.12 *1. The correspondence* $f: X \rightarrow Y$ *(*$X \subseteq \Re^n, Y \subseteq \Re$*) is* upper hemi-continuous (u.h.c.) at $\mathbf{x}^*$

$\Longleftrightarrow$

for every open set N *containing the set* $f(\mathbf{x}^*)$, $\exists \delta > 0$ *s.t.* $\| \mathbf{x} - \mathbf{x}^* \| < \delta \Longrightarrow f(\mathbf{x}) \subseteq N$.

(Upper hemi-continuity basically means that the graph of the correspondence is a closed and connected set.)

2. *The correspondence* $f: X \rightarrow Y$ *(*$X \subseteq \Re^n, Y \subseteq \Re$*) is* lower hemi-continuous (l.h.c.) at $\mathbf{x}^*$

$\Longleftrightarrow$

for every open set N *intersecting the set* $f(\mathbf{x}^*)$, $\exists \delta > 0$ *s.t.* $\| \mathbf{x} - \mathbf{x}^* \| < \delta \Longrightarrow f(\mathbf{x})$ *intersects* N.

3. *The correspondence* $f: X \rightarrow Y$ *(*$X \subseteq \Re^n, Y \subseteq \Re$*) is* continuous (at $\mathbf{x}^*$)

$\Longleftrightarrow$

it is both upper hemi-continuous and lower hemi-continuous (at $\mathbf{x}^*$*)*

(There are a couple of pictures from **?** to illustrate these definitions.)

2.4 Partial and Total Derivatives

Definition 2.4.1 *The* (total) derivative *or Jacobean of a real-valued function of* N *variables is the* N*-dimensional row vector of its partial derivatives. The Jacobean of a vector-valued function with values in* $\Re^M$ *is an* $M \times N$ *matrix of partial derivatives whose* j*th row is the Jacobean of the* j*th component function.*

Definition 2.4.2 *The* gradient *of a real-valued function is the transpose of its Jacobean.*

Definition 2.4.3 *A function is said to be* differentiable at **x** *if all its partial derivatives exist at* **x**.

Definition 2.4.4 *The function* $f : X \rightarrow Y$ *is* differentiable

$\Longleftrightarrow$

it is differentiable at every point of its domain

Definition 2.4.5 *The* Hessian matrix *of a real-valued function is the (usually symmetric) square matrix of its second order partial derivatives.*

Note that if $f : \Re^n \rightarrow \Re$, then, strictly speaking, the second derivative (Hessian) of f is the derivative of the vector-valued function

$$(f')^\top : \Re^n \rightarrow \Re^n : \mathbf{x} \mapsto (f'(\mathbf{x}))^\top .$$

Students always need to be warned about the differences in notation between the case of $n = 1$ and the case of $n > 1$. Statements and shorthands that make sense in univariate calculus must be modified for multivariate calculus.

2.5 The Chain Rule and Product Rule

Theorem 2.5.1 (The Chain Rule) *Let $g : \Re^n \rightarrow \Re^m$ and $f : \Re^m \rightarrow \Re^p$ be continuously differentiable functions and let $h : \Re^n \rightarrow \Re^p$ be defined by*

$$h(\mathbf{x}) \equiv f(g(\mathbf{x})) .$$

Then

$$\underbrace{h'(\mathbf{x})}_{p \times n} = \underbrace{f'(g(\mathbf{x}))}_{p \times m} \underbrace{g'(\mathbf{x})}_{m \times n} .$$

Proof This is easily shown using the Chain Rule for partial derivatives.

Q.E.D.

One of the most common applications of the Chain Rule is the following:
Let $g : \Re^n \rightarrow \Re^m$ and $f : \Re^{m+n} \rightarrow \Re^p$ be continuously differentiable functions, let $\mathbf{x} \in \Re^n$, and define $h : \Re^n \rightarrow \Re^p$ by:

$$h(\mathbf{x}) \equiv f(g(\mathbf{x}), \mathbf{x}) .$$

The univariate Chain Rule can then be used to calculate $\frac{\partial h^i}{\partial x_j}(\mathbf{x})$ in terms of partial derivatives of f and g for $i = 1, \ldots, p$ and $j = 1, \ldots, n$:

$$\frac{\partial h^i}{\partial x_j}(\mathbf{x}) = \sum_{k=1}^{m} \frac{\partial f^i}{\partial x_k}(g(\mathbf{x}), \mathbf{x}) \frac{\partial g^k}{\partial x_j}(\mathbf{x}) + \sum_{k=m+1}^{m+n} \frac{\partial f^i}{\partial x_k}(g(\mathbf{x}), \mathbf{x}) \frac{\partial x^k}{\partial x_j}(\mathbf{x}) . \quad (2.5.1)$$

Note that

$$\frac{\partial x^k}{\partial x_j}(\mathbf{x}) = \delta_j^k \equiv \begin{cases} 1 & \text{if } k = j \\ 0 & \text{otherwise} \end{cases} ,$$

which is known as the *Kronecker Delta*. Thus all but one of the terms in the second summation in (2.5.1) vanishes, giving:

$$\frac{\partial h^i}{\partial x_j}(\mathbf{x}) = \sum_{k=1}^{m} \frac{\partial f^i}{\partial x_k}(g(\mathbf{x}), \mathbf{x}) \frac{\partial g^k}{\partial x_j}(\mathbf{x}) + \frac{\partial f^i}{\partial x_j}(g(\mathbf{x}), \mathbf{x}) .$$

Stacking these scalar equations in matrix form and factoring yields:

$$\begin{pmatrix} \frac{\partial h^1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial h^1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial h^p}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial h^p}{\partial x_n}(\mathbf{x}) \end{pmatrix} =$$

$$\begin{pmatrix} \frac{\partial f^1}{\partial x_1}(g(\mathbf{x}),\mathbf{x}) & \dots & \frac{\partial f^1}{\partial x_m}(g(\mathbf{x}),\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^p}{\partial x_1}(g(\mathbf{x}),\mathbf{x}) & \dots & \frac{\partial f^p}{\partial x_m}(g(\mathbf{x}),\mathbf{x}) \end{pmatrix} \begin{pmatrix} \frac{\partial g^1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial g^1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial g^m}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial g^m}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

$$+ \begin{pmatrix} \frac{\partial f^1}{\partial x_{m+1}}(g(\mathbf{x}),\mathbf{x}) & \dots & \frac{\partial f^1}{\partial x_{m+n}}(g(\mathbf{x}),\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f^p}{\partial x_{m+1}}(g(\mathbf{x}),\mathbf{x}) & \dots & \frac{\partial f^p}{\partial x_{m+n}}(g(\mathbf{x}),\mathbf{x}) \end{pmatrix}. \tag{2.5.2}$$

Now, by partitioning the total derivative of f as

$$\underbrace{f'(\cdot)}_{p\times(m+n)} = \begin{bmatrix} \underbrace{D_{\mathbf{g}} f(\cdot)}_{p\times m} & \underbrace{D_{\mathbf{x}} f(\cdot)}_{p\times n} \end{bmatrix}, \tag{2.5.3}$$

we can use (2.5.2) to write out the total derivative $h'(\mathbf{x})$ as a product of partitioned matrices:

$$h'(\mathbf{x}) = D_{\mathbf{g}} f(g(\mathbf{x}),\mathbf{x})\, g'(\mathbf{x}) + D_{\mathbf{x}} f(g(\mathbf{x}),\mathbf{x}). \tag{2.5.4}$$

Theorem 2.5.2 (Product Rule for Vector Calculus) *The multivariate Product Rule comes in two versions:*

1. *Let $f, g\colon \Re^m \to \Re^n$ and define $h\colon \Re^m \to \Re$ by*

$$\underbrace{h(\mathbf{x})}_{1\times 1} \equiv \underbrace{(f(\mathbf{x}))^\top}_{1\times n} \underbrace{g(\mathbf{x})}_{n\times 1}.$$

   Then

$$\underbrace{h'(\mathbf{x})}_{1\times m} = \underbrace{(g(\mathbf{x}))^\top}_{1\times n} \underbrace{f'(\mathbf{x})}_{n\times m} + \underbrace{(f(\mathbf{x}))^\top}_{1\times n} \underbrace{g'(\mathbf{x})}_{n\times m}.$$

2. *Let $f\colon \Re^m \to \Re$ and $g\colon \Re^m \to \Re^n$ and define $h\colon \Re^m \to \Re^n$ by*

$$\underbrace{h(\mathbf{x})}_{n\times 1} \equiv \underbrace{f(\mathbf{x})}_{1\times 1} \underbrace{g(\mathbf{x})}_{n\times 1}.$$

   Then

$$\underbrace{h'(\mathbf{x})}_{n\times m} = \underbrace{g(\mathbf{x})}_{n\times 1} \underbrace{f'(\mathbf{x})}_{1\times m} + \underbrace{f(\mathbf{x})}_{1\times 1} \underbrace{g'(\mathbf{x})}_{n\times m}.$$

Proof This is easily shown using the Product Rule from univariate calculus to calculate the relevant partial derivatives and then stacking the results in matrix form.

Q.E.D.

2.6 The Implicit Function Theorem

Theorem 2.6.1 (Implicit Function Theorem) *Let $g\colon \Re^n \to \Re^m$, where $m < n$. Consider the system of m scalar equations in n variables, $g(\mathbf{x}^*) = \mathbf{0}_m$.*
Partition the n-dimensional vector $\mathbf{x}$ as $(\mathbf{y}, \mathbf{z})$ where $\mathbf{y} = (x_1, x_2, \ldots, x_m)$ is m-dimensional and $\mathbf{z} = (x_{m+1}, x_{m+2}, \ldots, x_n)$ is $(n-m)$-dimensional. Similarly, partition the total derivative of g at $\mathbf{x}^*$ as

$$\begin{array}{ccccc} g'(\mathbf{x}^*) & = & [D_{\mathbf{y}}g & & D_{\mathbf{z}}g] \\ (m \times n) & & (m \times m) & & (m \times (n-m)) \end{array} \tag{2.6.1}$$

We aim to solve these equations for the first m variables, $\mathbf{y}$, which will then be written as functions, $h(\mathbf{z})$ of the last $n-m$ variables, $\mathbf{z}$.
Suppose g is continuously differentiable in a neighbourhood of $\mathbf{x}^*$, and that the $m \times m$ matrix:

$$D_{\mathbf{y}}g \equiv \begin{pmatrix} \frac{\partial g^1}{\partial x_1}(\mathbf{x}^*) & \ldots & \frac{\partial g^1}{\partial x_m}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g^m}{\partial x_1}(\mathbf{x}^*) & \ldots & \frac{\partial g^m}{\partial x_m}(\mathbf{x}^*) \end{pmatrix}$$

formed by the first m columns of the total derivative of g at $\mathbf{x}^*$ is non-singular.
Then $\exists$ neighbourhoods Y of $\mathbf{y}^*$ and Z of $\mathbf{z}^*$, and a continuously differentiable function $h\colon Z \to Y$ such that

1. $\mathbf{y}^* = h(\mathbf{z}^*)$,
2. $g(h(\mathbf{z}), \mathbf{z}) = \mathbf{0} \quad \forall \mathbf{z} \in Z$, *and*
3. $h'(\mathbf{z}^*) = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g$.

Proof The full proof of this theorem, like that of Brouwer's Fixed Point Theorem later, is beyond the scope of this course. However, part 3 follows easily from material in Section 2.5. The aim is to derive an expression for the total derivative $h'(\mathbf{z}^*)$ in terms of the partial derivatives of g, using the Chain Rule.
We know from part 2 that

$$f(\mathbf{z}) \equiv g(h(\mathbf{z}), \mathbf{z}) = \mathbf{0}_m \quad \forall \mathbf{z} \in Z.$$

Thus

$$f'(\mathbf{z}) \equiv \mathbf{0}_{m \times (n-m)} \quad \forall \mathbf{z} \in Z,$$

in particular at $\mathbf{z}^*$. But we know from (2.5.4) that

$$f'(\mathbf{z}) = D_{\mathbf{y}}g h'(\mathbf{z}) + D_{\mathbf{z}}g.$$

Hence

$$D_{\mathbf{y}}gh'(\mathbf{z}) + D_{\mathbf{z}}g = \mathbf{0}_{m\times(n-m)}$$

and, since the statement of the theorem requires that $D_{\mathbf{y}}g$ is invertible,

$$h'(\mathbf{z}^*) = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g,$$

as required.

Q.E.D.

To conclude this section, consider the following two examples:

1. the equation $g(x,y) \equiv x^2 + y^2 - 1 = 0$.

   Note that $g'(x,y) = (2x\ 2y)$.

   We have $h(y) = \sqrt{1-y^2}$ or $h(y) = -\sqrt{1-y^2}$, each of which describes a single-valued, differentiable function on $(-1,1)$. At $(x,y) = (0,1)$, $\frac{\partial g}{\partial x} = 0$ and $h(y)$ is undefined (for $y > 1$) or multi-valued (for $y < 1$) in any neighbourhood of $y = 1$.

2. the system of linear equations $g(\mathbf{x}) \equiv \mathbf{B}\mathbf{x} = \mathbf{0}$, where $\mathbf{B}$ is an $m \times n$ matrix.

   We have $g'(\mathbf{x}) = \mathbf{B} \quad \forall\mathbf{x}$ so the implicit function theorem applies provided the equations are linearly independent.

2.7 Directional Derivatives

Definition 2.7.1 *Let X be a vector space and $\mathbf{x} \neq \mathbf{x}' \in X$. Then*

1. *for $\lambda \in \Re$ and particularly for $\lambda \in [0,1]$, $\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'$ is called a* convex combination *of $\mathbf{x}$ and $\mathbf{x}'$.*

2. *$L = \{\lambda\mathbf{x} + (1-\lambda)\mathbf{x}' : \lambda \in \Re\}$ is the* line from $\mathbf{x}'$, where $\lambda = 0$, to $\mathbf{x}$, where $\lambda = 1$, in X.

3. *The* restriction of the function $f: X \to \Re$ to the line L *is the function*

$$f|_L: \Re \to \Re: \lambda \mapsto f(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}').$$

4. *If f is a differentiable function, then the* directional derivative of f at $\mathbf{x}'$ in the direction from $\mathbf{x}'$ to $\mathbf{x}$ *is* $f|_L'(0)$.

- We will endeavour, wherever possible, to stick to the convention that $\mathbf{x}'$ denotes the point at which the derivative is to be evaluated and $\mathbf{x}$ denotes the point in the direction of which it is measured.[1]

- Note that, by the Chain Rule,

$$f|_L'(\lambda) = f'(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}')(\mathbf{x} - \mathbf{x}') \tag{2.7.1}$$

  and hence the directional derivative

$$f|_L'(0) = f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}') . \tag{2.7.2}$$

- The ith partial derivative of f at $\mathbf{x}$ is the directional derivative of f at $\mathbf{x}$ in the direction from $\mathbf{x}$ to $\mathbf{x} + \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith standard basis vector. In other words, partial derivatives are a special case of directional derivatives or directional derivatives a generalisation of partial derivatives.

- As an exercise, consider the interpretation of the directional derivatives at a point in terms of the rescaling of the parameterisation of the line L.

- Note also that, returning to first principles,

$$f|_L'(0) = \lim_{\lambda\to 0} \frac{f(\mathbf{x}' + \lambda(\mathbf{x} - \mathbf{x}')) - f(\mathbf{x}')}{\lambda}. \tag{2.7.3}$$

- Sometimes it is neater to write $\mathbf{x} - \mathbf{x}' \equiv \mathbf{h}$. Using the Chain Rule, it is easily shown that the second derivative of $f|_L$ is

$$f|_L''(\lambda) = \mathbf{h}^\top f''(\mathbf{x}' + \lambda\mathbf{h})\mathbf{h}$$

  and

$$f|_L''(0) = \mathbf{h}^\top f''(\mathbf{x}')\mathbf{h}.$$

2.8 Taylor's Theorem: Deterministic Version

This should be fleshed out following **?**.

Readers are presumed to be familiar with single variable versions of Taylor's Theorem. In particular recall both the second order exact and infinite versions.

An interesting example is to approximate the discount factor using powers of the interest rate:

$$\frac{1}{1+i} = 1 - i + i^2 - i^3 + i^4 + \dots \tag{2.8.1}$$

[1]There may be some lapses in this version.

We will also use two multivariate versions of Taylor's theorem which can be obtained by applying the univariate versions to the restriction to a line of a function of n variables.

Theorem 2.8.1 (Taylor's Theorem) *Let $f : X \rightarrow \Re$ be twice differentiable, $X \subseteq \Re^n$. Then for any $\mathbf{x}, \mathbf{x}' \in X$, $\exists \lambda \in (0,1)$ such that*

$$f(\mathbf{x}) = f(\mathbf{x}') + f'(\mathbf{x}')(\mathbf{x}-\mathbf{x}') + \frac{1}{2}(\mathbf{x}-\mathbf{x}')^\top f''(\mathbf{x}'+\lambda(\mathbf{x}-\mathbf{x}'))(\mathbf{x}-\mathbf{x}'). \quad (2.8.2)$$

Proof Let L be the line from $\mathbf{x}'$ to $\mathbf{x}$.
Then the univariate version tells us that there exists $\lambda \in (0,1)$[2] such that

$$f|_L(1) = f|_L(0) + f|'_L(0) + \frac{1}{2} f|''_L(\lambda). \quad (2.8.3)$$

Making the appropriate substitutions gives the multivariate version in the theorem.

Q.E.D.

The (infinite) Taylor series expansion does not necessarily converge at all, or to $f(x)$. Functions for which it does are called *analytic*. **?** is an example of a function which is not analytic.

2.9 The Fundamental Theorem of Calculus

This theorem sets out the precise rules for cancelling integration and differentiation operations.

Theorem 2.9.1 (Fundamental Theorem of Calculus) *The integration and differentiation operators are inverses in the following senses:*

1. $$\frac{d}{db}\int_a^b f(x)dx = f(b)$$

2. $$\int_a^b f'(x)dx = f(b) - f(a)$$

This can be illustrated graphically using a picture illustrating the use of integration to compute the area under a curve.

[2] Should this not be the closed interval?

Chapter 3

CONVEXITY AND OPTIMISATION

3.1 Introduction

[To be written.]

3.2 Convexity and Concavity

3.2.1 Definitions

Definition 3.2.1 *A subset X of a vector space is a* convex set
$\Longleftrightarrow$
$\forall \mathbf{x}, \mathbf{x}' \in X, \lambda \in [0,1],\ \lambda\mathbf{x} + (1-\lambda)\mathbf{x}' \in X.$

Theorem 3.2.1 *A sum of convex sets, such as*

$$X + Y \equiv \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\},$$

is also a convex set.

Proof The proof of this result is left as an exercise.

Q.E.D.

Definition 3.2.2 *Let $f : X \to Y$ where X is a convex subset of a real vector space and $Y \subseteq \Re$. Then*

1. f *is a* convex function

   $\iff$

   $\forall \mathbf{x} \neq \mathbf{x}' \in X, \lambda \in (0,1)$

$$f\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'\right) \leq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}'). \tag{3.2.1}$$

   (This just says that a function of several variables is convex if its restriction to every *line segment in its domain is a convex function of one variable in the familiar sense.)*

2. f *is a* concave function

   $\iff$

   $\forall \mathbf{x} \neq \mathbf{x}' \in X, \lambda \in (0,1)$

$$f\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'\right) \geq \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}'). \tag{3.2.2}$$

3. f *is affine*

   $\iff$

   f *is both convex and concave.*[1]

Note that the conditions (3.2.1) and (3.2.2) could also have been required to hold (equivalently)

$$\forall \mathbf{x}, \mathbf{x}' \in X, \lambda \in [0,1]$$

since they are satisfied as equalities $\forall f$ when $\mathbf{x} = \mathbf{x}'$, when $\lambda = 0$ and when $\lambda = 1$.
Note that f is convex $\iff$ $-f$ is concave.

Definition 3.2.3 *Again let $f : X \to Y$ where X is a convex subset of a real vector space and $Y \subseteq \Re$. Then*

1. f *is a* strictly convex function

   $\iff$

   $\forall \mathbf{x} \neq \mathbf{x}' \in X, \lambda \in (0,1)$

$$f\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'\right) < \lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}').$$

[1] A *linear* function is an affine function which also satisfies $f(\mathbf{0}) = 0$.

2. f *is a* strictly concave function

   $\Longleftrightarrow$

   $\forall \mathbf{x} \neq \mathbf{x}' \in X, \lambda \in (0,1)$

$$f\left(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}'\right) > \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{x}').$$

Note that there is no longer any flexibility as regards allowing $\mathbf{x} = \mathbf{x}'$ or $\lambda = 0$ or $\lambda = 1$ in these definitions.

3.2.2 Properties of concave functions

Note the connection between convexity of a function of several variables and convexity of the restrictions of that function to any line in its domain: the former is convex if and only if all the latter are.
Note that a function on a multidimensional vector space, X, is convex if and only if the restriction of the function to the line L is convex for every line L in X, and similarly for concave, strictly convex, and strictly concave functions.
Since every convex function is the mirror image of a concave function, and *vice versa*, every result derived for one has an obvious corollary for the other. In general, we will consider only concave functions, and leave the derivation of the corollaries for convex functions as exercises.
Let $f : X \to \Re$ and $g : X \to \Re$ be concave functions. Then

1. If $a, b > 0$, then $af + bg$ is concave.

2. If $a < 0$, then af is convex.

3. $\min\{f, g\}$ is concave

The proofs of the above properties are left as exercises.

Definition 3.2.4 *Consider the real-valued function* $f\colon X \to Y$ *where* $Y \subseteq \Re$.

1. *The upper contour sets of* f *are the sets* $\{\mathbf{x} \in X : f(\mathbf{x}) \geq \alpha\}$ *(*$\alpha \in \Re$*).*

2. *The level sets or indifference curves of* f *are the sets* $\{\mathbf{x} \in X : f(\mathbf{x}) = \alpha\}$ *(*$\alpha \in \Re$*).*

3. *The lower contour sets of* f *are the sets* $\{\mathbf{x} \in X : f(\mathbf{x}) \leq \alpha\}$ *(*$\alpha \in \Re$*).*

In Definition 3.2.4, X does not have to be a (real) vector space.

Theorem 3.2.2 *The upper contour sets* $\{\mathbf{x} \in X : f(\mathbf{x}) \geq \alpha\}$ *of a concave function are convex.*

Proof This proof is probably in a problem set somewhere.

Q.E.D.

Consider as an aside the two-good consumer problem. Note in particular the implications of Theorem 3.2.2 for the shape of the indifference curves corresponding to a concave utility function. Concave u is a sufficient but not a necessary condition for convex upper contour sets.

3.2.3 Convexity and differentiability

In this section, we show that there are a total of three ways of characterising concave functions, namely the definition above, a theorem in terms of the first derivative (Theorem 3.2.3) and a theorem in terms of the second derivative or Hessian (Theorem 3.2.4).

Theorem 3.2.3 *[Convexity criterion for differentiable functions.] Let* $f : X \to \Re$ *be differentiable,* $X \subseteq \Re^n$ *an open, convex set. Then:*
f *is (strictly) concave*
$\iff$
$\forall \mathbf{x} \neq \mathbf{x}' \in X$,

$$f(\mathbf{x}) \leq (<) f(\mathbf{x}') + f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}'). \tag{3.2.3}$$

Theorem 3.2.3 says that a function is concave if and only if the tangent hyperplane at any point lies completely above the graph of the function, or that a function is concave if and only if for any two distinct points in the domain, the directional derivative at one point in the direction of the other exceeds the jump in the value of the function between the two points. (See Section 2.7 for the definition of a directional derivative.)

Proof (See **?**.)

1. We first prove that the weak version of inequality 3.2.3 is necessary for concavity, and then that the strict version is necessary for strict concavity.

   Choose $\mathbf{x}, \mathbf{x}' \in X$.

(a) Suppose that f is concave.
Then, for $\lambda \in (0, 1)$,

$$f\left(\mathbf{x}' + \lambda(\mathbf{x} - \mathbf{x}')\right) \geq f(\mathbf{x}') + \lambda\left(f(\mathbf{x}) - f(\mathbf{x}')\right). \tag{3.2.4}$$

Subtract $f(\mathbf{x}')$ from both sides and divide by λ:

$$\frac{f\left(\mathbf{x}' + \lambda(\mathbf{x} - \mathbf{x}')\right) - f(\mathbf{x}')}{\lambda} \geq f(\mathbf{x}) - f(\mathbf{x}'). \tag{3.2.5}$$

Now consider the limits of both sides of this inequality as $\lambda \to 0$. The LHS tends to $f'\left(\mathbf{x}'\right)\left(\mathbf{x} - \mathbf{x}'\right)$ by definition of a directional derivative (see (2.7.2) and (2.7.3) above). The RHS is independent of λ and does not change. The result now follows easily for concave functions. However, 3.2.5 remains a weak inequality even if f is a strictly concave function.

(b) Now suppose that f is strictly concave and $\mathbf{x} \neq \mathbf{x}'$.
Since f is also concave, we can apply the result that we have just proved to $\mathbf{x}'$ and $\mathbf{x}'' \equiv \frac{1}{2}\left(\mathbf{x} + \mathbf{x}'\right)$ to show that

$$f'(\mathbf{x}')(\mathbf{x}'' - \mathbf{x}') \geq f(\mathbf{x}'') - f(\mathbf{x}'). \tag{3.2.6}$$

Using the definition of strict concavity (or the strict version of inequality (3.2.4)) gives:

$$f(\mathbf{x}'') - f(\mathbf{x}') > \frac{1}{2}\left(f(\mathbf{x}) - f(\mathbf{x}')\right). \tag{3.2.7}$$

Combining these two inequalities and multiplying across by 2 gives the desired result.

2. Conversely, suppose that the derivative satisfies inequality (3.2.3). We will deal with concavity. To prove the theorem for strict concavity, just replace all the weak inequalities ($\geq$) with strict inequalities ($>$), as indicated.

   Set $\mathbf{x}' = \lambda\mathbf{x} + (1-\lambda)\mathbf{x}''$. Then, applying the hypothesis of the proof in turn to $\mathbf{x}$ and $\mathbf{x}'$ and to $\mathbf{x}''$ and $\mathbf{x}'$ yields:

$$f(\mathbf{x}) \leq f(\mathbf{x}') + f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}') \tag{3.2.8}$$

and

$$f(\mathbf{x}'') \leq f(\mathbf{x}') + f'(\mathbf{x}')(\mathbf{x}'' - \mathbf{x}') \tag{3.2.9}$$

   A convex combination of (3.2.8) and (3.2.9) gives:

$$\lambda f(\mathbf{x}) + (1-\lambda)f(\mathbf{x}'')$$

$$\begin{aligned} &\leq f(\mathbf{x}') + f'(\mathbf{x}')\left(\lambda\left((\mathbf{x}-\mathbf{x}')\right) + (1-\lambda)\left((\mathbf{x}''-\mathbf{x}')\right)\right) \\ &= f(\mathbf{x}'), \end{aligned} \tag{3.2.10}$$

since

$$\lambda\left((\mathbf{x}-\mathbf{x}')\right) + (1-\lambda)\left((\mathbf{x}''-\mathbf{x}')\right) = \lambda\mathbf{x} + (1-\lambda)\mathbf{x}'' - \mathbf{x}' = \mathbf{0}_n. \tag{3.2.11}$$

(3.2.10) is just the definition of concavity as required.

Q.E.D.

Theorem 3.2.4 *[Concavity criterion for twice differentiable functions.] Let $f : X \to \Re$ be twice continuously differentiable (C^2), $X \subseteq \Re^n$ open and convex. Then:*

1. *f is concave*

   $\Longleftrightarrow$

   $\forall \mathbf{x} \in X$, the Hessian matrix $f''(\mathbf{x})$ is negative semidefinite.

2. *$f''(\mathbf{x})$ negative definite $\forall \mathbf{x} \in X$*

   $\Rightarrow$

   f is strictly concave.

The fact that the condition in the second part of this theorem is sufficient but not necessary for concavity inspires the search for a counter-example, in other words for a function which is strictly concave but has a second derivative which is only negative semi-definite and not strictly negative definite. The standard counter-example is given by $f(x) = x^{2n}$ for any integer $n > 1$.

Proof We first use Taylor's theorem to demonstrate the sufficiency of the condition on the Hessian matrices. Then we use the Fundamental Theorem of Calculus (Theorem 2.9.1) and a proof by contrapositive to demonstrate the necessity of this condition in the concave case for $n = 1$. Then we use this result and the Chain Rule to demonstrate necessity for $n > 1$. Finally, we show how these arguments can be modified to give an alternative proof of sufficiency for functions of one variable.

1. Suppose first that $f''(\mathbf{x})$ is negative semi-definite $\forall \mathbf{x} \in X$. Recall Taylor's Theorem above (Theorem 2.8.1).

   It follows that $f(\mathbf{x}) \leq f(\mathbf{x}') + f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}')$. Theorem 3.2.3 shows that f is then concave. A similar proof will work for negative definite Hessian and strictly concave function.

2. To demonstrate necessity, we must consider separately first functions of a single variable and then functions of several variables.

   (a) First consider a function of a single variable. Instead of trying to show that concavity of f implies a negative semi-definite (*i.e.* non-positive) second derivative $\forall x \in X$, we will prove the contrapositive. In other words, we will show that if there is any point $x^* \in X$ where the second derivative is positive, then f is locally strictly convex around x^* and so cannot be concave.

   So suppose $f''(x^*) > 0$. Then, since f is twice continuously differentiable, $f''(x) > 0$ for all x in some neighbourhood of x^*, say (a, b). Then f' is an increasing function on (a, b). Consider two points in (a, b), $x < x''$ and let $x' = \lambda x + (1 - \lambda)x'' \in X$, where $\lambda \in (0, 1)$. Using the fundamental theorem of calculus,

$$f(x') - f(x) = \int_x^{x'} f'(t)dt < f'(x')(x' - x)$$

   and

$$f(x'') - f(x') = \int_{x'}^{x''} f'(t)dt < f'(x')(x'' - x').$$

   Rearranging each inequality gives:

$$f(x) > f(x') + f'(x')(x - x')$$

   and

$$f(x'') > f(x') + f'(x')(x'' - x'),$$

   which are just the single variable versions of (3.2.8) and (3.2.9). As in the proof of Theorem 3.2.3, a convex combination of these inequalities reduces to

$$f(x') < \lambda f(x) + (1 - \lambda)f(x''),$$

   and hence f is locally strictly convex on (a, b).

   (b) Now consider a function of several variables. Suppose that f is concave and fix $\mathbf{x} \in X$ and $\mathbf{h} \in \Re^n$. (We use an $\mathbf{x}$, $\mathbf{x}+\mathbf{h}$ argument instead of an $\mathbf{x}$, $\mathbf{x}'$ argument to tie in with the definition of a negative definite matrix.) Then, at least for sufficiently small λ, $g(\lambda) \equiv f(\mathbf{x} + \lambda\mathbf{h})$ also defines a concave function (of one variable), namely the restriction of f to the line segment from $\mathbf{x}$ in the direction from $\mathbf{x}$ to $\mathbf{x} + \mathbf{h}$. Thus, using the result we have just proven for functions of one variable, g has non-positive second derivative. But we know from p. 25 above that $g''(0) = \mathbf{h}^\top f''(\mathbf{x})\mathbf{h}$, so $f''(\mathbf{x})$ is negative semi-definite.

3. For functions of one variable, the above arguments can give an alternative proof of sufficiency which does not require Taylor's Theorem. In fact, we have something like the following:

$$\begin{array}{lcl} f''(x) < 0 \text{ on } (a,b) & \Rightarrow & f \text{ locally strictly concave on } (a,b) \\ f''(x) \leq 0 \text{ on } (a,b) & \Rightarrow & f \text{ locally concave on } (a,b) \\ f''(x) > 0 \text{ on } (a,b) & \Rightarrow & f \text{ locally strictly convex on } (a,b) \\ f''(x) \geq 0 \text{ on } (a,b) & \Rightarrow & f \text{ locally convex on } (a,b) \end{array}$$

   The same results which we have demonstrated for the interval (a, b) also hold for the entire domain X (which of course is also just an open interval, as it is an open convex subset of $\Re$).

Q.E.D.

Theorem 3.2.5 *A non-decreasing twice differentiable concave transformation of a twice differentiable concave function (of several variables) is also concave.*

Proof The details are left as an exercise.

Q.E.D.

Note finally the implied hierarchy among different classes of functions:
negative definite Hessian $\subset$ strictly concave $\subset$ concave $=$ negative semidefinite Hessian.
As an exercise, draw a Venn diagram to illustrate these relationships (and add other classes of functions to it later on as they are introduced).
The second order condition above is reminiscent of that for optimisation and suggests that concave or convex functions will prove useful in developing theories of optimising behaviour. In fact, there is a wider class of useful functions, leading us to now introduce further definitions.

3.2.4 Variations on the convexity theme

Let $X \subseteq \Re^n$ be a convex set and $f : X \to \Re$ a real-valued function defined on X. In order (for reasons which shall become clear in due course) to maintain consistency with earlier notation, we adopt the convention when labelling vectors $\mathbf{x}$ and $\mathbf{x}'$ that $f(\mathbf{x}') \leq f(\mathbf{x})$.[2]

[2] There may again be some lapses in this version.

Definition 3.2.5 *Let*

$$C(\alpha) = \{\mathbf{x} \in X : f(\mathbf{x}) \geq \alpha\}.$$

Then $f : X \to \Re$ *is* quasiconcave
$\iff$
$\forall \alpha \in \Re,\ C(\alpha)$ *is a convex set.*

Theorem 3.2.6 *The following statements are equivalent to the definition of quasiconcavity:*

1. $\forall \mathbf{x} \neq \mathbf{x}' \in X$ *such that* $f(\mathbf{x}') \leq f(\mathbf{x})$ *and* $\forall \lambda \in (0, 1)$,
$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}') \geq f(\mathbf{x}').$$
2. $\forall \mathbf{x}, \mathbf{x}' \in X, \lambda \in [0, 1], f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}') \geq \min\{f(\mathbf{x}), f(\mathbf{x}')\}$.
3. *(If f is differentiable,)* $\forall \mathbf{x}, \mathbf{x}' \in X$ *such that* $f(\mathbf{x}) - f(\mathbf{x}') \geq 0$, $f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}') \geq 0$.

Proof 1. We begin by showing the equivalence between the definition and condition 1.[3]

(a) First suppose that the upper contour sets are convex. Let $\mathbf{x}$ and $\mathbf{x}' \in X$ and let $\alpha = \min\{f(\mathbf{x}), f(\mathbf{x}')\}$. Then $\mathbf{x}$ and $\mathbf{x}'$ are in $C(\alpha)$. By the hypothesis of convexity, for any $\lambda \in (0, 1)$, $\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}' \in C(\alpha)$. The desired result now follows.

(b) Now suppose that condition 1 holds. To show that $C(\alpha)$ is a convex set, we just take $\mathbf{x}$ and $\mathbf{x}' \in C(\alpha)$ and investigate whether $\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}' \in C(\alpha)$. But, by our previous result,

$$f(\lambda \mathbf{x} + (1 - \lambda)\mathbf{x}') \geq \min\{f(\mathbf{x}), f(\mathbf{x}')\} \geq \alpha$$

where the final inequality holds because $\mathbf{x}$ and $\mathbf{x}'$ are in $C(\alpha)$.

2. It is almost trivial to show the equivalence of conditions 1 and 2.

(a) In the case where $f(\mathbf{x}) \geq f(\mathbf{x}')$ or $f(\mathbf{x}') = \min\{f(\mathbf{x}), f(\mathbf{x}')\}$, there is nothing to prove. Otherwise, we can just reverse the labels $\mathbf{x}$ and $\mathbf{x}'$. The statement is true for $\mathbf{x} = \mathbf{x}'$ or $\lambda = 0$ or $\lambda = 1$ even if f is not quasiconcave.

[3] This proof may need to be rearranged to reflect the choice of a different equivalent characterisation to act as definition.

(b) The proof of the converse is even more straightforward and is left as an exercise.

3. Proving that condition 3 is equivalent to quasiconcavity for a differentiable function (by proving that it is equivalent to conditions 1 and 2) is much the trickiest part of the proof.

   (a) Begin by supposing that f satisfies conditions 1 and 2. Proving that condition 3 is necessary for quasiconcavity is the easier part of the proof (and appears as an exercise on one of the problem sets). Pick any $\lambda \in (0,1)$ and, without loss of generality, $\mathbf{x}$ and $\mathbf{x}'$ such that $f(\mathbf{x}') \leq f(\mathbf{x})$. By quasiconcavity,

$$f\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'\right) \geq f(\mathbf{x}'). \tag{3.2.12}$$

   Consider

$$f|_L(\lambda) = f\left(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}'\right) = f\left(\mathbf{x}' + \lambda(\mathbf{x}-\mathbf{x}')\right). \tag{3.2.13}$$

   We want to show that the directional derivative

$$f|_L'(0) = f'(\mathbf{x}')(\mathbf{x}-\mathbf{x}') \geq 0. \tag{3.2.14}$$

   But

$$f|_L'(0) = \lim_{\lambda\to 0} \frac{f\left(\mathbf{x}' + \lambda(\mathbf{x}-\mathbf{x}')\right) - f(\mathbf{x}')}{\lambda}. \tag{3.2.15}$$

   Since the right hand side is non-negative for small positive values of λ ($\lambda < 1$), the derivative must be non-negative as required.

   (b) Now the difficult part — to prove that condition 3 is a sufficient condition for quasiconcavity.

   Suppose the derivative satisfies the hypothesis of the theorem, but f is not quasiconcave. In other words, $\exists \mathbf{x}, \mathbf{x}', \lambda^*$ such that

$$f\left(\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}'\right) < \min\{f(\mathbf{x}), f(\mathbf{x}')\}, \tag{3.2.16}$$

   where without loss of generality $f(\mathbf{x}') \leq f(\mathbf{x})$. The hypothesis of the theorem applied first to $\mathbf{x}$ and $\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}'$ and then to $\mathbf{x}'$ and $\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}'$ tells us that:

$$f'\left(\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}'\right)\left(\mathbf{x} - (\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}')\right) \geq 0 \tag{3.2.17}$$
$$f'\left(\lambda^*\mathbf{x} + (1-\lambda^*)\mathbf{x}'\right)\left(\mathbf{x}' - (\lambda^*\mathbf{x}' + (1-\lambda^*)\mathbf{x})\right) \geq 0. \tag{3.2.18}$$

Multiplying the first inequality by $(1-\lambda^*)$ and the second by λ^* yields a pair of inequalities which can only be satisfied simultaneously if:

$$f'\left(\lambda^*\mathbf{x}+(1-\lambda^*)\mathbf{x}'\right)(\mathbf{x}'-\mathbf{x})=0. \tag{3.2.19}$$

In other words, $f|'_L(\lambda^*)=0$; we already know that $f|_L(\lambda^*)<f|_L(0)\leq f|_L(1)$. We can apply the same argument to any point where the value of $f|_L$ is less than $f|_L(0)$ to show that the corresponding part of the graph of $f|_L$ has zero slope, or is flat. But this is incompatible either with continuity of $f|_L$ or with the existence of a point where $f|_L(\lambda^*)$ is strictly less than $f|_L(0)$. So we have a contradiction as required.

Q.E.D.

In words, part 3 of Theorem 3.2.6 says that whenever a differentiable quasiconcave function has a higher value at $\mathbf{x}$ than at $\mathbf{x}'$, or the same value at both points, then the directional derivative of f at $\mathbf{x}'$ in the direction of $\mathbf{x}$ is non-negative. It might help to think about this by considering $n=1$ and separating out the cases $x>x'$ and $x<x'$.

Theorem 3.2.7 *Let $f: X\rightarrow\Re$ be quasiconcave and $g:\Re\rightarrow\Re$ be increasing. Then $g\circ f$ is a quasiconcave function.*

Proof This follows easily from the previous result. The details are left as an exercise.

Q.E.D.

Note the implications of Theorem 3.2.7 for utility theory. In particular, if preferences can be represented by a quasiconcave utility function, then they can be represented by a quasiconcave utility function only. This point will be considered again in a later section of the course.

Definition 3.2.6 *$f: X\rightarrow\Re$ is strictly quasiconcave*

$\Longleftrightarrow$

$\forall\mathbf{x}\neq\mathbf{x}'\in X$ *such that* $f(\mathbf{x})\geq f(\mathbf{x}')$ *and* $\forall\lambda\in(0,1)$, $f\left(\lambda\mathbf{x}+(1-\lambda)\mathbf{x}'\right)>f(\mathbf{x}')$.

Definition 3.2.7 *f is (strictly) quasiconvex*

$\Longleftrightarrow$

$-f$ *is (strictly) quasiconcave*[4]

[4]EC3080 ended here for Hilary Term 1998.

Definition 3.2.8 *f is pseudoconcave*
$\iff$
f is differentiable and quasiconcave and

$$f(\mathbf{x}) - f(\mathbf{x}') > 0 \quad \Rightarrow \quad f'(\mathbf{x}')\,(\mathbf{x} - \mathbf{x}') > 0.$$

Note that the last definition modifies slightly the condition in Theorem 3.2.6 which is equivalent to quasiconcavity for a differentiable function.
Pseudoconcavity will crop up in the second order condition for equality constrained optimisation.
We conclude this section by looking at a couple of the functions of several variables which will crop repeatedly in applications in economics later on.
First, consider the interesting case of the affine function

$$f\colon \Re^n \to \Re\colon \mathbf{x} \mapsto M - \mathbf{p}^\top \mathbf{x},$$

where $M \in \Re$ and $\mathbf{p} \in \Re^n$. This function is both concave and convex, but neither strictly concave nor strictly convex. Furthermore,

$$\begin{aligned} f\left(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{x}'\right) &= \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{x}') && (3.2.20)\\ &\geq \min\{f(\mathbf{x}), f(\mathbf{x}')\} && (3.2.21)\end{aligned}$$

and

$$\begin{aligned} (-f)\left(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{x}'\right) &= \lambda(-f)(\mathbf{x}) + (1-\lambda)(-f)(\mathbf{x}') && (3.2.22)\\ &\geq \min\{(-f)(\mathbf{x}), (-f)(\mathbf{x}')\}, && (3.2.23)\end{aligned}$$

so f is both quasiconcave and quasiconcave, but not strictly so in either case. f is, however, pseudoconcave (and pseudoconvex) since

$$\begin{aligned} f(\mathbf{x}) > f(\mathbf{x}') &\iff \mathbf{p}^\top \mathbf{x} < \mathbf{p}^\top \mathbf{x}' && (3.2.24)\\ &\iff \mathbf{p}^\top\left(\mathbf{x} - \mathbf{x}'\right) < 0 && (3.2.25)\\ &\iff -f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}') < 0 && (3.2.26)\\ &\iff f'(\mathbf{x}')(\mathbf{x} - \mathbf{x}') > 0. && (3.2.27)\end{aligned}$$

Finally, here are two graphs of Cobb-Douglas functions:[5]

[5] To see them you will have to have copied two `.WMF` files retaining their uppercase filenames and put them in the appropriate directory,
`C:/TCD/teaching/WWW/MA381/NOTES/`!

Graph of $z = x^{0.5}y^{0.5}$

Graph of $z = x^{-0.5}y^{1.5}$

3.3 Unconstrained Optimisation

Definition 3.3.1 *Let $X \subseteq \Re^n$, $f : X \to \Re$.*
Then f has a (strict) global maximum at $\mathbf{x}^*$ $\iff$ $\forall \mathbf{x} \in X, \mathbf{x} \neq \mathbf{x}^*, f(\mathbf{x}) \leq (<)f(\mathbf{x}^*)$.
Also f has a (strict) local maximum at $\mathbf{x}^*$ $\iff$ $\exists \epsilon > 0$ such that $\forall \mathbf{x} \in B\epsilon(\mathbf{x}^*), \mathbf{x} \neq \mathbf{x}^*, f(\mathbf{x}) \leq (<)f(\mathbf{x}^*)$.

Similarly for minima.

Theorem 3.3.1 *A continuous real-valued function on a compact subset of $\Re^n$ attains a global maximum and a global minimum.*

Proof Not given here. See **?**.

Q.E.D.

While this is a neat result for functions on compact domains, results in calculus are generally for functions on open domains, since the limit of the first difference

of the function at x, say, makes no sense if the function and the first difference are not defined in some open neighbourhood of x (some $B\epsilon(x)$).
The remainder of this section and the next two sections are each centred around three related theorems:

1. a theorem giving necessary or first order conditions which must be satisfied by the solution to an optimisation problem (Theorems 3.3.2, 3.4.1 and 3.5.1);

2. a theorem giving sufficient or second order conditions under which a solution to the first order conditions satisfies the original optimisation problem (Theorems 3.3.3, 3.4.2 and 3.5.2); and

3. a theorem giving conditions under which a known solution to an optimisation problem is the unique solution (Theorems 3.3.4, 3.4.3 and 3.5.3).

The results are generally presented for maximisation problems. However, any minimisation problem is easily turned into a maximisation problem by reversing the sign of the function to be minimised and maximising the function thus obtained.
Throughout the present section, we deal with the unconstrained optimisation problem

$$\max_{\mathbf{x}\in X} f(\mathbf{x}) \tag{3.3.1}$$

where $X \subseteq \Re^n$ and $f : X \to \Re$ is a real-valued function of several variables, called the *objective function* of Problem (3.3.1.)
(It is conventional to use the letter λ both to parameterise convex combinations and as a Lagrange multiplier. To avoid confusion, in this section we switch to the letter α for the former usage.)

Theorem 3.3.2 *Necessary (first order) condition for unconstrained maxima and minima.*
Let X be open and f differentiable with a local maximum or minimum at $\mathbf{x}^* \in X$. Then $f'(\mathbf{x}^*) = \mathbf{0}$, or f has a stationary point *at $\mathbf{x}^*$.*

Proof Without loss of generality, assume that the function has a local maximum at $\mathbf{x}^*$. Then $\exists \epsilon > 0$ such that, whenever $\|\mathbf{h}\| < \epsilon$,

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \leq 0.$$

It follows that, for $0 < h < \epsilon$,

$$\frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} \leq 0,$$

(where $\mathbf{e}_i$ denotes the ith standard basis vector) and hence that

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = \lim_{h \to 0} \frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} \leq 0. \tag{3.3.2}$$

Similarly, for $0 > h > -\epsilon$,

$$\frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} \geq 0,$$

and hence

$$\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = \lim_{h \to 0} \frac{f(\mathbf{x}^* + h\mathbf{e}_i) - f(\mathbf{x}^*)}{h} \geq 0. \tag{3.3.3}$$

Combining (3.3.2) and (3.3.3) yields the desired result.

Q.E.D.

The first order conditions are only useful for identifying optima in the interior of the domain of the objective function: Theorem 3.3.2 applies only to functions whose domain X is open. Other methods must be used to check for possible corner solutions or boundary solutions to optimisation problems where the objective function is defined on a domain that is not open.

Theorem 3.3.3 *Sufficient (second order) condition for unconstrained maxima and minima.*
Let $X \subseteq \Re^n$ be open and let $f : X \to \Re$ be a twice continuously differentiable function with $f'(\mathbf{x}^*) = \mathbf{0}$ and $f''(\mathbf{x}^*)$ negative definite.
Then f has a strict local maximum at $\mathbf{x}^*$.
Similarly for positive definite Hessians and local minima.

Proof Consider the second order Taylor expansion used previously in the proof of Theorem 3.2.4: for any $\mathbf{x} \in X$, $\exists s \in (0, 1)$ such that

$$f(\mathbf{x}) = f(\mathbf{x}^*) + f'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top f''(\mathbf{x}^* + s(\mathbf{x} - \mathbf{x}^*))(\mathbf{x} - \mathbf{x}^*).$$

or, since the first derivative vanishes at $\mathbf{x}^*$,

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top f''(\mathbf{x}^* + s(\mathbf{x} - \mathbf{x}^*))(\mathbf{x} - \mathbf{x}^*).$$

Since f'' is continuous, $f''(\mathbf{x}^* + s(\mathbf{x} - \mathbf{x}^*))$ will also be negative definite for $\mathbf{x}$ in some open neighbourhood of $\mathbf{x}^*$. Hence, for $\mathbf{x}$ in this neighbourhood, $f(\mathbf{x}) < f(\mathbf{x}^*)$ and f has a strict local maximum at $\mathbf{x}^*$.

Q.E.D.

The weak form of this result does not hold. In other words, semi-definiteness of the Hessian matrix at $\mathbf{x}^*$ is not sufficient to guarantee that f has any sort of maximum at $\mathbf{x}^*$. For example, if $f(x) = x^3$, then the Hessian is negative semi-definite at $x = 0$ but the function does not have a local maximum there (rather, it has a *point of inflexion*).

Theorem 3.3.4 *Uniqueness conditions for unconstrained maximisation.*
If

1. $\mathbf{x}^*$ *solves Problem (3.3.1) and*
2. f *is strictly quasiconcave (presupposing that* X *is a convex set),*

then $\mathbf{x}^*$ *is the unique (global) maximum.*

Proof Suppose not, in other words that $\exists \mathbf{x} \neq \mathbf{x}^*$ such that $f(\mathbf{x}) = f(\mathbf{x}^*)$.
Then, for any $\alpha \in (0, 1)$,

$$f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{x}^*) > \min\{f(\mathbf{x}), f(\mathbf{x}^*)\} = f(\mathbf{x}^*),$$

so f does not have a maximum at either $\mathbf{x}$ or $\mathbf{x}^*$.
This is a contradiction, so the maximum must be unique.

Q.E.D.

Theorem 3.3.5 *Tempting, but not quite true, corollaries of Theorem 3.3.3 are:*

- *Every stationary point of a twice continuously differentiable strictly concave function is a strict global maximum (and so there can be at most one stationary point).*
- *Every stationary point of a twice continuously differentiable strictly convex function is a strict global minimum.*

Proof If the Hessian matrix is positive/negative definite everywhere, then the argument in the proof of Theorem 3.3.3 can be applied for $\mathbf{x} \in X$ and not just for $\mathbf{x} \in B\epsilon(\mathbf{x}^*)$. If there are points at which the Hessian is merely semi-definite, then the proof breaks down.

Q.E.D.

Note that many strictly concave and strictly convex functions will have no stationary points, for example

$$f \colon \Re \to \Re \colon x \mapsto e^x.$$

3.4 Equality Constrained Optimisation: The Lagrange Multiplier Theorems

Throughout this section, we deal with the equality constrained optimisation problem

$$\begin{array}{rl} & \max_{\mathbf{x} \in X} f(\mathbf{x}) \\ \text{s.t.} & g(\mathbf{x}) = \mathbf{0}_m \end{array} \tag{3.4.1}$$

where $X \subseteq \Re^n$, $f : X \to \Re$ is a real-valued function of several variables, called the *objective function* of Problem (3.4.1) and $g : X \to \Re^m$ is a vector-valued function of several variables, called the *constraint function* of Problem (3.4.1); or, equivalently, $g^j : X \to \Re$ are real-valued functions for $j = 1, \ldots, m$. In other words, there are m scalar constraints represented by a single vector constraint:

$$\begin{pmatrix} g^1(\mathbf{x}) \\ \vdots \\ g^m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

We will introduce and motivate the *Lagrange multiplier method* which applies to such constrained optimisation problems with equality constraints. We will assume where appropriate that the objective function f and the m constraint functions $g^1, \ldots, g^m$ are all once or twice continuously differentiable.

The entire discussion here is again presented in terms of maximisation, but can equally be presented in terms of minimisation by reversing the sign of the objective function. Similarly, note that the signs of the constraint function(s) can be reversed without altering the underlying problem. We will see, however, that this also reverses the signs of the corresponding *Lagrange multipliers*. The significance of this effect will be seen from the formal results, which are presented here in terms of the usual three theorems.

Before moving on to those formal results, we briefly review the methodology for solving constrained optimisation problems which should be familiar from introductory and intermediate economic analysis courses.

If $\mathbf{x}^*$ is a solution to Problem (3.4.1), then there exist Lagrange multipliers,[6] $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_m)$, such that

$$f'(\mathbf{x}^*) + \boldsymbol{\lambda}^\top g'(\mathbf{x}^*) = \mathbf{0}_n.$$

Thus, to find the constrained optimum, we proceed as if optimising the *Lagrangean*:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}^*) + \boldsymbol{\lambda}^\top g(\mathbf{x}^*).$$

Note that

[6] As usual, a row of numbers separated by commas is used as shorthand for a column vector.

1. $\mathcal{L} = f$ whenever $g = \mathbf{0}$ and
2. $g = \mathbf{0}$ where $\mathcal{L}$ is optimised.

Roughly speaking, this is why the constrained optimum of f corresponds to the optimum of $\mathcal{L}$.
The Lagrange multiplier method involves the following four steps:

1. introduce the m Lagrange multipliers, $\boldsymbol{\lambda} \equiv (\lambda_1, \ldots, \lambda_m)$.
2. Define the Lagrangean $\mathcal{L}$: $X \times \Re^m \to \Re$, where

$$X \times \Re^m \equiv \{(\mathbf{x}, \boldsymbol{\lambda}) : \mathbf{x} \in X, \boldsymbol{\lambda} \in \Re^m\},$$

   by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \equiv f(\mathbf{x}) + \boldsymbol{\lambda}^\top g(\mathbf{x}).$$

3. Find the stationary points of the Lagrangean, *i.e.* set $\mathcal{L}'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$. Since the Lagrangean is a function of $n+m$ variables, this gives $n+m$ first order conditions. The first n are

$$f'(\mathbf{x}) + \boldsymbol{\lambda}^\top g'(\mathbf{x}) = \mathbf{0}$$

   or

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j \frac{\partial g^j}{\partial x_i}(\mathbf{x}) = 0 \quad i = 1, \ldots, n.$$

   The last m are just the original constraints,

$$g(\mathbf{x}) = \mathbf{0}$$

   or

$$g^j(\mathbf{x}) = 0 \quad j = 1, \ldots, m.$$

4. Finally, the second order conditions must be checked.

As an example, consider maximisation of a utility function representing Cobb-Douglas preferences subject to a budget constraint.
The first n first order or Lagrangean conditions say that the total derivative (or gradient) of f at $\mathbf{x}$ is a linear combination of the total derivatives (or gradients) of the constraint functions at $\mathbf{x}$.
Consider a picture with $n = 2$ and $m = 1$.
Since the directional derivative along a tangent to a level set or indifference curve is zero at the point of tangency, $\mathbf{x}$, (the function is at a maximum or minimum

along the tangent) or $f'(\mathbf{x})(\mathbf{x}' - \mathbf{x}) = 0$, the gradient vector, $f'(\mathbf{x})^\top$, must be perpendicular to the direction of the tangent, $\mathbf{x}' - \mathbf{x}$.
At the optimum, the level sets of f and g have a common tangent, so $f'(\mathbf{x})$ and $g'(\mathbf{x})$ are collinear, or $f'(\mathbf{x}) = -\lambda g'(\mathbf{x})$. It can also be seen with a little thought that for the solution to be a local constrained maximum, λ must be positive if g is quasiconcave and negative if g is quasiconvex (in either case, the constraint curve is the boundary of a convex set).
We now consider the equality constrained optimisation problem in more depth.

Theorem 3.4.1 *First order (necessary) conditions for optimisation with equality constraints.*
Consider problem (3.4.1) or the corresponding minimisation problem.
If

1. $\mathbf{x}^*$ *solves this problem (which implies that* $g(\mathbf{x}^*) = \mathbf{0}$*),*
2. f *and* g *are continuously differentiable, and*
3. *the* $m \times n$ *matrix*

$$g'(\mathbf{x}^*) = \begin{pmatrix} \frac{\partial g^1}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g^1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g^m}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g^m}{\partial x_n}(\mathbf{x}^*) \end{pmatrix}$$

is of rank m *(i.e. there are no redundant constraints, both in the sense that there are fewer constraints than variables and in the sense that the constraints which are present are 'independent'),*

then $\exists \boldsymbol{\lambda}^* \in \Re^m$ *such that* $f'(\mathbf{x}^*) + \boldsymbol{\lambda}^{*\top} g'(\mathbf{x}^*) = \mathbf{0}$ *(i.e. in* $\Re^n$*,* $f'(\mathbf{x}^*)$ *is in the* $m-$*dimensional subspace generated by the* m *vectors* $g^{1'}(\mathbf{x}^*), \dots, g^{m'}(\mathbf{x}^*)$*).*

Proof The idea is to solve $g(\mathbf{x}^*) = \mathbf{0}$ for m variables as a function of the other $n - m$ and to substitute the solution into the objective function to give an unconstrained problem with $n - m$ variables.
For this proof, we need Theorem 2.6.1 on p. 2.6.1 above, the Implicit Function Theorem. Using this theorem, we must find the m weights $\lambda_1, \dots, \lambda_m$ to prove that $f'(\mathbf{x}^*)$ is a linear combination of $g^{1'}(\mathbf{x}^*), \dots, g^{m'}(\mathbf{x}^*)$.
Without loss of generality, we assume that the first m columns of $g'(\mathbf{x}^*)$ are linearly independent (if not, then we merely relabel the variables accordingly).
Now we can partition the vector $\mathbf{x}^*$ as $(\mathbf{y}^*, \mathbf{z}^*)$ and, using the notation of the Implicit Function Theorem, find a neighbourhood Z of $\mathbf{z}^*$ and a function h defined on Z such that

$$g(h(\mathbf{z}), \mathbf{z}) = \mathbf{0} \quad \forall \mathbf{z} \in Z$$

and also

$$h'(\mathbf{z}^*) = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g.$$

Now define a new objective function $F\colon Z \to \Re$ by

$$F(\mathbf{z}) \equiv f(h(\mathbf{z}), \mathbf{z}).$$

Since $\mathbf{x}^*$ solves the constrained problem $\max_{\mathbf{x}\in X} f(\mathbf{x})$ subject to $g(\mathbf{x}) = \mathbf{0}$, it follows that $\mathbf{z}^*$ solves the unconstrained problem $\max_{\mathbf{z}\in Z} F(\mathbf{z})$. (This is easily shown using a proof by contradiction argument.)
Hence, $\mathbf{z}^*$ satisfies the first order conditions for unconstrained maximisation of F, namely

$$F'(\mathbf{z}^*) = \mathbf{0}.$$

Applying the Chain Rule in exactly the same way as in the proof of the Implicit Function Theorem yields an equation which can be written in shorthand as:

$$D_{\mathbf{y}}fh'(\mathbf{z}) + D_{\mathbf{z}}f = \mathbf{0}.$$

Substituting for $h'(\mathbf{z})$ gives:

$$D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g = D_{\mathbf{z}}f.$$

We can also partition $f'(\mathbf{x})$ as

$$\begin{pmatrix} D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1} D_{\mathbf{y}}g & D_{\mathbf{z}}f \end{pmatrix}$$

Substituting for the second sub-matrix yields:

$$\begin{aligned} f'(\mathbf{x}) &= \begin{pmatrix} D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1} D_{\mathbf{y}}g & D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g \end{pmatrix} \\ &= D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1} \begin{pmatrix} D_{\mathbf{y}}g & D_{\mathbf{z}}g \end{pmatrix} \\ &= -\boldsymbol{\lambda}^{\top} g'(\mathbf{x}) \end{aligned}$$

where we define

$$\boldsymbol{\lambda} \equiv -D_{\mathbf{y}}f(D_{\mathbf{y}}g)^{-1}.$$

Q.E.D.

Theorem 3.4.2 *Second order (sufficient or concavity) conditions for maximisation with equality constraints.*
If

1. *f and g are differentiable,*

2. *$f'(\mathbf{x}^*) + \boldsymbol{\lambda}^{*\top} g'(\mathbf{x}^*) = \mathbf{0}$ (i.e. the first order conditions are satisfied at $\mathbf{x}^*$),*

3. *$\lambda_j^* \geq 0$ for $j = 1, \ldots, m$,*

4. *f is pseudoconcave, and*

5. *g^j is quasiconcave for $j = 1, \ldots, m$,*

then $\mathbf{x}^*$ solves the constrained maximisation problem.

It should be clear that non-positive Lagrange multipliers and quasiconvex constraint functions can take the place of non-negative Lagrange multipliers and quasiconcave Lagrange multipliers to give an alternative set of second order conditions.

Proof Suppose that the second order conditions are satisfied, but that $\mathbf{x}^*$ is not a constrained maximum. We will derive a contradiction.
Since $\mathbf{x}^*$ is not a maximum, $\exists \mathbf{x} \neq \mathbf{x}^*$ such that $g(\mathbf{x}) = \mathbf{0}$ but $f(\mathbf{x}) > f(\mathbf{x}^*)$.
By pseudoconcavity, $f(\mathbf{x}) - f(\mathbf{x}^*) > 0$ implies that $f'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) > 0$.
Since the constraints are satisfied at both $\mathbf{x}$ and $\mathbf{x}^*$, we have $g(\mathbf{x}^*) = g(\mathbf{x}) = \mathbf{0}$.
By quasiconcavity of the constraint functions (see Theorem 3.2.6), $g^j(\mathbf{x}) - g^j(\mathbf{x}^*) = 0$ implies that $g^{j\prime}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$.
By assumption, all the Lagrange multipliers are non-negative, so

$$f'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + \boldsymbol{\lambda}^{*\top} g'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) > 0.$$

Rearranging yields:

$$\left[f'(\mathbf{x}^*) + \boldsymbol{\lambda}^{*\top} g'(\mathbf{x}^*)\right](\mathbf{x} - \mathbf{x}^*) > 0.$$

But the first order condition guarantees that the LHS of this inequality is zero (not positive), which is the required contradiction.

Q.E.D.

Theorem 3.4.3 *Uniqueness condition for equality constrained maximisation.*
If

1. *$\mathbf{x}^*$ is a solution,*

2. *f is strictly quasiconcave, and*

3. *g^j is an affine function (i.e. both convex and concave) for $j = 1, \ldots, m$,*

then $\mathbf{x}^*$ is the unique (global) maximum.

Proof The uniqueness result is also proved by contradiction. Note that it does not require any differentiability assumption.

- We first show that the feasible set is convex.

  Suppose $\mathbf{x} \neq \mathbf{x}^*$ are two distinct solutions.

  Consider the convex combination of these two solutions $\mathbf{x}\alpha \equiv \alpha\mathbf{x}+(1-\alpha)\,\mathbf{x}^*$.

  Since each g^j is affine and $g^j\left(\mathbf{x}^*\right) = g^j\left(\mathbf{x}\right) = 0$, we have

  $$g^j\left(\mathbf{x}\alpha\right) = \alpha g^j\left(\mathbf{x}\right) + (1-\alpha)\, g^j\left(\mathbf{x}^*\right) = 0.$$

  In other words, $\mathbf{x}\alpha$ also satisfies the constraints.

- To complete the proof, we find the required contradiction:

  Since f is strictly quasiconcave and $f\left(\mathbf{x}^*\right) = f\left(\mathbf{x}\right)$, it must be the case that $f\left(\mathbf{x}\alpha\right) > f\left(\mathbf{x}^*\right)$.

Q.E.D.

The construction of the obvious corollaries for minimisation problems is left as an exercise.
We conclude this section with

Theorem 3.4.4 (Envelope Theorem.) *Consider the modified constrained optimisation problem:*

$$\max_{\mathbf{x}} f\left(\mathbf{x}, \boldsymbol{\alpha}\right) \quad \text{subject to} \quad g\left(\mathbf{x}, \boldsymbol{\alpha}\right) = \mathbf{0}, \tag{3.4.2}$$

where $\mathbf{x} \in \Re^n$, $\boldsymbol{\alpha} \in \Re^q$, $f\colon \Re^{n+q} \rightarrow \Re$ *and* $g\colon \Re^{n+q} \rightarrow \Re^m$ *(i.e. as usual f is the real-valued objective function and g is a vector of m real-valued constraint functions, but either or both can depend on* exogenous *or* control *variables* $\boldsymbol{\alpha}$ *as well as on the* endogenous *or* choice *variables* $\mathbf{x}$*).*
Suppose that the standard conditions for application of the Lagrange multiplier theorems are satisfied.
Let $\mathbf{x}^*\left(\boldsymbol{\alpha}\right)$ *denote the optimal choice of* $\mathbf{x}$ *for given* $\boldsymbol{\alpha}$ *(*$\mathbf{x}^*\colon \Re^q \rightarrow \Re^n$ *is called the* optimal response function*) and let* $M\left(\boldsymbol{\alpha}\right)$ *denote the maximum value attainable by f for given* $\boldsymbol{\alpha}$ *(*$M\colon \Re^q \rightarrow \Re$ *is called the* envelope function*).*
Then the partial derivative of M with respect to α_i is just the partial derivative of the relevant Lagrangean, $f + \boldsymbol{\lambda}^\top g$*, with respect to α_i, evaluated at the optimal value of* $\mathbf{x}$*. The dependence of the vector of Lagrange multipliers,* $\boldsymbol{\lambda}$*, on the vector* $\boldsymbol{\alpha}$ *should be ignored in calculating the last-mentioned partial derivative.*

Proof The Envelope Theorem can be proved in the following steps:

1. Write down the identity relating the functions M, f and $\mathbf{x}^*$:

$$M(\boldsymbol{\alpha}) \equiv f(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha})$$

2. Use (2.5.4) to derive an expression for the partial derivatives $\frac{\partial M}{\partial \alpha_i}$ of M in terms of the partial derivatives of f and $\mathbf{x}^*$:

$$M'(\boldsymbol{\alpha}) = D_{\mathbf{x}} f(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \mathbf{x}^{*\prime}(\boldsymbol{\alpha}) + D_{\boldsymbol{\alpha}} f(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}).$$

3. The first order (necessary) conditions for constrained optimisation say that

$$D_{\mathbf{x}} f(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = -\boldsymbol{\lambda}^\top D_{\mathbf{x}} g(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha})$$

   and allow us to eliminate the $\frac{\partial f}{\partial x_i}$ terms from this expression.

4. Apply (2.5.4) again to the identity $g(\mathbf{x}^*, \boldsymbol{\alpha}) = \mathbf{0}_m$ to obtain

$$D_{\mathbf{x}} g(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \mathbf{x}^{*\prime}(\boldsymbol{\alpha}) + D_{\boldsymbol{\alpha}} g(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = \mathbf{0}_{m \times q}.$$

   Finally, use this result to eliminate the $\frac{\partial g}{\partial x_i}$ terms from your new expression for $\frac{\partial M}{\partial \alpha_i}$.

   Combining all these results gives:

$$M'(\boldsymbol{\alpha}) = D_{\boldsymbol{\alpha}} f(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}) + \boldsymbol{\lambda}^\top D_{\boldsymbol{\alpha}} g(\mathbf{x}^*(\boldsymbol{\alpha}), \boldsymbol{\alpha}),$$

   which is the required result.

Q.E.D.

In applications in economics, the most frequently encountered applications will make sufficient assumptions to guarantee that

1. $\mathbf{x}^*$ satisfies the first order conditions with each $\lambda_i \geq 0$;
2. the Hessian $f''(\mathbf{x}^*)$ is a negative definite ($n \times n$) matrix; and
3. g is a linear function,

so that $\mathbf{x}^*$ is the unique optimal solution to the equality constrained optimisation problem.

3.5 Inequality Constrained Optimisation: The Kuhn-Tucker Theorems

Throughout this section, we deal with the inequality constrained optimisation problem

$$\begin{aligned} &\max_{\mathbf{x}\in X} f(\mathbf{x}) \\ \text{s.t.} \quad & g^i(\mathbf{x}) \geq 0, \quad i = 1, 2, \ldots, m \end{aligned} \tag{3.5.1}$$

where once again $X \subseteq \Re^n$, $f : X \to \Re$ is a real-valued function of several variables, called the *objective function* of Problem (3.5.1) and $g : X \to \Re^m$ is a vector-valued function of several variables, called the *constraint function* of Problem (3.5.1).
Before presenting the usual theorems formally, we need some graphical motivation concerning the interpretation of Lagrange multipliers.
Suppose that the constraint functions are given by

$$g(\mathbf{x}, \boldsymbol{\alpha}) = \boldsymbol{\alpha} - h(\mathbf{x}). \tag{3.5.2}$$

The Lagrangean is

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^\top (\boldsymbol{\alpha} - h(\mathbf{x})). \tag{3.5.3}$$

Thus, using the Envelope Theorem, it is easily seen that the rate of change of the envelope function $M(\boldsymbol{\alpha})$ with respect to the 'level' of the ith underlying constraint function h^i is:

$$\frac{\partial M}{\partial \alpha_i} = \frac{\partial \mathcal{L}}{\partial \alpha_i} = \lambda_i. \tag{3.5.4}$$

Thus

- when $\lambda_i = 0$, the envelope function is at its maximum, or the objective function at its unconstrained maximum, and the constraint is not binding.
- when $\lambda_i < 0$, the envelope function is decreasing
- when $\lambda_i > 0$, the envelope function is increasing

Now consider how the nature of the inequality constraint change as α_i increases (as illustrated, assuming f quasiconcave as usual and h^i quasiconvex or g^i quasiconcave, so that the relationship between α_i and λ_i is negative)

$$h^i(\mathbf{x}) \leq \alpha_i \tag{3.5.5}$$

Type of constraint	**Derivative of objective fn.**	**Lagrange Multiplier**	**Constraint fn.**
Binding/active	$f' \leq 0$	$\lambda \geq 0$	$g = 0$
Non-binding/inactive	$f' = 0$	$\lambda = 0$	$g > 0$

Table 3.1: Sign conditions for inequality constrained optimisation

(or $g^i(\mathbf{x}, \boldsymbol{\alpha}) \geq 0$). For values of α_i such that $\lambda_i > 0$, this constraint is strictly binding. For values of α_i such that $\lambda_i = 0$, this constraint is just binding. For values of α_i such that $\lambda_i < 0$, this constraint is non-binding.
Note that in this setup,

$$\frac{\partial^2 M}{\partial \alpha_i^2} = \frac{\partial \lambda_i}{\partial \alpha_i} < 0, \tag{3.5.6}$$

so that the envelope function is strictly concave (up to the unconstrained optimum, and constant beyond it).
Thus we will find that part of the *necessity* conditions below is that the Lagrange multipliers be non-negative. (For equality constrained optimisation, the signs were important only when dealing with second order conditions).
Consider also at this stage the first order conditions for maximisation of a function of one variable subject to a non-negativity constraint:

$$\max_x f(x) \text{ s.t. } x \geq 0.$$

They can be expressed as:

$$f'(x^*) \leq 0 \tag{3.5.7}$$
$$f'(x^*) = 0 \quad \text{if } x > 0 \tag{3.5.8}$$

The various sign conditions which we have looked at are summarised in Table 3.1.

Theorem 3.5.1 *Necessary (first order) conditions for optimisation with inequality constraints.*
If

1. $\mathbf{x}^*$ *solves Problem (3.5.1), with*

$$g^i(\mathbf{x}^*) = 0, i = 1, 2, \ldots, b$$

and

$$g^i(\mathbf{x}^*) > 0, i = b + 1, \ldots, m$$

(in other words, the first b constraints are binding *(active) at $\mathbf{x}^*$ and the last $n - b$ are* non-binding *(inactive) at $\mathbf{x}^*$, renumbering the constraints if necessary to achieve this),*

2. *f and g are continuously differentiable, and*

3. *the $b \times n$ submatrix of $g'(\mathbf{x}^*)$,*

$$\begin{pmatrix} \frac{\partial g^1}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g^1}{\partial x_n}(\mathbf{x}^*) \\ \vdots & \ddots & \vdots \\ \frac{\partial g^b}{\partial x_1}(\mathbf{x}^*) & \dots & \frac{\partial g^b}{\partial x_n}(\mathbf{x}^*) \end{pmatrix},$$

   is of full rank b (i.e. there are no redundant binding constraints, both in the sense that there are fewer binding constraints than variables and in the sense that the constraints which are binding are 'independent'),

then $\exists \boldsymbol{\lambda} \in \Re^m$ such that $f'(\mathbf{x}^*) + \boldsymbol{\lambda}^\top g'(\mathbf{x}^*) = \mathbf{0}$, with $\lambda_i \geq 0$ for $i = 1, 2, \dots, m$ and $g^i(\mathbf{x}^*) = 0$ if $\lambda_i > 0$.

Proof The proof is similar to that of Theorem 3.4.1 for the equality constrained case. It can be broken into seven steps.

1. Suppose $\mathbf{x}^*$ solves Problem (3.5.1).

   We begin by restricting attention to a neighbourhood $B_\epsilon(\mathbf{x}^*)$ throughout which the non-binding constraints remain non-binding, *i.e.*

$$g^i(\mathbf{x}) > 0 \ \forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*), i = b+1, \dots, m. \tag{3.5.9}$$

   Such a neighbourhood exists since the constraint functions are continuous. Since $\mathbf{x}^*$ solves Problem (3.5.1) by assumption, it also solves the following problem:

$$\begin{array}{rl} & \max_{\mathbf{x} \in B_\epsilon(\mathbf{x}^*)} f(\mathbf{x}) \\ \text{s.t.} & g^i(\mathbf{x}) \geq 0, \quad i = 1, 2, \dots, b. \end{array} \tag{3.5.10}$$

   In other words, since the non-binding constraints are non-binding $\forall \mathbf{x} \in B_\epsilon(\mathbf{x}^*)$ by construction, we can ignore them if we confine our search for a maximum to this neighbourhood. We will return to the non-binding constraints in the very last step of this proof, but until then g will be taken to refer to the vector of b binding constraint functions only and $\boldsymbol{\lambda}$ to the vector of b Kuhn-Tucker multipliers corresponding to these binding constraints.

2. We now introduce slack variables $\mathbf{s} \equiv (s_1, \dots, s_b)$, one corresponding to each binding constraint, and consider the following equality constrained maximisation problem:

$$\max_{\mathbf{x} \in B_\epsilon(\mathbf{x}^*), \mathbf{s} \in \Re^b_+} f(\mathbf{x})$$

$$\text{s.t.} \quad G(\mathbf{x}, \mathbf{s}) = \mathbf{0}_b \tag{3.5.11}$$

where $G^i(\mathbf{x}, \mathbf{s}) \equiv g^i(\mathbf{x}) - s_i$, $i = 1, 2, \ldots, b$.

Since $\mathbf{x}^*$ solves Problem (3.5.10) and all b constraints in that problem are binding at $\mathbf{x}^*$, it can be seen that $(\mathbf{x}^*, \mathbf{0}_b)$ solves this new problem. For consistency of notation, we define $\mathbf{s}^* \equiv \mathbf{0}_b$.

3. We proceed with Problem (3.5.11) as in the Lagrange case. In other words, we use the Implicit Function Theorem to solve the system of b equations in $n + b$ unknowns,

$$G(\mathbf{x}, \mathbf{s}) = \mathbf{0}_b, \tag{3.5.12}$$

for the first b variables in terms of the last n. To do this, we partition the vector of choice and slack variables three ways:

$$(\mathbf{x}, \mathbf{s}) \equiv (\mathbf{y}, \mathbf{z}, \mathbf{s}) \tag{3.5.13}$$

where $\mathbf{y} \in \Re^b$ and $\mathbf{z} \in \Re^{n-b}$, and correspondingly partition the matrix of partial derivatives evaluated at the optimum:

$$\begin{aligned} G'(\mathbf{y}^*, \mathbf{z}^*, \mathbf{s}^*) &= G'(\mathbf{x}^*, \mathbf{s}^*) && (3.5.14) \\ &= \begin{pmatrix} D_{\mathbf{y}}G & D_{\mathbf{z},\mathbf{s}}G \end{pmatrix} && (3.5.15) \\ &= \begin{pmatrix} D_{\mathbf{y}}G & D_{\mathbf{z}}G & D_{\mathbf{s}}G \end{pmatrix} && (3.5.16) \\ &= \begin{pmatrix} D_{\mathbf{y}}g & D_{\mathbf{z}}g & -\mathbf{I}_b \end{pmatrix}. && (3.5.17) \end{aligned}$$

The rank condition allows us to apply the Implicit Function Theorem and to find a function $h : \Re^n \longrightarrow \Re^b$ such that $\mathbf{y} = h(\mathbf{z}, \mathbf{s})$ is a solution to $G(\mathbf{y}, \mathbf{z}, \mathbf{s}) = \mathbf{0}$ with

$$h'(\mathbf{z}^*, \mathbf{s}^*) = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z},\mathbf{s}}G. \tag{3.5.18}$$

(3.5.18) can in turn be partitioned to yield

$$D_{\mathbf{z}}h = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}G = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{z}}g \tag{3.5.19}$$

and

$$D_{\mathbf{s}}h = -(D_{\mathbf{y}}g)^{-1} D_{\mathbf{s}}G = (D_{\mathbf{y}}g)^{-1} \mathbf{I}_b = (D_{\mathbf{y}}g)^{-1}. \tag{3.5.20}$$

4. This solution can be substituted into the original objective function f to create a new objective function F defined by

$$F(\mathbf{z}, \mathbf{s}) \equiv f(h(\mathbf{z}, \mathbf{s}), \mathbf{z}) \tag{3.5.21}$$

and another new maximisation problem where there are only (implicit) non-negativity constraints:

$$\max_{\mathbf{z}\in B_\epsilon(\mathbf{z}^*),\mathbf{s}\in\Re^b_+} F(\mathbf{z},\mathbf{s}) \tag{3.5.22}$$

It should be clear that $\mathbf{z}^*, \mathbf{0}_b$ solves Problem (3.5.22). The first order conditions for Problem (3.5.22) are just that the partial derivatives of F with respect to the remaining $n-b$ choice variables equal zero (according to the first order conditions for unconstrained optimisation), while the partial derivatives of F with respect to the b slack variables must be less than or equal to zero.

5. The Kuhn-Tucker multipliers can now be found exactly as in the Lagrange case. We know that

$$D_{\mathbf{z}}F = D_{\mathbf{y}}fD_{\mathbf{z}}h + D_{\mathbf{z}}f\mathbf{I}_{n-b} \quad = \quad \mathbf{0}_{n-b}.$$

Substituting for $D_{\mathbf{z}}h$ from (3.5.19) gives:

$$D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}D_{\mathbf{z}}g = D_{\mathbf{z}}f.$$

We can also partition $f'(\mathbf{x})$ as

$$\begin{pmatrix} D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}D_{\mathbf{y}}g & D_{\mathbf{z}}f \end{pmatrix}$$

Substituting for the second sub-matrix yields:

$$\begin{aligned} f'(\mathbf{x}) &= \begin{pmatrix} D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}D_{\mathbf{y}}g & D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}D_{\mathbf{z}}g \end{pmatrix} \\ &= D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}\begin{pmatrix} D_{\mathbf{y}}g & D_{\mathbf{z}}g \end{pmatrix} \\ &\equiv -\boldsymbol{\lambda}^\top g'(\mathbf{x}) \end{aligned}$$

where we define the Kuhn-Tucker multipliers corresponding to the binding constraints, $\boldsymbol{\lambda}$, by

$$\boldsymbol{\lambda} \equiv -D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1}. \tag{3.5.23}$$

6. Next, we calculate the partial derivatives of F with respect to the slack variables and show that they can be less than or equal to zero if and only if the Kuhn-Tucker multipliers corresponding to the binding constraints are greater than or equal to zero. This can be seen by differentiating both sides of (3.5.21) with respect to s to obtain:

$$D_{\mathbf{s}}F = D_{\mathbf{y}}fD_{\mathbf{s}}h + D_{\mathbf{z}}f\mathbf{0}_{(n-b)\times b} \tag{3.5.24}$$

$$= D_{\mathbf{y}}f\left(D_{\mathbf{y}}g\right)^{-1} \tag{3.5.25}$$

$$= -\boldsymbol{\lambda}, \tag{3.5.26}$$

where we have used (3.5.20) and (3.5.23).

7. Finally just set the Kuhn-Tucker multipliers corresponding to the non-binding constraints equal to zero.

Q.E.D.

Theorem 3.5.2 *Second order (sufficient or concavity) conditions for optimisation with inequality constraints.*
If

1. f *and* g *are differentiable,*

2. $\exists \boldsymbol{\lambda} \in \Re^m$ *such that* $f'(\mathbf{x}^*) + \boldsymbol{\lambda}^\top g'(\mathbf{x}^*) = \mathbf{0}$*, with* $\lambda_i \geq 0$ *for* $i = 1, 2, \ldots, m$ *and* $g^i(\mathbf{x}^*) = 0$ *if* $\lambda_i > 0$ *(i.e. the first order conditions are satisfied at* $\mathbf{x}^*$*),*

3. f *is pseudoconcave, and*

4. $g^i(\mathbf{x}^*) = 0, i = 1, 2, \ldots, b$*;* $g^i(\mathbf{x}^*) > 0, i = b+1, \ldots, m$ *and* g^j *is quasiconcave for* $j = 1, \ldots, b$ *(i.e. the binding constraint functions are quasiconcave),*

then $\mathbf{x}^*$ *solves the constrained maximisation problem.*

Proof The proof just requires the first order conditions to be reduced to

$$f'(\mathbf{x}^*) + \sum_{i=1}^{b} \lambda_i g^{i\prime}(\mathbf{x}^*) = \mathbf{0} \qquad (3.5.27)$$

from where it is virtually identical to that for the Lagrange case, and so it is left as an exercise.

Q.E.D.

Theorem 3.5.3 *Uniqueness condition for inequality constrained optimisation.*
If

1. $\mathbf{x}^*$ *is a solution,*

2. f *is strictly quasiconcave, and*

3. g^j *is a quasiconcave function for* $j = 1, \ldots, m$*,*

then $\mathbf{x}^*$ *is the unique (global) optimal solution.*

Proof The proof is again similar to that for the Lagrange case and is left as an exercise. The point to note this time is that the feasible set with equality constraints was convex if the constraint functions were affine, whereas the feasible set with inequality constraints is convex if the constraint functions are quasiconcave. This is because the feasible set (where all the inequality constraints are satisfied simultaneously) is the intersection of m upper contour sets of quasiconcave functions, or the intersection of m convex sets.

Q.E.D.

The last important result on optimisation, the Theorem of the Maximum, is closely related to the Envelope Theorem. Before proceeding to the statement of the theorem, the reader may want to review Definition 2.3.12.

Theorem 3.5.4 (Theorem of the maximum) *Consider the modified inequality constrained optimisation problem:*

$$\max_{\mathbf{x}} f(\mathbf{x}, \boldsymbol{\alpha}) \quad \text{subject to} \quad g^i(\mathbf{x}, \boldsymbol{\alpha}) \geq 0, i = 1, \ldots, m \tag{3.5.28}$$

where $\mathbf{x} \in \Re^n$, $\boldsymbol{\alpha} \in \Re^q$, $f: \Re^{n+q} \to \Re$ *and* $g: \Re^{n+q} \to \Re^m$.
Let $\mathbf{x}^*(\boldsymbol{\alpha})$ *denote the optimal choice of* $\mathbf{x}$ *for given* $\boldsymbol{\alpha}$ ($\mathbf{x}^*: \Re^q \to \Re^n$) *and let* $M(\boldsymbol{\alpha})$ *denote the maximum value attainable by* f *for given* $\boldsymbol{\alpha}$ ($M: \Re^q \to \Re$).
If

1. f *is continuous;*
2. *the range of* f *is closed and bounded; and*
3. *the constraint set is a non-empty, compact-valued, continuous correspondence of* $\boldsymbol{\alpha}$,

then

1. M *is a continuous (single-valued) function; and*
2. $\mathbf{x}^*$ *is an upper hemi-continuous correspondence, and hence is continuous if it is a continuous (single-valued) function.*

Proof The proof of this theorem is omitted, along with some of the more technical material on continuity of (multi-valued) correspondences.

Q.E.D.

Theorem 3.5.4 will be used in consumer theory to prove such critical results as the continuity of demand functions derived from the maximisation of continuous utility functions.
The following are two frequently encountered examples illustrating the use of the Kuhn-Tucker theorems in economics.[7]

[7] The calculations should be left as exercises.

1. The canonical quadratic programming problem.

   Find the vector $\mathbf{x} \in \Re^n$ which maximises the value of the quadratic form $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ subject to the m linear inequality constraints $\mathbf{g}^{i\top}\mathbf{x} \geq \alpha_i$, where $\mathbf{A} \in \Re^{n\times n}$ is negative definite and $\mathbf{g}^i \in \Re^n$ for $i = 1, \ldots, m$.

   The objective function can always be rewritten as a quadratic form in a symmetric (negative definite) matrixl, since as $\mathbf{x}^\top\mathbf{A}\mathbf{x}$ is a scalar,

$$\mathbf{x}^\top\mathbf{A}\mathbf{x} = \left(\mathbf{x}^\top\mathbf{A}\mathbf{x}\right)^\top \tag{3.5.29}$$
$$= \mathbf{x}^\top\mathbf{A}^\top\mathbf{x} \tag{3.5.30}$$
$$= \frac{1}{2}\left(\mathbf{x}^\top\mathbf{A}\mathbf{x} + \mathbf{x}^\top\mathbf{A}^\top\mathbf{x}\right) \tag{3.5.31}$$
$$= \mathbf{x}^\top\left(\frac{1}{2}\left(\mathbf{A} + \mathbf{A}^\top\right)\right)\mathbf{x} \tag{3.5.32}$$

   and $\frac{1}{2}\left(\mathbf{A} + \mathbf{A}^\top\right)$ is always symmetric.

   Let $\mathbf{G}$ be the $m \times n$ matrix whose ith row is $\mathbf{g}^i$. $\mathbf{G}$ must have full rank if we are to apply the Kuhn-Tucker conditions.

   The Lagrangean is:

$$\mathbf{x}^\top\mathbf{A}\mathbf{x} + \boldsymbol{\lambda}^\top\left(\mathbf{G}\mathbf{x} - \boldsymbol{\alpha}\right). \tag{3.5.33}$$

   The first order conditions are:

$$2\mathbf{x}^\top\mathbf{A} + \boldsymbol{\lambda}^\top\mathbf{G} = \mathbf{0}_n \tag{3.5.34}$$

   or, transposing and multiplying across by $\frac{1}{2}\mathbf{A}^{-1}$:

$$\mathbf{x} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{G}^\top\boldsymbol{\lambda}. \tag{3.5.35}$$

   If the constraints are binding, then we will have:

$$\boldsymbol{\alpha} = -\frac{1}{2}\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\boldsymbol{\lambda}. \tag{3.5.36}$$

   Now we need the fact that $\mathbf{G}$ (and hence $\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top$) has full rank to solve for the Lagrange multipliers $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = -2\left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1}\boldsymbol{\alpha}. \tag{3.5.37}$$

   Now the sign conditions tell us that each component of $\boldsymbol{\lambda}$ must be non-negative. An easy fix is to let the Kuhn-Tucker multipliers be defined by:

$$\boldsymbol{\lambda}^* \equiv \max\left\{\mathbf{0}_m, -2\left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1}\boldsymbol{\alpha}\right\}, \tag{3.5.38}$$

where the $\max$ operator denotes component-by-component maximisation.

The effect of this is to knock out the non-binding constraints (those with negative Lagrange multipliers) from the original problem and the subsequent analysis.

We can now find the optimal $\mathbf{x}$ by substituting for $\boldsymbol{\lambda}$ in (3.5.35) the value of $\boldsymbol{\lambda}^*$ from (3.5.38).

In the case in which all the constraints are binding, the solution is:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{G}^\top \left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1} \boldsymbol{\alpha} \tag{3.5.39}$$

and the envelope function is given by:

$$\begin{aligned} \mathbf{x}^\top \mathbf{A}\mathbf{x} &= \boldsymbol{\alpha}^\top \left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1} \mathbf{G}\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{G}^\top \left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1} \boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^\top \left(\mathbf{G}\mathbf{A}^{-1}\mathbf{G}^\top\right)^{-1} \boldsymbol{\alpha} \\ &= -\frac{1}{2}\boldsymbol{\alpha}^\top \boldsymbol{\lambda}. \end{aligned} \tag{3.5.40}$$

The applications of this problem will include ordinary least squares and generalised least squares regression and the mean-variance portfolio choice problem in finance.

2. Maximising a Cobb-Douglas utility function subject to a budget constraint and non-negativity constraints.

   The applications of this problem will include choice under certainty, choice under uncertainty with log utility where the parameters are reinterpreted as probabilities, the extension to Stone-Geary preferences, and intertemporal choice with log utility, where the parameters are reinterpreted as time discount factors.

Further exercises consider the duals of each of the forgoing problems, and it is to the question of duality that we will turn in the next section.

3.6 Duality

Let $X \subseteq \Re^n$ and let $f, g : X \mapsto \Re$ be, respectively, pseudoconcave and pseudoconvex functions. Consider the envelope functions defined by the dual families of problems:

$$M(\alpha) \equiv \max_{\mathbf{x}} f(\mathbf{x}) \,\mathrm{s.t.}\, g(\mathbf{x}) \leq \alpha \tag{3.6.1}$$

and

$$N(\beta) \equiv \min_{\mathbf{x}} g(\mathbf{x}) \, \text{s.t.} \; f(\mathbf{x}) \geq \beta. \tag{3.6.2}$$

Suppose that these problems have solutions, say $\mathbf{x}^*(\alpha)$ and $\mathbf{x}^\dagger(\beta)$ respectively, and that the constraints bind at these points.
The first order conditions for the two problems are respectively:

$$f'(\mathbf{x}) - \lambda g'(\mathbf{x}) = \mathbf{0} \tag{3.6.3}$$

and

$$g'(\mathbf{x}) + \mu f'(\mathbf{x}) = \mathbf{0}, \tag{3.6.4}$$

where λ and μ are the relevant Lagrange multipliers.
Thus if $\mathbf{x}$ and $\lambda^* \neq 0$ solve (3.6.3), then $\mathbf{x}$ and $\mu^* \equiv -1/\lambda$ solve (3.6.4). However, for the $\mathbf{x}$ which solves the original problem (3.6.1) to also solve (3.6.2), it must also satisfy the constraint, or $f(\mathbf{x}) = \beta$. But we know that $f(\mathbf{x}) = M(\alpha)$. This allows us to conclude that:

$$\mathbf{x}^*(\alpha) = \mathbf{x}^\dagger(M(\alpha)). \tag{3.6.5}$$

Similarly,

$$\mathbf{x}^\dagger(\beta) = \mathbf{x}^*(N(\beta)). \tag{3.6.6}$$

Combining these equations leads to the conclusion that

$$\alpha = N(M(\alpha)) \tag{3.6.7}$$

and

$$\beta = M(N(\beta)). \tag{3.6.8}$$

In other words, the envelope functions for the two dual problems are inverse functions (over any range where the Lagrange multipliers are non-zero, *i.e.* where the constraints are binding). Thus, either α or β or indeed λ or μ can be used to parameterise either family of problems.
We will see many examples of these principles in the applications in the next part of the book. In particular, duality will be covered further in the context of its applications to consumer theory in Section 4.6.

Part II

APPLICATIONS

Chapter 4

CHOICE UNDER CERTAINTY

4.1 Introduction

[To be written.]

4.2 Definitions

There are two possible types of *economy* which we could analyse:

- a *pure exchange economy*, in which households are endowed directly with goods, but there are no firms, there is no production, and economic activity consists solely of pure exchanges of an initial aggregate endowment; and

- a *production economy*, in which households are further indirectly endowed with, and can trade, shares in the profit or loss of firms, which can use part of the initial aggregate endowment as inputs to production processes whose outputs are also available for trade and consumption.

Economies of these types comprise:

- H *households* or *consumers* or *agents* or (later) *investors* or, merely, *individuals*, indexed by the subscript[1] h;

- N *goods* or *commodities*, indexed by the superscript n; and

- (in the case of a production economy only) F *firms*, indexed by f.

[1]Notation in what follows is probably far from consistent in regard to superscripts and subscripts and in regard to ith or nth good and needs fixing.

This chapter concentrates on the theory of optimal consumer choice and of equilibrium in a pure exchange economy. The theory of optimal production decisions and of equilibrium in an economy with production is mathematically similar.
Goods can be distinguished from each other in many ways:

- obviously, by intrinsic physical characteristics, *e.g.* apples or oranges.

- by the time at which they are consumed, *e.g.* an Easter egg delivered before Easter Sunday or an Easter egg delivered after Easter Sunday. (While all trading takes place simultaneously in the model, consumption can be spread out over many periods.)

- by the state of the world in which they are consumed, *e.g.* the service provided by an umbrella on a wet day or the service which it provides on a dry day. (Typically, the state of the world can be any point ω in a sample space Ω.)

The important characteristics of household h are that it is faced with the choice of a *consumption vector* or *consumption plan* or *consumption bundle*, $\mathbf{x}_h = \left(x_h^1, \ldots, x_h^N\right)$, from a (closed, convex) *consumption set*, X_h. Typically, $X_h = \Re_+^N$. More generally, consumer h's consumption set might require a certain subsistence consumption of some commodities, such as water, and rule out points of $\Re_+^N$ not meeting this requirement. The household's endowments are denoted $\mathbf{e}_h \in \Re_+^N$ and can be traded.
In a production economy, the shareholdings of households in firms are denoted $\mathbf{c}_h \in \Re^F$,
A consumer's net demand is denoted $\mathbf{z}_h \equiv \mathbf{x}_h - \mathbf{e}_h \in \Re^N$.
Each consumer is assumed to have a *preference relation* or *(weak) preference ordering* which is a binary relation on the consumption set X_h (**?**, Chapter 7).
Since each household will have different preferences, we should really denote household h's preference relation $\succeq_h$, but the subscript will be omitted for the time being while we consider a single household. Similarly, we will assume for the time being that each household chooses from the same consumption set, $\mathcal{X}$, although this is not essential.
Recall (see **?**) that a binary relation R on $\mathcal{X}$ is just a subset R of $\mathcal{X} \times \mathcal{X}$ or a collection of pairs (x, y) where $x \in \mathcal{X}$ and $y \in \mathcal{X}$.
If $(x, y) \in R$, we usually just write xRy.
Thus $x \succeq y$ means that either x is preferred to y or the consumer is indifferent between the two (*i.e.* that x is at least as good as y).
The following properties of a general relation R on a general set X are often of interest:

1. A relation R is *reflexive*

   $\iff$

   $xRx \; \forall x \in X$

2. A relation R is *symmetric*

   $\iff$

   $xRy \Rightarrow yRx$

3. A relation R is *transitive*

   $\iff$

   $xRy, yRz \Longrightarrow xRz$

4. A relation R is *complete*

   $\iff$

   $\forall x, y \in X$ either xRy or yRx (or both) (in other words a complete relation orders the whole set)

An *indifference relation*, $\sim$, and a *strict preference relation*, $\succ$, can be derived from every preference relation:

1. $x \succ y$ means $x \succeq y$ but not $y \succeq x$

2. $x \sim y$ means $x \succeq y$ and $y \succeq x$.

The *utility function* $u : \mathcal{X} \to \Re$ *represents* the preference relation $\succeq$ if

$$u(x) \geq u(y) \iff x \succeq y.$$

If $f \colon \Re \to \Re$ is a monotonic increasing function and u represents the preference relation $\succeq$, then $f \circ u$ also represents $\succeq$, since

$$f\left(u(x)\right) \geq f\left(u(y)\right) \iff u(x) \geq u(y) \iff x \succeq y.$$

If $\mathcal{X}$ is a countable set, then there exists a utility function representing any preference relation on $\mathcal{X}$. To prove this, just write out the consumption plans in $\mathcal{X}$ in order of preference, and assign numbers to them, assigning the same number to any two or more consumption plans between which the consumer is indifferent.
If $\mathcal{X}$ is an uncountable set, then there may not exist a utility function representing every preference relation on $\mathcal{X}$.

4.3 Axioms

We now consider six axioms which it are frequently assumed to be satisfied by preference relations when considering *consumer choice under certainty*. (Note that symmetry would not be a very sensible axiom!) Section 5.5.1 will consider further axioms that are often added to simplify the analysis of *consumer choice under uncertainty*. After the definition of each axiom, we will give a brief rationale for its use.

Axiom 1 (Completeness) *A (weak) preference relation is complete.*

Completeness means that the consumer is never agnostic.

Axiom 2 (Reflexivity) *A (weak) preference relation is reflexive.*

Reflexivity means that each bundle is at least as good as itself.

Axiom 3 (Transitivity) *A (weak) preference relation is transitive.*

Transitivity means that preferences are rational and consistent.

Axiom 4 (Continuity) *The preference relation $\succeq$ is continuous i.e. for all consumption plans $\mathbf{y} \in \mathcal{X}$ the sets $B_{\mathbf{y}} \equiv \{\mathbf{x} \in \mathcal{X} : \mathbf{x} \succeq \mathbf{y}\}$ and $W_{\mathbf{y}} = \{\mathbf{x} \in \mathcal{X} : \mathbf{y} \succeq \mathbf{x}\}$ are closed sets.*

Consider the picture when $N = 2$:

We will see shortly that $B_{\mathbf{y}}$, the set of consumption plans which are better than or as good as $\mathbf{y}$, and $W_{\mathbf{y}}$, the set of consumption plans which are worse than or as good as $\mathbf{y}$, are just the upper contour sets and lower contour sets respectively of utility functions, if such exist.
E.g. consider *lexicographic preferences*:

Lexicographic preferences violate the continuity axiom. A consumer with such preference prefers more of commodity 1 regardless of the quantities of other commodities, more of commodity 2 if faced with a choice between two consumption plans having the same amount of commodity 1, and so on.
In the picture, the consumption plan $\mathbf{y}$ lies in the lower contour set $W_{\mathbf{x}^*}$ but $B\epsilon\,(\mathbf{y})$ never lies completely in $W_{\mathbf{x}^*}$ for any ϵ. Thus, lower contour sets are not open, and upper contour sets are not closed.
Theorems on the existence of continuous utility functions have been proven by Gerard Debreu, Nobel laureate, whose proof used Axioms 1–4 only (see **?** or **?**) and by Hal Varian, whose proof was simpler by virtue of adding an additional axiom (see **?**).

Theorem 4.3.1 (Debreu) *If $\mathcal{X}$ is a closed and convex set and $\succeq$ is a complete, reflexive, transitive and continuous preference relation on $\mathcal{X}$, then $\exists$ a continuous utility function $u: \mathcal{X} \rightarrow \Re$ representing $\succeq$.*

Proof For the proof of this theorem, see **?**.

Q.E.D.

Axiom 5 (Greed) *Greed is incorporated into consumer behaviour by assuming either*

1. Strong monotonicity:

   If $\mathcal{X} = \Re^N_+$, then $\succeq$ is said to be strongly monotonic *iff whenever $x_n \geq y_n \forall n$ but $\mathbf{x} \neq \mathbf{y}$, $\mathbf{x} \succ \mathbf{y}$. [$\mathbf{x} = (x_1, \ldots, x_N)$ &c.]; or*

2. Local non-satiation:

   $\forall \mathbf{x} \in \mathcal{X}, \epsilon > 0, \exists \mathbf{x}' \in B\epsilon(\mathbf{x})$ *s.t.* $\mathbf{x}' \succ \mathbf{x}$

The strong monotonicity axiom is a much stronger restriction on preferences than local non-satiation; however, it greatly simplifies the proof of existence of utility functions.
We will prove the existence, but not the continuity, part of the following weaker theorem.

Theorem 4.3.2 (Varian) *If $\mathcal{X} = \Re^N_+$ and $\succeq$ is a complete, reflexive, transitive, continuous and strongly monotonic preference relation on $\mathcal{X}$, then $\exists$ a continuous utility function $u: \mathcal{X} \rightarrow \Re$ representing $\succeq$.*

Proof (of existence only (**?**, p.97))
Pick a benchmark consumption plan, *e.g.* $\mathbf{1} \equiv (1, 1, \ldots, 1)$.
The idea is that the utility of $\mathbf{x}$ is the multiple of the benchmark consumption plan to which $\mathbf{x}$ is indifferent.
By strong monotonicity, the sets $\{t \in \Re : t\mathbf{1} \succeq \mathbf{x}\}$ and $\{t \in \Re : \mathbf{x} \succeq t\mathbf{1}\}$ are both non-empty.
By continuity of preferences, both are closed (intersection of a ray through the origin and a closed set); and by completeness, they cover $\Re$.
By connectedness of $\Re$, they intersect in at least one point, $u(\mathbf{x})$ say, and $\mathbf{x} \sim u(\mathbf{x})\mathbf{1}$.
Now

$$\begin{aligned} \mathbf{x} \succeq \mathbf{y} &\iff u(\mathbf{x})\mathbf{1} \succeq u(\mathbf{y})\mathbf{1} \\ &\iff u(\mathbf{x}) \geq u(\mathbf{y}) \end{aligned}$$

where the first equivalence follows from transitivity of preferences and the second from strong monotonicity.
The assumption that preferences are reflexive is not used in establishing existence of the utility function, so it can be inferred that it is required to establish continuity.

Q.E.D.

The rule which the consumer will follow is to choose the most preferred bundle from the set of affordable alternatives (*budget set*), in other words the bundle at which the utility function is maximised subject to the *budget constraint*, if one exists.
We know that an optimal choice will exist if the utility function is continuous and the budget set is closed and bounded.
If the utility function is differentiable, we can go further and use calculus to find the maximum.
So we usually assume differentiability.
If u is a concave utility function, then $f \circ u$, which also represents the same preferences, is not necessarily a concave function (unless f itself is a convex function). In other words, concavity of a utility function is a property of the particular representation and not of the underlying preferences. Notwithstanding this, convexity of preferences is important, as indicated by the use of one or other of the following axioms, each of which relates to the preference relation itself and not to the particular utility function chosen to represent it.

Axiom 6 (Convexity) *There are two versions of this axiom:*

1. Convexity:

   The preference relation $\succeq$ *is* convex $\iff$

   $$\mathbf{x} \succeq \mathbf{y} \Longrightarrow \lambda\mathbf{x} + (1-\lambda)\,\mathbf{y} \succeq \mathbf{y}.$$

2. Strict convexity:

   The preference relation $\succeq$ *is* strictly convex $\iff$

   $$\mathbf{x} \succeq \mathbf{y} \Longrightarrow \lambda\mathbf{x} + (1-\lambda)\,\mathbf{y} \succ \mathbf{y}.$$

The difference between the two versions of the convexity axiom basically amounts to ruling out linear segments in indifference curves in the strict case.

Theorem 4.3.3 *The preference relation* $\succeq$ *is (strictly) convex if and only if every utility function representing* $\succeq$ *is a (strictly) quasiconcave function.*

Proof In either case, both statements are equivalent to saying that

$$u(\mathbf{x}) \geq u(\mathbf{y}) \Longrightarrow u\left(\lambda\mathbf{x} + (1-\lambda)\,\mathbf{y}\right) \geq (>) u(\mathbf{y}). \tag{4.3.1}$$

Q.E.D.

Theorem 4.3.4 *All upper contour sets of a utility function representing convex preferences are convex sets (i.e. indifference curves are convex to the origin for convex preferences).*

It can be seen that convexity of preferences is a generalisation of the two-good assumption of a diminishing marginal rate of substitution.

Axiom 7 (Inada Condition) *This axiom seems to be relegated to some other category in many existing texts and its attribution to Inada is also difficult to trace. It is not easy to see how to express it in terms of the underlying preference relation so perhaps it cannot be elevated to the status of an axiom. Should it be expressed as:*

$$\lim_{x_i \to 0} \frac{\partial u}{\partial x_i} = \infty$$

or as

$$\lim_{x_i \to \infty} \frac{\partial u}{\partial x_i} = 0?$$

The Inada condition will be required to rule out corner solutions in the consumer's problem. Intuitively, it just says that indifference curves may be asymptotic to the axes but never reach them.

4.4 Optimal Response Functions: Marshallian and Hicksian Demand

4.4.1 The consumer's problem

A consumer with consumption set X_h, endowment vector $\mathbf{e}_h \in X_h$, shareholdings $\mathbf{c}_h \in \Re^F$ and preference ordering $\succeq_h$ represented by utility function u_h who desires to trade his endowment at prices $\mathbf{p} \in \Re^N_+$ faces an inequality constrained optimisation problem:

$$\max_{\mathbf{x} \in X_h} u_h(\mathbf{x}) \text{ s.t. } \mathbf{p}^\top \mathbf{x} \leq \mathbf{p}^\top \mathbf{e}_h + \mathbf{c}_h^\top \mathbf{\Pi}(\mathbf{p}) \equiv M_h \qquad (4.4.1)$$

where $\mathbf{\Pi}(\mathbf{p})$ is the vector of the F firms' maximised profits when prices are $\mathbf{p}$. Constraining $\mathbf{x}$ to lie in the consumption set normally just means imposing non-negativity constraints on the problem. From a mathematical point of view, the source of income is irrelevant, and in particular the distinction between pure exchange and production economy is irrelevant. Thus, income can be represented by M in either case.

Since the constraint functions are linear in the choice variables $\mathbf{x}$, the Kuhn-Tucker theorem on second order conditions (Theorem 3.5.2) can be applied, provided that the utility function u_h is pseudo-concave. In this case, the first order conditions identify a maximum.

The Lagrangian, using multipliers λ for the budget constraint and $\boldsymbol{\mu} \in \Re^N$ for the non-negativity constraints, is

$$u_h(\mathbf{x}) + \lambda\left(M - \mathbf{p}^\top \mathbf{x}\right) + \boldsymbol{\mu}^\top \mathbf{x}. \tag{4.4.2}$$

The first order conditions are given by the (N-dimensional) vector equation:

$$u_h'(\mathbf{x}) + \lambda(-\mathbf{p}) + \boldsymbol{\mu} = \mathbf{0}_N \tag{4.4.3}$$

and the sign condition $\lambda \geq 0$ with $\lambda > 0$ if the budget constraint is binding. We also have $\boldsymbol{\mu} = \mathbf{0}_N$ unless one of the non-negativity constraints is binding: Axiom 7 would rule out this possibility.

Now for each $\mathbf{p} \in \Re^N_{++}$ (ruling out bads, or goods with negative prices, and even (see below) free goods), $\mathbf{e}_h \in X_h$, and $\mathbf{c}_h \in \Re^F$, or for each $\mathbf{p}, M_h$ combination, there is a corresponding solution to the consumer's utility maximisation problem, denoted $\mathbf{x}_h(\mathbf{p}, \mathbf{e}_h, \mathbf{c}_h)$ or $\mathbf{x}_h(\mathbf{p}, M_h)$. The function (correspondence) $\mathbf{x}_h$ is often called a Marshallian demand function (correspondence).

If the utility function u_h is also strictly quasiconcave (*i.e.* preferences are strictly convex), then the conditions of the Kuhn-Tucker theorem on uniqueness (Theorem 3.5.3) are satisfied. In this case, the consumer's problem has a unique solution for given prices and income, so that the optimal response correspondence is a single-valued demand function. On the other hand, the weak form of the convexity axiom would permit a multi-valued demand correspondence.

4.4.2 The No Arbitrage Principle

Definition 4.4.1 *An* arbitrage opportunity *means the opportunity to acquire a consumption vector or its constituents, directly or indirectly, at one price, and to sell the same consumption vector or its constituents, directly or indirectly, at a higher price.*

Theorem 4.4.1 (The No Arbitrage Principle) *Arbitrage opportunities do not exist in equilibrium in an economy in which at least one agent has preferences which exhibit local non-satiation.*[2]

[2] The No Arbitrage Principle is also known as the *No Free Lunch Principle*, or the *Law of One Price*

Proof If preferences exhibit local non-satiation, then Marshallian demand is not well-defined if the price vector permits arbitrage opportunities.
If the no arbitrage principle doesn't hold, then any individual can increase wealth without bound by exploiting the available arbitrage opportunity on an infinite scale. Thus there is no longer a budget constraint and, since local non-satiation rules out bliss points, utility too can be increased without bound.
When we come to consider equilibrium, we will see that if even one individual has preferences exhibiting local non-satiation, then equilibrium prices can not permit arbitrage opportunities.

Q.E.D.

Examples are usually in the financial markets, in a multi-period context, *e.g.* covered interest parity, term structure of interest rates, *&c.* The most powerful application is in the derivation of option-pricing formulae, since options can be shown to be identical to various synthetic portfolios made up of the underlying security and the riskfree security.
The simple rule for figuring out how to exploit arbitrage opportunities is 'buy low, sell high'. With interest rates and currencies, for example, this may be a non-trivial calculation.
Exercise: If the interest rate for one-year deposits or loans is r_1 per annum compounded annually, the interest rate for two-year deposits or loans is r_2 per annum compounded annually and the forward interest rate for one year deposits or loans beginning in one year's time is f_{12} per annum compounded annually, calculate the relationship that must hold between these three rates if there are to be no arbitrage opportunities.
Solution: $(1 + r_1)(1 + f_{12}) = (1 + r_2)^2$.

4.4.3 Other Properties of Marshallian demand

Other noteworthy properties of Marshallian demand include the following:

1. If preferences exhibit local non-satiation, then the budget constraint is binding.

   This is because no consumption vector in the interior of the budget set can maximise utility as some nearby consumption vector will always be both preferred and affordable. At the optimum, on the budget hyperplane, the nearby consumption vector which is preferred will not be affordable.

2. If $\mathbf{p}$ includes a zero price ($p^n = 0$ for some n), then $\mathbf{x}_h(\mathbf{p}, M_h)$ may not be well defined.

This is because, at least in the case of strongly monotonic preferences, the consumer will seek to acquire and consume an infinite amount of the free good, thereby increasing utility without bound. For this reason, it is neater to define Marshallian demand only on the open non-negative orthant in $\Re^N$, namely $\Re^N_{++}$.

3. The demand $\mathbf{x}_h(\mathbf{p}, M_h)$ is independent of the representation u_h of the underlying preference relation $\succeq_h$ which is used in the statement of the consumer's problem.

4. $\mathbf{x}_h(\mathbf{p}, M_h)$ is homogenous of degree 0 in $\mathbf{p}, M_h$.

   In other words, if all prices and income are multiplied by $\alpha > 0$, then demand does not change:

$$\mathbf{x}_h(\alpha\mathbf{p}, \alpha M_h) = \mathbf{x}_h(\mathbf{p}, M_h). \tag{4.4.4}$$

5. Demand functions are continuous.

   This follows from the theorem of the maximum (Theorem 3.5.4). It follows that small changes in prices or income will lead to small changes in quantities demanded.

4.4.4 The dual problem

Consider also the (dual) expenditure minimisation problem:

$$\min_{\mathbf{x}} \mathbf{p}^\top \mathbf{x} \; s.t. \; u_h(\mathbf{x}) \geq \bar{u}. \tag{4.4.5}$$

In other words, what happens if expenditure is minimised subject to a certain level of utility, $\bar{u}$, being attained?
The solution (optimal response function) is called the Hicksian or compensated demand function (or correspondence) and is usually denoted $\mathbf{h}_h(\mathbf{p}, \bar{u})$.
There should really be a more general discussion of duality, based on the mean-variance problem as well as the utility maximisation/expenditure minimisation problems.
If the local non-satiation axiom holds, then the constraints are binding in both the utility-maximisation and expenditure-minimisation problems, and we have a number of duality relations. In particular, there will be a one-to-one correspondence between income M and utility $\bar{u}$ for a given price vector $\mathbf{p}$. The expenditure function and the indirect utility function will then act as a pair of inverse envelope functions mapping utility levels to income levels and *vice versa* respectively.

A full page table setting out exactly the parallels between the two problems is called for here.
The following duality relations (or fundamental identities as **?** calls them) will prove extremely useful later on:

$$e(\mathbf{p}, v(\mathbf{p}, M)) = M \tag{4.4.6}$$

$$v(\mathbf{p}, e(\mathbf{p}, \bar{u})) = \bar{u} \tag{4.4.7}$$

$$\mathbf{x}(\mathbf{p}, M) = \mathbf{h}(\mathbf{p}, v(\mathbf{p}, M)) \tag{4.4.8}$$

$$\mathbf{h}(\mathbf{p}, \bar{u}) = \mathbf{x}(\mathbf{p}, e(\mathbf{p}, \bar{u})) \tag{4.4.9}$$

These are just equations (3.6.5)–(3.6.8) adapted to the notation of the consumer's problem.

4.4.5 Properties of Hicksian demands

As in the Marshallian approach, if preferences are strictly convex, then any solution to the expenditure minimisation problem is unique and the Hicksian demands are well-defined single valued functions.
It's worth going back to the uniqueness proof with this added interpretation. If two different consumption vectors minimise expenditure, then they both cost the same amount, and any convex combination of the two also costs the same amount. But by strict convexity, a convex combination yields higher utility, and nearby there must, by continuity, be a cheaper consumption vector still yielding utility $\bar{u}$.
If preferences are not strictly convex, then Hicksian demands may be correspondences rather than functions.
Hicksian demands are homogenous of degree 0 in prices:

$$\mathbf{h}_h(\alpha \mathbf{p}, \bar{u}) = \mathbf{h}_h(\mathbf{p}, \bar{u}) . \tag{4.4.10}$$

4.5 Envelope Functions: Indirect Utility and Expenditure

Now consider the envelope functions corresponding to the two approaches:

1. The indirect utility function:

$$v_h(\mathbf{p}, M) \equiv u_h(\mathbf{x}_h(\mathbf{p}, M)) \tag{4.5.1}$$

2. The expenditure function:

$$e_h(\mathbf{p}, \bar{u}) \equiv \mathbf{p}^\top \mathbf{h}_h(\mathbf{p}, \bar{u}) . \tag{4.5.2}$$

Sometimes we meet two other related functions:

3. The money metric utility function

$$m_h(\mathbf{p}, \mathbf{x}) \equiv e_h(\mathbf{p}, u_h(\mathbf{x})) \tag{4.5.3}$$

is the (least) cost at prices $\mathbf{p}$ of being as well off as with the consumption vector $\mathbf{x}$.

4. The money metric indirect utility function

$$\mu_h(\mathbf{p}; \mathbf{q}, M) \equiv e_h(\mathbf{p}, v_h(\mathbf{q}, M)) \tag{4.5.4}$$

is the (least) cost at prices $\mathbf{p}$ of being as well off as if prices were $\mathbf{q}$ and income was M.

The following are interesting properties of the indirect utility function:

1. By the Theorem of the Maximum, the indirect utility function is continuous for positive prices and income.

2. The indirect utility function is non-increasing in $\mathbf{p}$ and non-decreasing in M.

3. The indirect utility function is quasi-convex in prices.

   To see this, let $B(\mathbf{p})$ denote the budget set when prices are $\mathbf{p}$ and let $\mathbf{p}\lambda \equiv \lambda\mathbf{p} + (1-\lambda)\mathbf{p}'$.

   Then $B(\mathbf{p}\lambda) \subseteq (B(\mathbf{p}) \cup B(\mathbf{p}'))$.

   Suppose this was not the case, *i.e.* for some $\mathbf{x}$, $\mathbf{p}\lambda^\top\mathbf{x} \leq M$ but $\mathbf{p}^\top\mathbf{x} > M$ and $\mathbf{p}'^\top\mathbf{x} > M$. Then taking a convex combination of the last two inequalities yields
   $$\lambda\mathbf{p}^\top\mathbf{x} + (1-\lambda)\mathbf{p}'^\top\mathbf{x} > M,$$
   which contradicts the first inequality.

   It follows that the maximum value of $u_h(\mathbf{x})$ on the subset $B(\mathbf{p}\lambda)$ is less than or equal to its maximum value on the superset $B(\mathbf{p}) \cup B(\mathbf{p}')$.

   In terms of the indirect utility function, this says that

   $$v_h(\mathbf{p}\lambda, M) \leq \max\{v_h(\mathbf{p}, M), v_h(\mathbf{p}', M)\},$$

   or that v_h is quasiconvex.

4. $v_h(\mathbf{p}, M)$ is homogenous of degree zero in $\mathbf{p}, M$, or

   $$v_h(\lambda\mathbf{p}, \lambda M) = v_h(\mathbf{p}, M).$$

The following are interesting properties of the expenditure function:

1. The expenditure function is continuous.

2. The expenditure function itself is non-decreasing in prices, since raising the price of one good while holding the prices of all other goods constant can not reduce the minimum cost of attaining a fixed utility level.

3. The expenditure function is concave in prices.

   To see this, we just fix two price vectors $\mathbf{p}$ and $\mathbf{p}'$ and consider the value of the expenditure function at the convex combination $\mathbf{p}\lambda \equiv \lambda\mathbf{p} + (1-\lambda)\,\mathbf{p}'$.

$$\begin{aligned} e(\mathbf{p}\lambda, \bar{u}) &= (\mathbf{p}\lambda)^\top \mathbf{h}(\mathbf{p}\lambda, \bar{u}) && (4.5.5) \\ &= \lambda\mathbf{p}^\top \mathbf{h}(\mathbf{p}\lambda, \bar{u}) + (1-\lambda)(\mathbf{p}')^\top \mathbf{h}(\mathbf{p}\lambda, \bar{u}) && (4.5.6) \\ &\geq \lambda\mathbf{p}^\top \mathbf{h}(\mathbf{p}, \bar{u}) + (1-\lambda)(\mathbf{p}')^\top \mathbf{h}(\mathbf{p}', \bar{u}) && (4.5.7) \\ &= \lambda e(\mathbf{p}, \bar{u}) + (1-\lambda)\, e(\mathbf{p}', \bar{u}), && (4.5.8) \end{aligned}$$

   where the inequality follows because the cost of a suboptimal bundle for the given prices must be greater than the cost of the optimal (expenditure-minimising) consumption vector for those prices.

4. The expenditure function is homogenous of degree 1 in prices:

$$e_h(\alpha\mathbf{p}, \bar{u}) = \alpha e_h(\mathbf{p}, \bar{u}). \qquad (4.5.9)$$

4.6 Further Results in Demand Theory

In this section, we present four important theorems on demand functions and the corresponding envelope functions. *Shephard's Lemma* will allow us to recover Hicksian demands from the expenditure function. Similarly, *Roy's Identity* will allow us to recover Marshallian demands from the indirect utility function. The *Slutsky symmetry condition* and the *Slutsky equation* provide further insights into the properties of consumer demand.

Theorem 4.6.1 (Shephard's Lemma.)

$$\begin{aligned} \frac{\partial e_h}{\partial p^n}(\mathbf{p}, \bar{u}) &= \frac{\partial}{\partial p^n}\Big(\mathbf{p}^\top\mathbf{x} + \lambda\left(u_h(\mathbf{x}) - \bar{u}\right)\Big) && (4.6.1) \\ &= x^n && (4.6.2) \end{aligned}$$

which, when evaluated at the optimum, is just $\mathbf{h}_h^n(\mathbf{p}, \bar{u})$.
In other words, the partial derivatives of the expenditure function with respect to prices are the corresponding Hicksian demand functions.

Proof By differentiating the expenditure function with respect to the price of good n and applying the envelope theorem (Theorem 3.4.4), we obtain *Shephard's Lemma*:
(To apply the envelope theorem, we should be dealing with an equality constrained optimisation problem; however, if we assume local non-satiation, we know that the budget constraint or utility constraint will always be binding, and so the inequality constrained expenditure minimisation problem is essentially and equality constrained problem.)

Q.E.D.

Theorem 4.6.2 (Roy's Identity.) *Marshallian demands may be recovered from the indirect utility function using:*

$$\mathbf{x}^n(\mathbf{p}, M) = -\frac{\frac{\partial v}{\partial p^n}(\mathbf{p}, M)}{\frac{\partial v}{\partial M}(\mathbf{p}, M)}. \tag{4.6.3}$$

Proof For Roy's Identity, see **?**.
It is obtained by differentiating equation (4.4.7) with respect to p^n, using the Chain Rule:

$$v(\mathbf{p}, e(\mathbf{p}, \bar{u})) = \bar{u}$$

implies that

$$\frac{\partial v}{\partial p^n}(\mathbf{p}, e(\mathbf{p}, \bar{u})) + \frac{\partial v}{\partial M}(\mathbf{p}, e(\mathbf{p}, \bar{u}))\frac{\partial e}{\partial p^n}(\mathbf{p}, \bar{u}) = 0 \tag{4.6.4}$$

and using Shephard's Lemma gives:

$$\frac{\partial v}{\partial p^n}(\mathbf{p}, e(\mathbf{p}, \bar{u})) + \frac{\partial v}{\partial M}(\mathbf{p}, e(\mathbf{p}, \bar{u}))\,\mathbf{h}^n(\mathbf{p}, \bar{u}) = 0 \tag{4.6.5}$$

Hence

$$\mathbf{h}^n(\mathbf{p}, \bar{u}) = -\frac{\frac{\partial v}{\partial p^n}(\mathbf{p}, e(\mathbf{p}, \bar{u}))}{\frac{\partial v}{\partial M}(\mathbf{p}, e(\mathbf{p}, \bar{u}))} \tag{4.6.6}$$

and expressing this last equation in terms of the relevant level of income M rather than the corresponding value of utility $\bar{u}$:

$$\mathbf{x}^n(\mathbf{p}, M) = -\frac{\frac{\partial v}{\partial p^n}(\mathbf{p}, M)}{\frac{\partial v}{\partial M}(\mathbf{p}, M)}. \tag{4.6.7}$$

Q.E.D.

Theorem 4.6.3 (Slutsky symmetry condition.) *All cross-price substitution effects are symmetric:*

$$\frac{\partial h_h^n}{\partial p^m} = \frac{\partial h_h^m}{\partial p^n}. \tag{4.6.8}$$

Proof From Shephard's Lemma, we can easily derive the *Slutsky symmetry conditions*, assuming that the expenditure function is twice continuously differentiable, and hence that

$$\frac{\partial^2 e_h}{\partial p^m \partial p^n} = \frac{\partial^2 e_h}{\partial p^n \partial p^m}. \tag{4.6.9}$$

Since $h_h^m = \frac{\partial e_h}{\partial p^m}$ and $h_h^n = \frac{\partial e_h}{\partial p^n}$, and the result follows.

Q.E.D.

The next result doesn't really have a special name of its own.

Theorem 4.6.4 *Since the expenditure function is concave in prices (see p. 75), the corresponding Hessian matrix is negative semi-definite. In particular, its diagonal entries are non-positive, or*

$$\frac{\partial^2 e_h}{\partial (p^n)^2} \leq 0, \quad n = 1, \ldots, N. \tag{4.6.10}$$

Using Shephard's Lemma, it follows that

$$\frac{\partial h_h^n}{\partial p^n} \leq 0, \quad n = 1, \ldots, N. \tag{4.6.11}$$

In other words, Hicksian demand functions, unlike Marshallian demand functions, are uniformly decreasing in own price. Another way of saying this is that own price substitution effects are always negative.

Theorem 4.6.5 (Slutsky equation.) *The total effect of a price change on (Marshallian) demand can be decomposed as follows into a substitution effect and an income effect:*

$$\frac{\partial x^m}{\partial p^n}(\mathbf{p}, M) = \frac{\partial h^m}{\partial p^n}(\mathbf{p}, \bar{u}) - \frac{\partial x^m}{\partial M}(\mathbf{p}, M)\, \mathbf{h}^n(\mathbf{p}, \bar{u}), \tag{4.6.12}$$

where $\bar{u} \equiv V(\mathbf{p}, M)$.

Before proving this, let's consider the signs of the various terms in the Slutsky equation and look at what it means in a two-good example.

By Theorem 4.6.4, we know that own price substitution effects are always non-positive.

[This is still on a handwritten sheet.]

Proof Differentiating both sides of the lth component of (4.4.9) with respect to p^n, using the Chain Rule, will yield the so-called *Slutsky equation* which decomposes the total effect on demand of a price change into an income effect and a substitution effect.
Differentiating the RHS of (4.4.9) with respect to p^n yields:

$$\frac{\partial x^m}{\partial p^n}\left(\mathbf{p}, e\left(\mathbf{p}, \bar{u}\right)\right) + \frac{\partial x^m}{\partial M}\left(\mathbf{p}, e\left(\mathbf{p}, \bar{u}\right)\right)\frac{\partial e}{\partial p^n}\left(\mathbf{p}, \bar{u}\right). \qquad (4.6.13)$$

To complete the proof:

1. set this equal to $\frac{\partial h^m}{\partial p^n}\left(\mathbf{p}, \bar{u}\right)$
2. substitute from Shephard's Lemma
3. define $M \equiv e\left(\mathbf{p}, \bar{u}\right)$ (which implies that $\bar{u} \equiv V\left(\mathbf{p}, M\right)$)

Q.E.D.

4.7 General Equilibrium Theory

4.7.1 Walras' law

Walras ...[3]

4.7.2 Brouwer's fixed point theorem

4.7.3 Existence of equilibrium

4.8 The Welfare Theorems

4.8.1 The Edgeworth box

4.8.2 Pareto efficiency

Definition 4.8.1 *A feasible allocation* $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_H)$ *is* Pareto efficient *if there does not exist any* **feasible** *way of reallocating the same initial aggregate endowment,* $\sum_{h=1}^{H} \mathbf{x}_h$*, which makes one individual better off without making any other worse off.*

Definition 4.8.2 $\mathbf{X}$ *is* Pareto dominated by $\mathbf{X}' = (\mathbf{x}'_1, \ldots, \mathbf{x}'_H)$ *if* $\sum_{h=1}^{H} \mathbf{x}_h = \sum_{h=1}^{H} \mathbf{x}'_h$, $\mathbf{x}'_h \succeq_h \mathbf{x}_h \; \forall h$ *and* $\mathbf{x}'_h \succ_h \mathbf{x}_h$ *for at least one* h.

[3] This material still exists only in handwritten form in Alan White's EC3080 notes from 1991-2. One thing missing from the handwritten notes is Kakutani's Fixed Point Theorem which should be quoted from **?**.

4.8.3 The First Welfare Theorem

(See **?**.)

Theorem 4.8.1 (First Welfare Theorem) *If the pair $(\mathbf{p}, \mathbf{X})$ is an equilibrium (for given preferences, $\succeq_h$, which exibit local non-satiation and given endowments, $\mathbf{e}_h$, $h = 1, \ldots, H$), then $\mathbf{X}$ is a Pareto efficient allocation.*

Proof The proof is by contradiction. Suppose that $\mathbf{X}$ is an equilibrium allocation which is Pareto dominated by a feasible allocation $\mathbf{X}'$.
If individual h is strictly better off under $\mathbf{X}'$ or $\mathbf{x}'_h \succ_h \mathbf{x}_h$, then it follows that individual h cannot afford $\mathbf{x}'_h$ at the equilibrium prices $\mathbf{p}$ or

$$\mathbf{p}^\top \mathbf{x}'_h > \mathbf{p}^\top \mathbf{x}_h = \mathbf{p}^\top \mathbf{e}_h. \tag{4.8.1}$$

The latter equality is just the budget constraint, which is binding since we have assumed local non-satiation.
Similarly, if individual h is indifferent between $\mathbf{X}$ and $\mathbf{X}'$ or $\mathbf{x}'_h \sim_h \mathbf{x}_h$, then it follows that

$$\mathbf{p}^\top \mathbf{x}'_h \geq \mathbf{p}^\top \mathbf{x}_h = \mathbf{p}^\top \mathbf{e}_h, \tag{4.8.2}$$

since if $\mathbf{x}'_h$ cost strictly less than $\mathbf{x}_h$, then by local non-satiation some consumption vector near enough to $\mathbf{x}'_h$ to also cost less than $\mathbf{x}_h$ would be strictly preferred to $\mathbf{x}_h$ and $\mathbf{x}_h$ would not maximise utility given the budget constraint.
Summing (4.8.1) and (4.8.2) over households yields

$$\mathbf{p}^\top \sum_{h=1}^{H} \mathbf{x}'_h > \mathbf{p}^\top \sum_{h=1}^{H} \mathbf{x}_h = \mathbf{p}^\top \sum_{h=1}^{H} \mathbf{e}_h, \tag{4.8.3}$$

(where the equality is essentially Walras' Law).
But since $\mathbf{X}'$ is feasible we must have for each good n

$$\sum_{h=1}^{H} x'^n_h \leq \sum_{h=1}^{H} e^n_h$$

and, hence, multiplying by prices and summing over all goods,

$$\mathbf{p}^\top \sum_{h=1}^{H} \mathbf{x}'_h \leq \mathbf{p}^\top \sum_{h=1}^{H} \mathbf{e}_h. \tag{4.8.4}$$

But (4.8.4) contradicts the inequality in (4.8.3), so no such Pareto dominant allocation $\mathbf{X}'_h$ can exist.

Q.E.D.

Before proceeding to the second welfare theorem, we need to say a little bit about separating hyperplanes.

4.8.4 The Separating Hyperplane Theorem

Definition 4.8.3 *The set* $\left\{\mathbf{z} \in \Re^N : \mathbf{p}^\top \mathbf{z} = \mathbf{p}^\top \mathbf{z}^*\right\}$ *is the* hyperplane *through* $\mathbf{z}^*$ *with normal* $\mathbf{p}$.

Note that any hyperplane divides $\Re^N$ into two closed half-spaces,

$$\left\{\mathbf{z} \in \Re^N : \mathbf{p}^\top \mathbf{z} \leq \mathbf{p}^\top \mathbf{z}^*\right\}$$

and

$$\left\{\mathbf{z} \in \Re^N : \mathbf{p}^\top \mathbf{z} \geq \mathbf{p}^\top \mathbf{z}^*\right\}.$$

The intersection of these two closed half-spaces is the hyperplane itself.
In two dimensions, a hyperplane is just a line; in three dimensions, it is just a plane.
The idea behind the separating hyperplane theorem is quite intuitive: if we take any point on the boundary of a convex set, we can find a hyperplane through that point so that the entire convex set lies on one side of that hyperplane. We will essentially be applying this notion to the upper contour sets of quasiconcave utility functions, which are of course convex sets. We will interpret the separating hyperplane as a budget hyperplane, and the normal vector as a price vector, so that at those prices nothing giving higher utility than the cutoff value is affordable.

Theorem 4.8.2 (Separating Hyperplane Theorem) *If Z is a convex subset of $\Re^N$ and $\mathbf{z}^* \in Z, \mathbf{z}^* \notin \text{int } Z$, then $\exists \mathbf{p}^* \neq \mathbf{0}$ in $\Re^N$ such that $\mathbf{p}^{*\top}\mathbf{z}^* \leq \mathbf{p}^{*\top}\mathbf{z}\ \forall \mathbf{z} \in Z$, or Z is contained in one of the closed half-spaces associated with the hyperplane through $\mathbf{z}^*$ with normal $\mathbf{p}^*$.*

Proof Not given. See **?**

Q.E.D.

4.8.5 The Second Welfare Theorem

(See **?**.)
We make slightly stronger assumptions than are essential for the proof of this theorem. This allows us to give an easier proof.

Theorem 4.8.3 (Second Welfare Theorem) *If all individual preferences are strictly convex, continuous and strictly monotonic, and if $\mathbf{X}^*$ is a Pareto efficient allocation such that all households are allocated positive amounts of all goods ($x_h^{*g} > 0\ \forall g = 1, \ldots, N; h = 1, \ldots, H$), then a reallocation of the initial aggregate endowment can yield an equilibrium where the allocation is $\mathbf{X}^*$.*

Proof There are four main steps in the proof.

1. First we construct a set of utility-enhancing *endowment perturbations*, and use the separating hyperplane theorem to find prices at which no such endowment perturbation is affordable.

   We need to use the fact (Theorem 3.2.1) that a sum of convex sets, such as

   $$X + Y \equiv \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X, \mathbf{y} \in Y\},$$

   is also a convex set.

   Given an aggregate initial endowment $\mathbf{x}^* = \sum_{h=1}^{H} \mathbf{x}_h^*$, we interpret any vector of the form $\mathbf{z} = \sum_{h=1}^{H} \mathbf{x}_h - \mathbf{x}^*$ as an endowment perturbation. Now consider the set of all ways of changing the aggregate endowment without making anyone worse off:

   $$Z \equiv \left\{\mathbf{z} \in \Re^N : \exists x_h^n \geq 0 \; \forall g, h \; s.t. \; u_h(\mathbf{x}_h) \geq u_h(\mathbf{x}_h^*) \; \& \; \mathbf{z} = \sum_{h=1}^{H} \mathbf{x}_h - \mathbf{x}^*\right\}. \tag{4.8.5}$$

   Z is a sum of convex sets provided that preferences are assumed to be convex:

   $$Z = \sum_{h=1}^{H} X_h - \{\mathbf{x}^*\}$$

   where

   $$X_h \equiv \{\mathbf{x}_h : u_h(\mathbf{x}_h) \geq u_h(\mathbf{x}_h^*)\}.$$

2. Next, we need to show that the zero vector is in the set Z, but not in the interior of Z.

   To show that $\mathbf{0} \in Z$, we just set $\mathbf{x}_h = \mathbf{x}_h^*$ and observe that $\mathbf{0} = \sum_{h=1}^{H} \mathbf{x}_h^* - \mathbf{x}^*$.[4]

   The zero vector is not, however, in the interior of Z, since then Z would contain some vector, say $\mathbf{z}^*$, in which all components were strictly negative. In other words, we could take away some of the aggregate endowment of every good without making anyone worse off than under the allocation $\mathbf{X}^*$. But by then giving $-\mathbf{z}^*$ back to one individual, he or she could be made better off without making anyone else worse off, contradicting Pareto optimality, again using the assumption that preferences are strictly monotonic.

[4]Note also (although I'm no longer sure why this is important) that budget constraints are binding and that there are no free goods (by the monotonicity assumption).

So, applying the Separating Hyperplane Theorem with $\mathbf{z}^* = \mathbf{0}$, we have a price vector $\mathbf{p}^*$ such that $0 = \mathbf{p}^{*\top}\mathbf{0} \leq \mathbf{p}^{*\top}\mathbf{z}\ \forall \mathbf{z} \in Z$.

Since preferences are monotonic, the set Z must contain all the standard unit basis vectors ($(1, 0, \ldots, 0)$, *&c.*). This fact can be used to show that all components of $\mathbf{p}^*$ are non-negative, which is essential if it is to be interpreted as an equilibrium price vector.

3. Next, we specify one way of redistributing the initial endowment in order that the desired prices and allocation emerge as a competitive equilibrium. All we need to do is value endowments at the equilibrium prices, and redistribute the aggregate endowment of each good to consumers in proportion to their share in aggregate wealth computed in this way.

4. Finally, we confirm that utility is maximised by the given Pareto efficient allocation, $\mathbf{X}^*$, at these prices. As usual, the proof is by contradiction: the details are left as an exercise.

Q.E.D.

4.8.6 Complete markets

The First Welfare Theorem tells us that competitive equilibrium allocations are Pareto optimal if markets are complete. If there are missing markets, then competitive trading may not lead to a Pareto optimal allocation. We can use the Edgeworth Box diagram to illustrate the simplest possible version of this principle.

4.8.7 Other characterizations of Pareto efficient allocations

There are a total of five equivalent characterisations of Pareto efficient allocations.

Theorem 4.8.4 *Each of the following is an equivalent description of the set of allocations which are Pareto efficient:*

1. *by definition, feasible allocations such that no other allocation strictly increases at least one individual's utility without decreasing the utility of any other individual;*

2. *by the Welfare Theorems, equilibrium allocations for all possible distributions of the fixed initial aggregate endowment;*

3. *in two dimensions, allocations lying on the contract curve in the Edgeworth box;*

4. *allocations which solve:* [5]

$$\max_{\{\mathbf{x}_h : h=1,\ldots,H\}} \sum_{h=1}^{H} \lambda_h \left[u_h(\mathbf{x}_h) \right] \tag{4.8.6}$$

subject to the feasibility constraints

$$\sum_{h=1}^{H} \mathbf{x}_h = \sum_{h=1}^{H} \mathbf{e}_h \tag{4.8.7}$$

for some non-negative weights $\{\lambda_h\}_{h=1}^{H}$.

5. *allocations which maximise the utility of a* representative agent *given by*

$$\sum_{h=1}^{H} \lambda_h \left[u_h(\mathbf{x}_h) \right] \tag{4.8.8}$$

where $\{\lambda_h\}_{h=1}^{H}$ *are again any non-negative weights.*

Proof If an allocation is not Pareto efficient, then the Pareto-dominating allocation gives a higher value of the objective function in the above problem for all possible weights.
If an allocation is Pareto efficient, then the relative weights for which the above objective function is maximized are the ratios of the Lagrange multipliers from the problem of maximizing any individual's utility subject to the constraint that all other individuals' utilities are unchanged:

$$\max u_1(\mathbf{x}_1) \tag{4.8.9}$$

s.t.

$$u_h(\mathbf{x}_h) = u_h(\mathbf{x}_h^*) \; h = 2, \ldots, H \tag{4.8.10}$$

since these two problems will have the same necessary and sufficient first order conditions.
The absolute weights corresponding to a particular allocation are not unique, as they can be multiplied by any positive constant without affecting the maximum.
Different absolute weights (or Lagrange multipliers) arise from fixing different individuals' utilities in the last problem, but the relative weights will be the same.

[5] The solution here would be unique if the underlying utility function were concave, since linear combinations of concave functions with non-negative weights are concave, and the constraints specify a convex set on which the objective function has a unique optimum. This argument can not be used with merely quasiconcave utility functions.

Q.E.D.

Note that corresponding to each Pareto efficient allocation there is at least one:

1. set of non-negative weights defining
   (a) the objective function in 4. and
   (b) the representative agent in 5.

   and
2. initial allocation leading to the competitive equilibrium in 2.

4.9 Multi-period General Equilibrium

In Section 4.2, it was pointed out that the objects of choice can be differentiated not only by their physical characteristics, but also both by the time at which they are consumed and by the state of nature in which they are consumed. These distinctions were suppressed in the intervening sections but are considered again in this section and in Section 5.4 respectively.

The multi-period model should probably be introduced at the end of Chapter 4 but could also be left until Chapter 7. For the moment this brief introduction is duplicated in both chapters.

Discrete time multi-period investment problems serve as a stepping stone from the single period case to the continuous time case.

The main point to be gotten across is the derivation of interest rates from equilibrium prices: spot rates, forward rates, term structure, etc.

This is covered in one of the problems, which illustrates the link between prices and interest rates in a multiperiod model.

Chapter 5

CHOICE UNDER UNCERTAINTY

5.1 Introduction

[To be written.]

5.2 Review of Basic Probability

Economic theory has, over the years, used many different, sometimes overlapping, sometimes mutually exclusive, approaches to the analysis of choice under uncertainty. This chapter deals with choice under uncertainty exclusively in a single period context. Trade takes place at the beginning of the period and uncertainty is resolved at the end of the period. This framework is sufficient to illustrate the similarities and differences between the most popular approaches.

When we consider consumer choice under uncertainty, consumption plans will have to specify a fixed consumption vector for each possible *state of nature* or *state of the world*. This just means that each consumption plan is a *random vector*. Let us review the associated concepts from basic probability theory: probability space; random variables and vectors; and stochastic processes.

Let Ω denote the set of all possible states of the world, called the *sample space*.

A collection of states of the world, $A \subseteq \Omega$, is called an *event*.

Let $\mathcal{A}$ be a collection of events in Ω. The function $P : \mathcal{A} \to [0, 1]$ is a *probability function* if

1. (a) $\Omega \in \mathcal{A}$

   (b) $A \in \mathcal{A} \Rightarrow \Omega - A \in \mathcal{A}$

   (c) $A_i \in \mathcal{A}$ for $i = 1, \ldots, \infty \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

   (*i.e.* $\mathcal{A}$ is a *sigma-algebra* of events) and

2. (a) $P(\Omega) = 1$
   (b) $P(\Omega - A) = 1 - P(A)\,\forall A \in \mathcal{A}$ (redundant assumption)
   (c) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ when $A_1, A_2, \ldots$ are pairwise disjoint events in $\mathcal{A}$.

$(\Omega, \mathcal{A}, P)$ is then called a *probability space*
Note that the certainty case we considered already is just the special case of uncertainty in which the set Ω has only one element.
We will consider these concepts in more detail when we come to intertemporal models.
Suppose we are given such a probability space.
The function $\tilde{x} : \Omega \to \Re$ is a *random variable* (r.v.) if $\forall x \in \Re\ \{\omega \in \Omega : \tilde{x}(\omega) \leq x\} \in \mathcal{A}$, *i.e.* a function is a random variable if we know the probability that the value of the function is less than or equals any given real number.
The function $F_{\tilde{x}} : \Re \to [0,1] : x \mapsto \Pr(\tilde{x} \leq x) \equiv P(\{\omega \in \Omega : \tilde{x}(\omega) \leq x\})$ is known as the *cumulative distribution function* (c.d.f.) of the random variable $\tilde{x}$.
The convention of using a tilde over a letter to denote a random variable is common in financial economics; in other fields capital letters may be reserved for random variables. In either case, small letters usually denote particular real numbers (*i.e.* particular values of the random variable).
A *random vector* is just a vector of random variables. It can also be thought of as a vector-valued function on the sample space Ω.
A *stochastic process* is a collection of random variables or random vectors indexed by time, *e.g.* $\{\tilde{x}_t : t \in T\}$ or just $\{\tilde{x}_t\}$ if the time interval is clear from the context.
For the purposes of this part of the course, we will assume that the index set consists of just a finite number of times *i.e.* that we are dealing with discrete time stochastic processes.
Then a stochastic process whose elements are N-dimensional random vectors is equivalent to an $N|T|$-dimensional random vector.
The (joint) c.d.f. of a random vector or stochastic process is the natural extension of the one-dimensional concept.
Random variables can be *discrete*, *continuous* or *mixed*. The *expectation* (mean, average) of a discrete r.v., $\tilde{x}$, with possible values $x_1, x_2, x_3, \ldots$ is given by

$$E[\tilde{x}] \equiv \sum_{i=1}^{\infty} x_i Pr(\tilde{x} = x_i).$$

For a continuous random variable, the summation is replaced by an integral.

The *covariance* of two random variables $\tilde{x}$ and $\tilde{y}$ is given by

$$\mathrm{Cov}\left[\tilde{x}, \tilde{y}\right] \equiv E\left[\left(\tilde{x} - e\left[\tilde{x}\right]\right)\left(\tilde{y} - E\left[\tilde{y}\right]\right)\right].$$

The covariance of a random variable with itself is called its *variance*.
The expectation of a random vector is just the vector of the expectations of the component random variables.
The variance (*variance-covariance matrix*) of a random vector is the (symmetric, positive semi-definite) matrix of the covariances between the component random variables.
Given any two random variables $\tilde{x}$ and $\tilde{y}$, we can define a third random variable $\tilde{\epsilon}$ by

$$\tilde{\epsilon} \equiv \tilde{y} - \alpha - \beta\tilde{x}. \tag{5.2.1}$$

To specify $\tilde{\epsilon}$ completely, we can either specify α and β explicitly or fix them implicitly by imposing (two) conditions on $\tilde{\epsilon}$. We do the latter by insisting

1. $\tilde{\epsilon}$ and $\tilde{x}$ are *uncorrelated* (this is not the same as assuming statistical independence, except in special cases such as bivariate normality)

2. $E\left[\tilde{\epsilon}\right] = 0$

It follows that:

$$\beta = \frac{\mathrm{Cov}\left[\tilde{x}, \tilde{y}\right]}{\mathrm{Var}\left[\tilde{x}\right]} \tag{5.2.2}$$

and

$$\alpha = E\left[\tilde{y}\right] - \beta E\left[\tilde{x}\right]. \tag{5.2.3}$$

But what about the conditional expectation, $E\left[\tilde{y}|\tilde{x} = x\right]$? This is not equal to $\alpha + \beta x$, as one might expect, unless $E\left[\tilde{\epsilon}|\tilde{x} = x\right] = 0$. This requires statistical independence rather than the assumed lack of correlation. Again, a sufficient condition is multivariate normality.
The notion of the β of $\tilde{y}$ with respect to $\tilde{x}$ as given in (5.2.2) will recur frequently.
The final concept required from basic probability theory is the notion of a mixture of random variables. For lotteries which are discrete random variables, with payoffs $x_1, x_2, x_3, \ldots$ occuring with probabilities $\pi_1, \pi_2, \pi_3, \ldots$ respectively, we will use the notation:

$$\pi_1 x_1 \oplus \pi_2 x_2 \oplus \pi_3 x_3 \oplus \ldots$$

Similar notation will be used for compound lotteries (mixtures of random variables) where the payoffs themselves are further lotteries.
This might be a good place to talk about the MVN distribution and Stein's lemma.

5.3 Taylor's Theorem: Stochastic Version

We will frequently use the univariate Taylor expansion as applied to a function of a random variable expanded about the mean of the random variable.[1] Taking expectations on both sides of the Taylor expansion:

$$f(\tilde{x}) = f(E[\tilde{x}]) + \sum_{n=1}^{\infty} \frac{1}{n!} f^{(n)}(E[\tilde{x}])(\tilde{x} - E[\tilde{x}])^n \quad (5.3.1)$$

yields:

$$E[f(\tilde{x})] = f(E[\tilde{x}]) + \sum_{n=2}^{\infty} \frac{1}{n!} f^{(n)}(E[\tilde{x}]) m^n(\tilde{x}), \quad (5.3.2)$$

where

$$m^n(\tilde{x}) \equiv E\left[(\tilde{x} - E[\tilde{x}])^n\right]. \quad (5.3.3)$$

In particular,

$$m^1(\tilde{x}) = E\left[(\tilde{x} - E[\tilde{x}])^1\right] \equiv 0 \quad (5.3.4)$$

$$m^2(\tilde{x}) = E\left[(\tilde{x} - E[\tilde{x}])^2\right] \equiv \text{Var}[\tilde{x}] \quad (5.3.5)$$

$$m^3(\tilde{x}) = E\left[(\tilde{x} - E[\tilde{x}])^3\right] \equiv \text{Skew}[\tilde{x}] \quad (5.3.6)$$

and

$$m^4(\tilde{x}) = E\left[(\tilde{x} - E[\tilde{x}])^4\right] \equiv \text{Kurt}[\tilde{x}], \quad (5.3.7)$$

which allows us to start the summation in (5.3.2) at $n = 2$ rather than $n = 1$. Indeed, we can rewrite (5.3.2) as

$$E[f(\tilde{x})] = f(E[\tilde{x}]) + \frac{1}{2} f''(E[\tilde{x}])\text{Var}[\tilde{x}] + \frac{1}{6} f'''(E[\tilde{x}])\text{Skew}[\tilde{x}] + \frac{1}{24} f''''(E[\tilde{x}])\text{Kurt}[\tilde{x}] + \sum_{n=5}^{\infty} \frac{1}{n!} f^{(n)}(E[\tilde{x}]) m^n(\tilde{x}). \quad (5.3.8)$$

5.4 Pricing State-Contingent Claims

This part of the course draws on **?**, **?** and **?**.

The analysis of choice under uncertaintly will begin by reinterpreting the general equilibrium model of Chapter 4 so that goods can be differentiated by the state of nature in which they are consumed. Specifically, it will be assumed that the

[1] This section will eventually have to talk separately about kth order and infinite order Taylor expansions.

underlying sample space comprises a finite number of states of nature. A more thorough analysis of choice under uncertainty, allowing for infinite and continuous sample spaces and based on additional axioms of choice, follows in Section 5.5.
Consider a world with M possible states of nature (distinguished by a first subscript), markets for N securities (distinguished by a second subscript) and H consumers (distinguished by a superscript).[2]

Definition 5.4.1 *A* state contingent claim *or* Arrow-Debreu security *is a random variable or lottery which takes the value 1 in one particular state of nature and the value 0 in all other states.*

Definition 5.4.2 *A* complex security *is a random variable or lottery which can take on arbitrary values. The payoffs of a typical complex security will be represented by a column vector,* $\mathbf{y}_j \in \Re^M$*, where* y_{ij} *is the payoff in state* i *of security* j.

The set of all complex securities on a given finite sample space is an M-dimensional vector space and the M possible Arrow-Debreu securities constitute the standard basis for this vector space.
State contingent claims prices are determined by the market clearing equations in a general equilibrium model:

$$\text{Aggregate consumption in state } i = \text{Aggregate endowment in state } i.$$

Each individual will have an optimal consumption choice depending on endowments and preferences and conditional on the state of the world. Optimal future consumption is denoted

$$\mathbf{x}^* = \begin{pmatrix} x_1^* \\ x_2^* \\ \ldots \\ x_N^* \end{pmatrix}. \tag{5.4.1}$$

If there are N complex securities, then the investor must find a portfolio $\mathbf{w} = (w_1, \ldots, w_N)$ whose payoffs satisfy

$$x_i^* = \sum_{j=1}^{N} y_{ij} w_j.$$

Let $\mathbf{Y}$ be the $M \times N$ matrix[3] whose jth column contains the payoffs of the jth complex security in each of the M states of nature, *i.e.*

$$\mathbf{Y} \equiv (\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_N). \tag{5.4.2}$$

[2]Check for consistency in subscripting etc in what follows.
[3]Or maybe I mean its transpose.

Theorem 5.4.1 *If there are M complex securities ($M = N$) and the payoff matrix $\mathbf{Y}$ is non-singular, then markets are complete.*

Proof Suppose the optimal trade for consumer i state j is $x_{ij} - e_{ij}$. Then can invert $\mathbf{Y}$ to work out optimal trades in terms of complex securities.

Q.E.D.

An $(N + 1)$st security would be redundant.
Either a singular square matrix or $< N$ complex securities would lead to incomplete markets.
So far, we have made no assumptions about the form of the utility function, written purely as

$$u\left(x_0, x_1, x_2, \ldots, x_N\right),$$

where x_0 represents the quantity consumed at date 0 and x_i ($i > 0$) represents the quantity consumed at date 1 if state i materialises.

5.4.1 Completion of markets using options

Assume that there exists a *state index portfolio*, Y, yielding **different** non-zero payoffs in each state (*i.e.* a portfolio with a different payout in each state of nature, possibly one mimicking aggregate consumption). WLOG we can rank the states so that $Y_i < Y_j$ if $i < j$.
We now present some results, following **?**, showing conditions under which trading in a state index portfolio and in options on the state index portfolio can lead to the Pareto optimal complete markets equilibrium allocation. Now consider completion of markets using options on aggregate consumption.
In real-world markets, the number of linearly independent corporate securities is probably less than M.
However, options on corporate securities may be sufficient to form complete markets, and thereby ensure allocational (Pareto) efficiency for arbitrary preferences.
Further assume that $\exists\ M - 1$ European call options on Y with exercise prices $Y_1, Y_2, \ldots, Y_{M-1}$.
A **European call option** with exercise price K is an option to buy a security for K on a fixed date.
An **American call option** is an option to buy **on or before** the fixed date.
A **put option** is an option to sell.

Here, the original state index portfolio and the $M-1$ European call options yield the payoff matrix:

$$\begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_M \\ 0 & y_2-y_1 & y_3-y_1 & \cdots & y_M-y_1 \\ 0 & 0 & y_3-y_2 & \cdots & y_M-y_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_M-y_{M-1} \end{pmatrix} = \begin{pmatrix} \text{security } Y \\ \text{call option 1} \\ \text{call option 2} \\ \vdots \\ \text{call option } M-1 \end{pmatrix} \quad (5.4.3)$$

and as this matrix is non-singular, we have constructed a complete market. Instead of assuming a state index portfolio exists, we can assume identical probability beliefs and state-independent utility and complete markets in a similar manner (see below).

5.4.2 Restrictions on security values implied by allocational efficiency and covariance with aggregate consumption

Let $\begin{cases} C\omega = & \text{aggregate consumption in state } \omega \\ \Omega_k = & \{\omega \in \Omega : C\omega = k\} \end{cases}$

Let $\phi(k)$ be the value of the claim with payoffs

$$y_{k\omega} = \begin{cases} 1 & \text{if } C\omega = k \\ 0 & \text{otherwise} \end{cases} \quad (5.4.4)$$

and let the agreed probability of the event Ω_k (*i.e.* of aggregate consumption taking the value k) be:

$$\pi(k) = \sum_{\omega \in \Omega_k} \pi\omega \quad (5.4.5)$$

By time-additivity and state-independence of the utility function:

$$\phi\omega = \frac{\pi\omega u_i'(c_{i\omega})}{u_{i0}'(c_{i0})} \quad \forall \omega \in \Omega \quad (5.4.6)$$

The no arbitrage condition implies

$$\begin{aligned} \phi(k) &= \sum_{\omega \in \Omega_k} \phi\omega && (5.4.7) \\ &= \frac{u_i'(f_i(k))}{u_{i0}'(c_{i0})} \sum_{\omega \in \Omega_k} \pi\omega && (5.4.8) \\ &= \frac{u_i'(f_i(k))}{u_{i0}'(c_{i0})} \pi(k) && (5.4.9) \end{aligned}$$

where $f_i(k)$ denotes the i-th individual's equilibrium consumption in those states where aggregate consumption equals k.

	$x(0)$	$x(1)$	$x(2)$
$\tilde{C}=1$	1	0	0
$\tilde{C}=2$	2	1	0
$\tilde{C}=3$	3	2	1
.	.	.	.
.	.	.	.
.	.	.	.
$\tilde{C}=L$	L	$L-1$	$L-2$

Table 5.1: Payoffs for Call Options on the Aggregate Consumption

State-independence of the utility function is required for $f_i(k)$ to be well-defined. Therefore, an arbitrary security x has value:

$$S_x = \sum_{\omega\in\Omega} \phi\omega x\omega \quad (5.4.10)$$

$$= \sum_k \sum_{\omega\in\Omega_k} \phi\omega x\omega \quad (5.4.11)$$

$$= \sum_k \frac{u_i'(f_i(k))}{u_{i0}'(c_{i0})} \sum_{\omega\in\Omega_k} \pi\omega x\omega \quad (5.4.12)$$

$$= \sum_k \phi(k) \sum_{\omega\in\Omega_k} \frac{\pi\omega}{\pi(k)} x\omega \quad (5.4.13)$$

$$= \sum_k \phi(k) E[\tilde{x}|\tilde{C}=k] \quad (5.4.14)$$

5.4.3 Completing markets with options on aggregate consumption

Let $x(k)$ be the vector of payoffs in the various possible states on a European call option on aggregate consumption with one period to maturity and exercise price k.
Let $\{1, 2, \ldots, L\}$ be the set of possible values of aggregate consumption $C(\omega)$.
Then payoffs are as given in Table 5.1.
This all assumes

1. identical probability beliefs
2. time-additivity of u
3. state-independent u

5.4.4 Replicating elementary claims with a butterfly spread

Elementary claims against aggregate consumption can be constructed as follows, for example, for state 1, using a **butterfly spread**:

$$[x(0) - x(1)] - [x(1) - x(2)] \tag{5.4.15}$$

yields the payoff:

$$\begin{aligned}\left[\begin{pmatrix}1\\2\\3\\\vdots\\L\end{pmatrix} - \begin{pmatrix}0\\1\\2\\\vdots\\L-1\end{pmatrix}\right] - \left[\begin{pmatrix}0\\1\\2\\\vdots\\L-1\end{pmatrix} - \begin{pmatrix}0\\0\\1\\\vdots\\L-2\end{pmatrix}\right] &= \begin{pmatrix}1\\1\\1\\\vdots\\1\end{pmatrix} - \begin{pmatrix}0\\1\\1\\\vdots\\1\end{pmatrix}\\ &= \begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix}\end{aligned} \tag{5.4.16}$$

i.e. this replicating portfolio pays 1 iff aggregate consumption is 1, and 0 otherwise.
The prices of this, and the other elementary claims, must, by no arbitrage, equal the prices of the corresponding replicating portfolios.

5.5 The Expected Utility Paradigm

5.5.1 Further axioms

The objects of choice with which we are concerned in a world with uncertainty could still be called *consumption plans*, but we will acknowledge the additional structure now described by terming them *lotteries*.
If there are k physical commodities, a consumption plan must specify a k-dimensional vector, $\mathbf{x} \in \Re^k$, for each time and state of the world.
We assume a finite number of times, denoted by the set T.
The possible states of the world are denoted by the set Ω.
So a consumption plan or lottery is just a collection of $|T|$ k-dimensional random vectors, *i.e.* a stochastic process.
Again to distinguish the certainty and uncertainty cases, we let $\mathcal{L}$ denote the collection of lotteries under consideration; $\mathcal{X}$ will now denote the set of possible values of the lotteries in $\mathcal{L}$.

Preferences are now described by a relation on $\mathcal{L}$. We will continue to assume that preference relations are complete, reflexive, transitive, and continuous.
$\mathcal{X}$ can be identified with a subset of $\mathcal{L}$, in that each *sure thing* in $\mathcal{X}$ can be identified with the trivial lottery that pays off that sure thing with probability (w.p.) 1.
Although we have moved from a finite-dimensional to an infinite-dimensional problem by explicitly allowing a continuum of states of nature, it can be shown that the earlier theory of choice under certainty carries through to choice under uncertainty, in particular a preference relation can always be represented by a continuous utility function on $\mathcal{L}$.
However, we would like utility functions to have a stronger property than continuity, namely the expected utility property.

Axiom 8 (Substitution or Independence Axiom) *If $a \in (0,1]$ and $\tilde{p} \succ \tilde{q}$, then*

$$a\tilde{p} \oplus (1-a)\,\tilde{r} \succ a\tilde{q} \oplus (1-a)\,\tilde{r}\;\forall \tilde{r} \in \mathcal{L}.$$

Axiom 9 (Archimedian Axiom) *If $\tilde{p} \succ \tilde{q} \succ \tilde{r}$ then*

$$\exists a, b \in (0,1) \text{ s.t. } a\tilde{p} \oplus (1-a)\,\tilde{r} \succ \tilde{q} \succ b\tilde{p} \oplus (1-b)\,\tilde{r}.$$

(The Archimedian axiom is just a generalisation of the continuity axiom.)

Axiom 10 (Sure Thing Principle) *If probability is concentrated on a set of sure things which are preferred to q, then the associated consumption plan is also preferred to q.*

(The Sure Thing Principle is just a generalisation of the Substitution Axiom.)
Now let us consider the Allais paradox.
Suppose

$$1\pounds 1m. \succ 0.1\pounds 5m. \oplus 0.89\pounds 1m. \oplus 0.01\pounds 0.$$

Then, unless the substitution axiom is contradicted:

$$1\pounds 1m. \succ \frac{10}{11}\pounds 5m. \oplus \frac{1}{11}\pounds 0.$$

Finally, by the substitution axiom again,

$$0.11\pounds 1m. \oplus 0.89\pounds 0 \succ 0.1\pounds 5m. \oplus 0.9\pounds 0.$$

If these appears counterintuitive, then so does the independence axiom above.
One justification for persisting with the independence axiom is provided by **?**.

5.5.2 Existence of expected utility functions

A function $u: \Re^{k\times|T|} \rightarrow \Re$ can be thought of as a *utility function on sure things*.

Definition 5.5.1 *Let $V : \mathcal{L} \rightarrow \Re$ be a utility function representing the preference relation $\succeq$.*
Then $\succeq$ is said to have an expected utility representation *if there exists a utility function on sure things, u, such that*

$$\begin{aligned} V(\{\tilde{x}_t\}) &= E\left[u(\{\tilde{x}_t\})\right] \\ &= \int \ldots \int u(\{x_t\})\, dF_{\{\tilde{x}_t\}}(\{x_t\}) \end{aligned}$$

Such a representation will often be called a Von Neumann-Morgenstern (or VNM) utility function, *after its originators (**?**), or just an* expected utility function.

Any strictly increasing transformation of a VNM utility function represents the same preferences.
However, only increasing affine transformations:

$$f(x) = a + bx \ (b > 0)$$

retain the expected utility property.
Proof of this is left as an exercise.
We will now consider necessary and sufficient conditions on preference relations for an expected utility representation to exist.

Theorem 5.5.1 *If $\mathcal{X}$ contains only a finite number of possible values, then the substitution and Archimidean axioms are necessary and sufficient for a preference relation to have an expected utility representation.*

Proof We will just sketch the proof that the axioms imply the existence of an expected utility representation; the proof of the converse is left as an exercise.
For full details, see **?**.
Since $\mathcal{X}$ is finite, and unless the consumer is indifferent among all possible choices, there must exist maximal and minimal sure things, say p^+ and p_- respectively.
By the substitution axiom, and a simple inductive argument, these are maximal and minimal in $\mathcal{L}$ as well as in $\mathcal{X}$. (If $\mathcal{X}$ is not finite, then an inductive argument can no longer be used and the Sure Thing Principle is required.)
From the Archimedean axiom, it can be deduced that for every other lottery, $\tilde{p}$, there exists a unique $V(\tilde{p})$ such that

$$\tilde{p} \sim V(\tilde{p})\, p^+ \oplus (1 - V(\tilde{p}))\, p_-.$$

It is easily seen that V represents $\succeq$.
Linearity can be shown as follows:
We leave it as an exercise to deduce from the axioms that if $\tilde{x} \sim \tilde{y}$ and $\tilde{z} \sim \tilde{t}$ then

$$\pi\tilde{x} \oplus (1-\pi)\,\tilde{z} \sim \pi\tilde{y} \oplus (1-\pi)\,\tilde{t}.$$

Define $\tilde{z} \equiv \pi\tilde{x} \oplus (1-\pi)\,\tilde{y}$.
Then, using the definitions of $V(\tilde{x})$ and $V(\tilde{y})$,

$$\begin{aligned}
\tilde{z} &\sim \pi\tilde{x} \oplus (1-\pi)\,\tilde{y} \\
&\sim \pi\left(V(\tilde{x})\,p^+ \oplus (1-V(\tilde{x}))\,p_-\right) \oplus (1-\pi)\left(V(\tilde{y})\,p^+ \oplus (1-V(\tilde{y}))\,p_-\right) \\
&= (\pi V(\tilde{x}) + (1-\pi)\,V(\tilde{y}))\,p^+ \oplus (\pi\,(1-V(\tilde{x})) + (1-\pi)\,(1-V(\tilde{y})))\,p_-
\end{aligned}$$

It follows that

$$V(\pi\tilde{x} \oplus (1-\pi)\,\tilde{y}) = \pi V(\tilde{x}) + (1-\pi)\,V(\tilde{y})\,.$$

This shows linearity for compound lotteries with only two possible outcomes: by an inductive argument, every lottery can be reduced recursively to a two-outcome lottery when there are only a finite number of possible outcomes altogether. **Q.E.D.**

Theorem 5.5.2 *For more general $\mathcal{L}$, to these conditions must be added some technical conditions and the Sure Thing Principle.*

Proof We will not consider the proof of this more general theorem. It can be found in **?**.
Q.E.D.

Note that expected utility depends only on the distribution function of the consumption plan.
Two consumption plans having very different consumption patterns across states of nature but the same probability distribution give the same utility. *E.g.* if wet days and dry days are equally likely, then an expected utility maximiser is indifferent between any consumption plan and the plan formed by switching consumption between wet and dry days.
The basic *objects of choice* under expected utility are not consumption plans but classes of consumption plans with the same cumulative distribution function.
Chapter 6 will consider the problem of portfolio choice in considerable depth. This chapter, however, must continue with some basic analysis of the choice between one riskfree and one risky asset, following **?**.
Such an example is sufficient to show several things:

1. There is no guarantee that the portfolio choice problem has any finite or unique solution unless the expected utility function is concave.

2. probably local risk neutrality and stuff like that too.

5.6 Jensen's Inequality and Siegel's Paradox

Theorem 5.6.1 (Jensen's Inequality) *The expected value of a (strictly) concave function of a random variable is (strictly) less than the same concave function of the expected value of the random variable.*
$E\left[u\left(\tilde{W}\right)\right] \leq u\left(E\left[\tilde{W}\right]\right)$ *when u is concave*
Similarly, the expected value of a (strictly) convex function of a random variable is (strictly) greater than the same convex function of the expected value of the random variable.

Proof There are three ways of motivating this result, but only one provides a fully general and rigorous proof. Without loss of generality, consider the concave case.

1. One can reinterpret the defining inequality (3.2.1) in terms of a discrete random vector $\tilde{\mathbf{x}}$ taking on the value $\mathbf{x}$ with probability π and $\mathbf{x}'$ with probability $1-\pi$:

   $\forall \mathbf{x} \neq \mathbf{x}' \in X, \pi \in (0,1)$

   $$f\left(\pi\mathbf{x} + (1-\pi)\mathbf{x}'\right) \geq \pi f(\mathbf{x}) + (1-\pi) f(\mathbf{x}'), \qquad (5.6.1)$$

   which just says that

   $$f\left(E\left[\tilde{\mathbf{x}}\right]\right) \geq E\left[f\left(\tilde{\mathbf{x}}\right)\right]. \qquad (5.6.2)$$

   An inductive argument can be used to extend the result to all discrete r.v.s with a finite number of possible values, but runs into problems if the number of possible values is either countably or uncountably infinite.

2. Using a similar approach to that used with the Taylor series expansion in (5.3.2), take expectations on both sides of the first order condition for concavity (3.2.3) given by Theorem 3.2.3, where the two vectors considered are the mean $E\left[\tilde{\mathbf{x}}\right]$ and a generic value $\tilde{\mathbf{x}}$:

   $$f(\tilde{\mathbf{x}}) \leq f\left(E\left[\tilde{\mathbf{x}}\right]\right) + f'\left(E\left[\tilde{\mathbf{x}}\right]\right)\left(\tilde{\mathbf{x}} - E\left[\tilde{\mathbf{x}}\right]\right). \qquad (5.6.3)$$

   Taking expectations on both sides, the first order term will again disappear, once more yielding:

   $$f\left(E\left[\tilde{\mathbf{x}}\right]\right) \geq E\left[f\left(\tilde{\mathbf{x}}\right)\right]. \qquad (5.6.4)$$

3. One can also appeal to the second order condition for concavity, and the second order Taylor series expansion of f around $E\left[\tilde{\mathbf{x}}\right]$:

   $$E[f(\tilde{x})] \quad = \quad f(E[\tilde{x}]) + \frac{1}{2} f''(x^*) \mathrm{Var}\left[\tilde{x}\right], \qquad (5.6.5)$$

for some x^* in the support of $\tilde{x}$. However, this supposes that x^* is fixed, whereas in fact it varies with the value taken on by $\tilde{x}$, and is itself a random variable, correlated with $\tilde{x}$.

Accepting this (wrong) approximation, if f is concave, then by Theorem 3.2.4 the second derivative is non-positive and the variance is non-negative, so

$$E[f(\tilde{x})] \quad \leq \quad f(E[\tilde{x}]). \tag{5.6.6}$$

The arguments for convex functions, strictly concave functions and strictly concave functions are almost identical.

Q.E.D.

This result is often useful with functions such as $x \mapsto \ln x$ and $x \mapsto \frac{1}{x}$.
To get a feel for the extent to which $E[f(\tilde{x})]$ differs from $f(E[\tilde{x}])$, we can again use the following (wrong) second order Taylor approximation based on (5.3.8):

$$E[f(\tilde{x})] \approx f(E[\tilde{x}]) + \frac{1}{2} f''(E[\tilde{x}]) \mathrm{Var}\,[\tilde{x}]\,. \tag{5.6.7}$$

This shows that the difference is larger the larger is the curvature of f (as measured by the second derivative at the mean of $\tilde{x}$) and the larger is the variance of $\tilde{x}$.
One area in which this idea can be applied is the computation of present values based on replacing uncertain future discount factors with point estimates derived from expected future interest rates.
Another nice application of Jensen's Inequality in finance is:

Theorem 5.6.2 (Siegel's Paradox) *Current forward (relative) prices can not all equal expected future spot prices.*

Proof Let F_t be the current forward price and $\tilde{S}_{t+1}$ the unknown future spot price. If

$$E_t\left[\tilde{S}_{t+1}\right] = F_t,$$

then Jensen's Inequality tells us that

$$\frac{1}{\tilde{F}_t} = \frac{1}{E_t\left[\tilde{S}_{t+1}\right]} < E_t\left[\frac{1}{\tilde{S}_{t+1}}\right],$$

except in the degenerate case where $\tilde{S}_{t+1}$ is known with certainty at time t.
But since the reciprocals of relatives prices are also relative prices, we have shown that our initial hypothesis is untenable in terms of a different numeraire.

Q.E.D.

The original and most obvious application of Siegel's paradox is in the case of currency exchange rates. In that case, $\left\{\tilde{F}_t\right\}$ and $\left\{\tilde{S}_t\right\}$ are stochastic processes representing forward and spot exhange rates respectively. It seems reasonable to assume that forward exchange rates are good predictors of spot exchange rates in the future, say:

$$\tilde{F}_t = E_t\left[\tilde{S}_{t+1}\right].$$

But this is an internally inconsistent hypothesis.
However, Siegel's paradox applies equally well to any theory which uses current prices as a predictor of future values. Another such theory which is enormously popular is the Efficient Markets Hypothesis of **?**. In its general form, this says that current prices should *fully reflect* all available information about (expected) future values. Attempts to make the words *fully reflect* in any way mathematically rigorous quickly run into problems.

5.7 Risk Aversion

An individual is **risk averse** if he or she is unwilling to accept, or indifferent to, any actuarially fair gamble.
(An individual is **strictly risk averse** if he or she is unwilling to accept any actuarially fair gamble.)

Figure 1.17.1 goes here.

i.e. $\forall W_0$ and $\forall p, h_1, h_2$ such that $ph_1 + (1-p)\,h_2 = 0$

$$W_0 \succeq (\succ) W_0 + ph_1 + (1-p)\,h_2$$

i.e.

$$u\,(W_0) \geq (>) pu\,(W_0 + h_1) + (1-p)\,u\,(W_0 + h_2)$$

The following interpretation of the above definition of risk aversion is based on Jensen's Inequality (see Section 5.6).
In other words, a (strictly) risk averse individual is one whose VNM utility function is (strictly) concave.
Similarly, a (strictly) risk loving individual is one whose VNM utility function is (strictly) convex.
Finally, a risk neutral individual is one whose VNM utility function is affine.

Most functions do not fall into any of these categories, and represent behaviour which is *locally risk averse* at some wealth levels and *locally risk loving* at other wealth levels.

However, in most of what follows we will find it convenient to assume that individuals are globally risk averse.

Individuals who are globally risk averse will never gamble, in the sense that they will never have a bet unless they believe that the expected return on the bet is positive. Thus assuming global risk aversion (and rational expectations) rules out the existence of lotteries (except with large rollovers) and most other forms of betting and gaming.

We can distinguish between local and global risk aversion. An individual is locally risk averse at w if $u''(w) < 0$ and globally risk averse if $u''(w) < 0 \; \forall w$.

Individuals who gamble are not globally risk averse but may still be locally risk averse around their current wealth level.

Cut and paste relevant quotes from Purfield-Waldron papers in here.

Some people are more risk averse than others; some functions are more concave than others; how do we measure this?

The importance and usefulness of the Arrow-Pratt measures of risk aversion which we now define will become clearer as we proceed, in particular from the analysis of the portfolio choice problem:

Definition 5.7.1 *(The* Arrow-Pratt coefficient of) absolute risk aversion *is:*

$$R_A(w) = -u''(w)/u'(w)$$

which is the same for u and $au + b$.

Note that this varies with the level of wealth.

$u'(w)$ alone is meaningless, as u and u' can be multiplied by any positive constant and still represent the same preferences. However, the above ratio is independent of the expected utility function chosen to represent the preferences.

Definition 5.7.2 *(The* Arrow-Pratt coefficient of) relative risk aversion *is:*

$$R_R(w) = wR_A(w)$$

The utility function u exhibits *increasing (constant, decreasing) absolute risk aversion* (IARA, CARA, DARA) $\iff$

$$R'_A(w) > (=, <) \; 0 \; \forall w.$$

The utility function u exhibits *increasing (constant, decreasing) relative risk aversion* (IRRA, CRRA, DRRA) $\iff$

$$R_R'(w) > (=, <)\, 0\ \forall w.$$

Note:

- CARA or IARA $\Rightarrow$ IRRA
- CRRA or DRRA $\Rightarrow$ DARA

Here are some examples of utility functions and their risk measures:

- Quadratic utility (IARA, IRRA):

$$\begin{aligned}
u(w) &= w - \frac{b}{2}w^2, \qquad b > 0; \\
u'(w) &= 1 - bw; \\
u''(w) &= -b < 0 \\
R_A(w) &= \frac{b}{1-bw} \\
\frac{dR_A(w)}{dw} &= \frac{b^2}{(1-bw)^2} > 0
\end{aligned}$$

  In this case, marginal utility is positive and utility increasing if and only if $w < 1/b$. $1/b$ is called the bliss point of the quadratic utility function. For realism, $1/b$ should be rather large and thus b rather small.

- Negative exponential utility (CARA, IRRA):

$$\begin{aligned}
u(w) &= -e^{-bw}, \quad b > 0; \\
u'(w) &= be^{-bw} > 0; \\
u''(w) &= -b^2 e^{-bw} < 0 \\
R_A(w) &= b \\
\frac{dR_A(w)}{dw} &= 0
\end{aligned}$$

- Narrow power utility (CRRA, DARA):

$$u(w) = \frac{B}{B-1} w^{1-\frac{1}{B}},\ w > 0, B > 0, B \neq 1$$

The proofs in this case are left as an exercise. The solution is roughly as follows:[4]

$$\begin{aligned}
u(z) &= \frac{B}{B-1} z^{1-\frac{1}{B}},\ z>0, B>0 \\
u'(z) &= z^{-\frac{1}{B}} \\
u''(z) &= -\frac{1}{B} z^{-\frac{1}{B}-1} \\
R_A(z) &= \frac{1}{B} z^{-1} \\
R_R(z) &= \frac{1}{B} \\
\frac{dR_A(z)}{dz} &= -\frac{1}{B} z^{-2} < 0 \\
\frac{dR_R(z)}{dz} &= 0
\end{aligned}$$

- Extended power utility ():[5]

$$u(w) = \frac{1}{(C+1)B} (A+Bw)^{C+1}$$

Note that a risk neutral investor will seek to invest his or her entire wealth in the asset with the highest expected return. If all investors are risk neutral, then prices will adjust in equilibrium so that all securities have the same expected return. If there are risk averse or risk loving investors, then there is no reason for this result to hold, and in fact it almost certainly will not.

5.8 The Mean-Variance Paradigm

Three arguments are commonly used to motivate the mean-variance framework for analysis of the portfolio choice problem:

1. Taylor-approximated utility functions (see Section 2.8):

$$\begin{aligned}
u(\tilde{W}) &= u(E[\tilde{W}]) + u'(E[\tilde{W}])(\tilde{W} - E[\tilde{W}]) \\
&+ \frac{1}{2} u''(E[\tilde{W}])(\tilde{W} - E[\tilde{W}])^2 + R_3 \qquad (5.8.1) \\
R_3 &= \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[\tilde{W}])(\tilde{W} - E[\tilde{W}])^n \qquad (5.8.2)
\end{aligned}$$

[4]Add comments: $u'(w) > 0$, $u''(w) < 0$ and $u'(w) \to 1/w$ as $B \to 1$, so that $u(w) \to \ln w$.
[5]Check **?** for details of this one.

which implies:

$$E[u(\tilde{W})] = u(E[\tilde{W}]) + \frac{1}{2}u''(E[\tilde{W}])\sigma^2(\tilde{W}) + E[R_3] \quad (5.8.3)$$

where

$$E[R_3] = \sum_{n=3}^{\infty} \frac{1}{n!} u^{(n)}(E[\tilde{W}]) m^n(\tilde{W}) \quad (5.8.4)$$

It follows that the sign of the nth derivative of the utility function determines the direction of preference for the nth central moment of the probability distribution of terminal wealth. For example, a positive third derivative implies a preference for greater skewness. It can be shown fairly easily that an increasing utility function which exhibits non-increasing absolute risk aversion has a non-negative third derivative.

2. Quadratic utility:

$$\begin{aligned} E[u(\tilde{W})] &= E[\tilde{W}] - \frac{b}{2}E[\tilde{W}^2] && (5.8.5) \\ &= E[\tilde{W}] - \frac{b}{2}\left((E[\tilde{W}])^2 + \sigma^2(\tilde{W})\right) && (5.8.6) \\ &= u(E[\tilde{W}]) - \frac{b}{2}\mathrm{Var}[\tilde{W}] && (5.8.7) \end{aligned}$$

3. Normally distributed asset returns.

Note that the expected utility axioms are neither necessary nor sufficient to guarantee that the Taylor approximation to n moments is a valid representation of the utility function. Some counterexamples of both types are probably called for here, or maybe can be left as exercises.
Extracts from my PhD thesis can be used to talk about signing the first three coefficients in the Taylor expansion, and to speculate about further extensions to higher moments.

5.9 The Kelly Strategy

In a multi-period, discrete time, investment framework, investors will be concerned with both growth (return) and security (risk). There will be a trade-off between the two, and investors will be concerned with finding the optimal trade-off. This, of course, depends on preferences, but some useful benchmarks exist.
There are three ways of measuring growth:

1. the expected wealth at time t

2. the expected rate of growth of wealth up to time t
3. the expected first passage time to reach a critical level of wealth, say £1m

Likewise, there are three ways of measuring security:

1. the probability of reaching a critical level of wealth, say £1m, by a specific time t
2. the probability that the wealth process lies above some critical path, say above b_t at time t, $t = 1, 2, \ldots$
3. the probability of reaching some goal U before falling to some low level of wealth L

It can be shown that the ***?*** *strategy* both maximises the long run exponential growth rate and minimises the expected time to reach large goals.
All this is also covered in **?** which is reproduced in **?**.

5.10 Alternative Non-Expected Utility Approaches

Those not happy with the explanations of choice under uncertainty provided within the expected utility paradigm have proposed various alternatives in recent years. These include both vague qualitative waffle about fun and addiction and more formal approaches looking at maximum or minimum possible payoffs, irrational expectations, etc. Before proceeding, the reader might want to review Section 3.2.

Chapter 6

PORTFOLIO THEORY

6.1 Introduction

Portfolio theory is an important topic in the theory of choice under uncertainty. It deals with the problem facing an investor who must decide how to distribute an initial wealth of, say, W_0 among a number of single-period investment opportunities, called *securities* or *assets*.

The choice of *portfolio* will depend on both the investor's preferences and his beliefs about the uncertain payoffs of the various securities. A *mutual fund* is just a special type of (managed) portfolio.

The chapter begins by considering some issues of definition and measurement. Section 6.3 then looks at the portfolio choice problem in a general expected utility context. Section 6.4 considers the same problem from a mean-variance perspective. This leads on to a discussion of the properties of equilibrium security returns in Section 6.5.

6.2 Notation and preliminaries

6.2.1 Measuring rates of return

Good background reading for this section is **?**.

A rate of interest (growth, inflation, *&c.*) is not properly defined unless we state the time period to which it applies and the method of compounding to be used.

2% per annum is very different from 2% per month.

Table 6.1 illustrates what happens to £100 invested at 10% per annum as we change the interval of compounding. The final calculation in the table uses the

Compounded

Annually	£100	$\rightarrow$	£110
Semi-annually	£100	$\rightarrow$	£100 $\times (1.05)^2 = \pounds 110.25$
Quarterly	£100	$\rightarrow$	£100 $\times (1.025)^4 = \pounds 110.381\ldots$
Monthly	£100	$\rightarrow$	£100 $\times \left(1 + \frac{.10}{12}\right)^{12} = \pounds 110.471\ldots$
Weekly	£100	$\rightarrow$	£100 $\times \left(1 + \frac{.10}{52}\right)^{52} = \pounds 110.506\ldots$
Daily	£100	$\rightarrow$	£100 $\times \left(1 + \frac{.10}{365}\right)^{365} = \pounds 110.515\ldots$
Continuously	£100	$\rightarrow$	£100 $\times e^{0.10} = \pounds 110.517\ldots$

Table 6.1: The effect of an interest rate of 10% per annum at different frequencies of compounding.

fact that

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n = e^r$$

where $e \approx 2.7182\ldots$
This is sometimes used as the definition of e but others prefer to start with

$$e^r \equiv 1 + r + \frac{r^2}{2!} + \frac{r^3}{3!} + \ldots = \sum_{j=0}^{\infty} \frac{r^j}{j!}$$

where $n! \equiv n\,(n-1)\,(n-2)\ldots 3.2.1$ and $0! \equiv 1$ by convention.
There are five concepts which we need to be familiar with:

1. Discrete compounding:

$$P_t = \left(1+\frac{r}{n}\right)^{nt} P_0.$$

   We can solve this equation for any of five quantities given the other four:

   (a) present value

   (b) final value

   (c) implicit rate of return

   (d) time

   (e) interval of compounding

2. Continuous compounding:

$$P_t = e^{rt} P_0.$$

We can solve this equation for any of four quantities given the other three ((a)–(d) above).

Note also that the exponential function is its own derivative; that $y = 1+r$ is the tangent to $y = e^r$ at $r = 0$ and hence that $e^r > 1 + r \; \forall r$. In other words, given an initial value, continuous compounding yields a higher terminal value than discrete compounding for all interest rates, positive and negative, with equality for a zero interest rate only. Similarly, given a final value, continuous discounting yields a higher present value than does discrete.

3. Aggregating and averaging returns across portfolios ($\tilde{r}_{\mathbf{w}} = \mathbf{w}^\top \tilde{\mathbf{r}}$) and over time:

$$\begin{aligned}
P_t &= e^{rt} P_0 \\
P_{t+\Delta t} &= e^{r\Delta t} P_t \\
r &= \frac{\ln P_{t+\Delta t} - \ln P_t}{\Delta t} \\
r(s) &= \frac{d \ln P_s}{ds} \\
\int_0^t r(s) ds &= \int_0^t d \ln P_s = \ln P_t - \ln P_0 \\
P_t &= P_0 e^{\int_0^t r(s) ds}
\end{aligned}$$

Discretely compounded rates aggregate nicely across portfolios; continuously compounded rates aggregate nicely across time.

4. The *(net) present value (NPV)* of a stream of cash flows,

$$P_0, P_1, \ldots, P_T$$

is

$$NPV(r) \equiv \frac{P_0}{(1+r)^0} + \frac{P_1}{(1+r)^1} + \frac{P_2}{(1+r)^2} + \ldots + \frac{P_T}{(1+r)^T}.$$

5. The *internal rate of return (IRR)* of the stream of cash flows,

$$P_0, P_1, \ldots, P_T,$$

is the solution of the polynomial of degree T obtained by setting the NPV equal to zero:

$$\frac{P_0}{(1+r)^0} + \frac{P_1}{(1+r)^1} + \frac{P_2}{(1+r)^2} + \ldots + \frac{P_T}{(1+r)^T} = 0.$$

W_0	$=$	the investor's initial wealth
W_1	$=$	the investor's desired expected final wealth
N	$=$	number of risky assets
r_f	$=$	return on the riskfree asset
$\tilde{r}_j \in \Re$	$=$	return on jth risky asset
$\tilde{\mathbf{r}} \in \Re^N$	$=$	$(\tilde{r}_1, \ldots, \tilde{r}_N)$
$\mathbf{e} = E[\tilde{\mathbf{r}}] \in \Re^N$	$=$	vector of expected returns
$\mathbf{V} \in \Re^{N \times N}$	$=$	variance-covariance matrix of returns
$\mathbf{1}$	$=$	$(1, 1, \ldots, 1)$
	$=$	N-dimensional vector of 1s
$w_j \in \Re$	$=$	amount invested in jth risky asset
$\mathbf{w} \in \Re^N$	$=$	$(w_1, \ldots, w_n)$
$\tilde{r}_{\mathbf{w}} = \mathbf{w}^\top \tilde{\mathbf{r}}$	$=$	return on the portfolio $\mathbf{w}$
$\mu \equiv E[\tilde{r}_{\mathbf{w}}] \equiv \frac{W_1}{W_0} \in \Re$	$=$	the investor's desired expected return

Table 6.2: Notation for portfolio choice problem

In general, the polynomial defining the IRR has T (complex) roots. Conditions have been derived under which there is only one meaningful real root to this polynomial equation, in other words one corresponding to a positive IRR.[1]

Consider a quadratic example.

Simple rates of return are additive across portfolios, so we use them in one period cross sectional studies, in particular in this chapter.
Continuously compounded rates of return are additive across time, so we use them in multi-period single variable studies, such as in Chapter 7.
Consider as an example the problem of calculating mortgage repayments.

6.2.2 Notation

The investment opportunity set for the portfolio choice problem will generally consist of N risky assets. From time to time, we will add a riskfree asset. The notation used throughout this chapter is set out in Table 6.2. The presentation is in terms of a single period problem, and the unconditional distribution of returns. The analysis of the multiperiod, infinite horizon, discrete time problem, concentrating on the conditional distribution of the next period's returns given this period's, is quite similar.[2]

[1] These conditions are discussed in **?**.
[2] Make this into an exercise.

Definition 6.2.1 $\mathbf{w}$ *is said to be a* unit cost *or* normal portfolio *if its weights sum to 1 (*$\mathbf{w}^\top\mathbf{1} = 1$*).*

The portfolio held by an investor with initial wealth W_0 can be thought of either as a $\mathbf{w}$ with $\mathbf{w}^\top\mathbf{1} = W_0$ or as the corresponding normal portfolio, $\frac{1}{W_0}\mathbf{w}$. It will hopefully be clear from the context which meaning of 'portfolio' is intended.

Definition 6.2.2 $\mathbf{w}$ *is said to be a* zero cost *or* hedge portfolio *if its weights sum to 0 (*$\mathbf{w}^\top\mathbf{1} = 0$*).*

The vector of net trades carried out by an investor moving from the portfolio $\mathbf{w}_0$ to the portfolio $\mathbf{w}_1$ can be thought of as the hedge portfolio $\mathbf{w}_1 - \mathbf{w}_0$.

Definition 6.2.3 Short-selling *a security means owning a negative quantity of it.*

In practice short-selling means promising (credibly) to pay someone the same cash flows as would be paid by a security that one does not own, always being prepared, if required, to pay the current market price of the security to end the arrangement. Thus when short-selling is allowed $\mathbf{w}$ can have negative components; if short-selling is not allowed, then the portfolio choice problem will have non-negativity constraints, $w_i \geq 0$ for $i = 1, \ldots, N$.
A number of further comments are in order at this stage.

1. In the literature, initial wealth is often normalised to unity ($W_0 = 1$), but the development of the theory will be more elegant if we avoid this.

2. Since we will be dealing on occasion with hedge portfolios, we will in future avoid the concepts of *rate of return* and *net return*, which are usually thought of as the ratio of profit to initial investment. These terms are meaningless for a hedge portfolio as the denominator is zero.

   Instead, we will speak of the *gross return* on a portfolio or the *portfolio payoff*. This can be defined unambiguously as follows. There is no ambiguity about the payoff on one of the original securities, which is just the gross return per pound invested. The payoff on a unit cost or normal portfolio is equivalent to the gross return. It is just $\mathbf{w}^\top\tilde{\mathbf{r}}$. The payoff on a zero cost portfolio can also be defined as $\mathbf{w}^\top\tilde{\mathbf{r}}$.

   Note that where we initially worked with net rates of return ($\frac{P_1}{P_0} - 1$), we will deal henceforth with gross rates of return ($\frac{P_1}{P_0}$).

6.3 The Single-period Portfolio Choice Problem

6.3.1 The canonical portfolio problem

Unless otherwise stated, we assume that individuals:

1. have von Neumann-Morgenstern (VNM) utilities:
   i.e. preferences have the expected utility representation:

$$\begin{aligned} v(\tilde{z}) &= \mathrm{E}[u(\tilde{z})] \\ &= \int u(z) dF_{\tilde{z}}(z) \end{aligned}$$

   where v is the utility function for random variables (gambles, lotteries) and u is the utility function for sure things.

2. prefer more to less
   i.e. u is increasing:

$$u'(z) > 0 \ \ \forall z \tag{6.3.1}$$

3. are (strictly) risk-averse
   i.e. u is strictly concave:

$$u''(z) < 0 \ \ \forall z \tag{6.3.2}$$

Date 0 investment:

- w_j (pounds) in jth risky asset, $j = 1, \ldots, N$
- $(W_0 - \sum_j w_j)$ in risk free asset

Date 1 payoff:

- $w_j \tilde{r}_j$ from jth risky asset
- $(W_0 - \sum_j w_j) r_f$ from risk free asset

It is assumed here that there are no constraints on short-selling or borrowing (which is the same as short-selling the riskfree security).
The solution is found as follows:
Choose w_js to maximize expected utility of date 1 wealth,

$$\begin{aligned} \tilde{W} &= (W_0 - \sum_j w_j) r_f + \sum_j w_j \tilde{r}_j \\ &= W_0 r_f + \sum_j w_j (\tilde{r}_j - r_f) \end{aligned}$$

i.e.

$$\max_{\{w_j\}} f(w_1, \ldots, w_N) \equiv E[u(W_0 r_f + \sum_j w_j(\tilde{r}_j - r_f))]$$

The first order conditions are:

$$E[u'(\tilde{W})(\tilde{r}_j - r_f)] = 0 \quad \forall j. \tag{6.3.3}$$

The Hessian matrix of the objective function is:

$$\mathbf{A} \equiv E\left[u''\left(\tilde{W}\right)(\tilde{\mathbf{r}} - r_f\mathbf{1})(\tilde{\mathbf{r}} - r_f\mathbf{1})^\top\right]. \tag{6.3.4}$$

$\mathbf{h}^\top \mathbf{A}\mathbf{h} < 0 \, \forall \mathbf{h} \neq \mathbf{0}_N$ if and only if $u'' < 0$. Since we have assumed that investor behaviour is risk averse, $\mathbf{A}$ is a negative definite matrix and, by Theorem 3.2.4, f is a strictly concave (and hence strictly quasiconcave) function. Thus, under the present assumptions, Theorems 3.3.3 and 3.3.4 guarantee that the first order conditions have a unique solution. The trivial case in which the random returns are not really random at all can be ignored.

The rest of this section should be omitted until I figure out what is going on.

Another way of writing (6.3.3) is:

$$E[u'(\tilde{W})\tilde{r}_j] = E[u'(\tilde{W})]r_f \quad \forall j, \tag{6.3.5}$$

or

$$\text{Cov}\left[u'(\tilde{W}), \tilde{r}_j\right] + E[u'(\tilde{W})]E[\tilde{r}_j] = E[u'(\tilde{W})]r_f \quad \forall j, \tag{6.3.6}$$

or

$$E[\tilde{r}_j - r_f] = \frac{\text{Cov}\left[u'(\tilde{W}), \tilde{r}_j\right]}{E[u'(\tilde{W})]} \quad \forall j, \tag{6.3.7}$$

Suppose p_j is the price of the random payoff $\tilde{x}_j$. Then $\tilde{r}_j = \frac{\tilde{x}_j}{p_j}$ and

$$p_j = E\left[\frac{u'(\tilde{W})}{E[u'(\tilde{W})]r_f}\tilde{x}_j\right] \quad \forall j. \tag{6.3.8}$$

In other words, payoffs are valued by taking their expected present value, using the *stochastic discount factor* $\frac{u'(\tilde{W})}{E[u'(\tilde{W})]r_f}$, which ends up being the same for all investors. Practical corporate finance and theoretical asset pricing models to a large extent are (or should be) concerned with analysing this discount factor.

(We could consider here the explicit example with quadratic utility from the problem sets.)

6.3.2 Risk aversion and portfolio composition

For the moment, assume only one risky asset ($N = 1$).
We first consider the concept of *local risk neutrality*. The optimal investment in the risky asset is positive
$\Longleftrightarrow$
The objective function is increasing at $a = 0$
$\Longleftrightarrow$

$$f'(a) > 0 \tag{6.3.9}$$

$\Longleftrightarrow$

$$E[u'(W_0 r_f)(\tilde{r} - r_f)] > 0 \tag{6.3.10}$$

$\Longleftrightarrow$

$$u'(W_0 r_f) E[(\tilde{r} - r_f)] > 0 \tag{6.3.11}$$

$\Longleftrightarrow$

$$E[\tilde{r}] > E[r_f] = r_f$$

This is the property of local risk neutrality — a risk averse investor will always prefer a little of a risky asset paying a higher expected return than r_f to none of the risky asset.

Definition 6.3.1 *Let* $f: X \to \Re_{++}$ *be a positive-valued function defined on* $X \subseteq \Re^k_{++}$. *Then the* elasticity of f with respect to x_i at $\mathbf{x}^*$ *is*

$$\frac{x_i^*}{f(\mathbf{x})} \frac{\partial f}{\partial x_i}(\mathbf{x}^*).$$

Roughly speaking, the elasticity is just $\frac{\partial \ln f}{\partial \ln x_i,}$ or the slope of the graph of the function on log-log graph paper.
A function is said to be *inelastic* when the absolute value of the elasticity is less than unity; and *elastic* when the absolute value of the elasticity is greater than unity. The borderline case is called a *unit elastic* function.
One useful application of elasticity is in analysing the behaviour of the total revenue function associated with a particular inverse demand function, $P(Q)$. We have:

$$\frac{dP(Q)Q}{dQ} = q\frac{dP}{dQ} + P \tag{6.3.12}$$

$$= P\left(1 + \frac{1}{\eta}\right). \tag{6.3.13}$$

Hence, total revenue is constant or maximised or minimised where elasticity equals -1; increasing when elasticity is less than -1 (demand is elastic); and decreasing when elasticity is between 0 and -1 (demand is inelastic).
Now consider other properties of asset demands (assuming that $E[\tilde{r}] > r_f$):

- DARA ⇒ risky asset normal $\left(\frac{da}{dW_0} > 0\right)$
- CARA ⇒ $\left(\frac{da}{dW_0} = 0\right)$
- IARA ⇒ risky asset inferior $\left(\frac{da}{dW_0} < 0\right)$

Define the wealth elasticity of demand for the risky asset to be

$$\eta = \frac{W_0}{a}\frac{da}{dW_0}. \tag{6.3.14}$$

Then we have

$$\begin{aligned} \frac{d\left(\frac{a}{W_0}\right)}{dW_0} &= \frac{W_0\frac{da}{dW_0} - a}{W_0^2} \\ &= \frac{a}{W_0^2}(\eta - 1). \end{aligned}$$

Note that

$$\text{sign}\ \left(\frac{d\frac{a}{W_0}}{dW_0}\right) = \text{sign}\ \ (\eta - 1) \tag{6.3.15}$$

since by assuming a positive expected risk premium on the risky asset we guarantee (by local risk-neutrality) that a is positive.

- DRRA ⇒ increasing proportion of wealth invested in the risky asset $(\eta > 1)$
- CRRA ⇒ constant proportion of wealth invested in the risky asset $(\eta = 1)$
- IRRA ⇒ decreasing proportion of wealth invested in the risky asset $(\eta < 1)$

Theorem 6.3.1 *DARA ⇒ RISKY ASSET NORMAL*

Proof By implicit differentiation of the now familiar first order condition (6.3.3), which can be written:

$$E[u'(W_0 r_f + a(\tilde{r} - r_f))(\tilde{r} - r_f)] = 0, \tag{6.3.16}$$

we have

$$\frac{da}{dW_0} = \frac{E[u''(\tilde{W})(\tilde{r} - r_f)]r_f}{-E[u''(\tilde{W})(\tilde{r} - r_f)^2]}. \tag{6.3.17}$$

By concavity, the denominator is positive. Therefore:

$$\text{sign } (da/dW_0) = \text{sign } \{E[u''(\tilde{W})(\tilde{r} - r_f)]\} \tag{6.3.18}$$

We will show that both are positive.
For decreasing absolute risk aversion:[3]

$$\begin{aligned} \tilde{r} > r_f &\Rightarrow R_A(\tilde{W}) < R_A(W_0 r_f) \\ \tilde{r} \leq r_f &\Rightarrow R_A(\tilde{W}) \geq R_A(W_0 r_f) \end{aligned}$$

Multiplying both sides of each inequality by $-u'(\tilde{W})(\tilde{r} - r_f)$ gives respectively:

$$u''(\tilde{W})(\tilde{r} - r_f) > -R_A(W_0 r_f) u'(\tilde{W})(\tilde{r} - r_f) \tag{6.3.19}$$

in the event that $\tilde{r} > r_f$, and

$$u''(\tilde{W})(\tilde{r} - r_f) \geq -R_A(W_0 r_f) u'(\tilde{W})(\tilde{r} - r_f) \tag{6.3.20}$$

(the same result) in the event that $\tilde{r} \leq r_f$
Integrating over both events implies:

$$E[u''(\tilde{W})(\tilde{r} - r_f)] > -R_A(W_0 r_f) E[u'(\tilde{W})(\tilde{r} - r_f)], \tag{6.3.21}$$

provided that $\tilde{r} > r_f$ with positive probability.
The RHS of inequality (6.3.21) is 0 at the optimum, hence the LHS is positive as claimed.

Q.E.D.

The other results are proved similarly (exercise!).

6.3.3 Mutual fund separation

Commonly, investors delegate portfolio choice to mutual fund operators or managers. We are interested in conditions under which large groups of investors will agree on portfolio composition. For example, all investors with similar utility functions might choose the same portfolio, or all investors with similar probability beliefs might choose the same portfolio. More realistically, we may be able to define a group of investors whose portfolio choices all lie in a subspace of small dimension (say 2) of the N-dimensional portfolio space. The first such result is due to **?**.

[3] Think about whether separating out the case of $\tilde{r} = r_f$ is necessary.

Theorem 6.3.2 $\exists$ *Two fund monetary separation*
i.e. Agents with different wealths (but the same increasing, strictly concave, VNM utility) hold the same risky unit cost portfolio, $\mathbf{p}^*$ *say, (but may differ in the mix of the riskfree asset and risky portfolio)*
i.e. $\forall$ *portfolios* $\mathbf{p}$*, wealths* W_0*,* $\exists\lambda$ *s.t.*

$$E\left[u\left(W_0 r_f + \lambda W_0 \mathbf{p}^{*\top}\left(\tilde{\mathbf{r}} - r_f \mathbf{1}\right)\right)\right] \geq E\left[u\left(W_0 r_f + \mathbf{p}^{\top}\left(\tilde{\mathbf{r}} - r_f \mathbf{1}\right)\right)\right] \quad (6.3.22)$$

$$\Longleftrightarrow$$

Risk-tolerance $(1/R_A(z))$ *is linear (including constant)*
i.e. $\exists$ ***H****yperbolic* ***A****bsolute* ***R****isk* ***A****version (HARA, incl. CARA)*
i.e. the utility function is of one of these types:

- *Extended power:* $u(z) = \frac{1}{(C+1)B}(A + Bz)^{C+1}$
- *Logarithmic:* $u(z) = \ln(A + Bz)$
- *Negative exponential:* $u(z) = -\frac{A}{B}\exp\{Bz\}$

where A*,* B *and* C *are chosen to guarantee* $u' > 0$*,* $u'' < 0$*.*
i.e. marginal utility satisfies

$$u'(z) = (A + Bz)^C \quad \text{or} \quad u'(z) = A\exp\{Bz\} \quad (6.3.23)$$

where A*,* B *and* C *are again chosen to guarantee* $u' > 0$*,* $u'' < 0$*.*

Proof The proof that these conditions are necessary for two fund separation is difficult and tedious. The interested reader is referred to **?**.
We will show that $u'(z) = (A + Bz)^C$ is sufficient for two-fund separation.
The optimal dollar investments w_j are the unique solution to the first order conditions:

$$0 = E[u'(\tilde{W})\frac{\delta\tilde{W}}{\delta w_i}] \quad (6.3.24)$$

$$= E[(A + B\tilde{W})^C(\tilde{r}_i - r_f)] \quad (6.3.25)$$

$$= E[(A + BW_0 r_f + \sum_j Bw_j(\tilde{r}_j - r_f))^C(\tilde{r}_i - r_f)], \quad (6.3.26)$$

or equivalently to the system of equations

$$E[(1 + \sum_j \frac{Bw_j}{A + BW_0 r_f}(\tilde{r}_j - r_f))^C(\tilde{r}_i - r_f)] = 0 \quad (6.3.27)$$

or

$$E[(1+\sum_j x_j(\tilde{r}_j - r_f))^C(\tilde{r}_i - r_f)] = 0 \tag{6.3.28}$$

where

$$x_j = \frac{Bw_j}{A + BW_0 r_f}.$$

The unique solutions for x_j are clearly independent of W_0 which does not appear in (6.3.28).
Since A and B do not appear either, the unique solutions for x_j are also independent of those parameters.
However, they do depend on C.
But the risky portfolio weights are

$$\frac{w_i}{\sum_j w_j} = \frac{Bw_i/(A + BW_0 r_f)}{\sum_j Bw_j/(A + BW_0 r_f)} \tag{6.3.29}$$

$$= \frac{x_i}{\sum_j x_j} \tag{6.3.30}$$

and so are also independent of initial wealth.
Since the dollar investment in the jth risky asset satisfies:

$$w_j = x_j(\frac{A}{B} + W_0 r_f) \tag{6.3.31}$$

we also have in this case that the dollar investment in the common risky portfolio is a linear function of the initial wealth. The other sufficiency proofs are similar and are left as exercises.

Q.E.D.

Some humorous anecdotes about Cass may now follow.

6.4 Mathematics of the Portfolio Frontier

6.4.1 The portfolio frontier in $\Re^N$: risky assets only

The portfolio frontier

Definition 6.4.1 *The* (mean-variance) portfolio frontier *is the set of solutions to the mean-variance portfolio choice problem faced by a (risk-averse) investor with an initial wealth of W_0 who desires an expected final wealth of at least $W_1 \equiv \mu W_0$ (or, equivalently, an expected rate of return of μ), but with the smallest possible variance of final wealth.*

The mean-variance frontier can also be called the *two-moment portfolio frontier*, in recognition of the fact that the same approach can be extended (with difficulty) to higher moments.

The mean-variance portfolio frontier is a subset of the *portfolio space*, which is just $\Re^N$. However, introductory treatments generally present it (without proof) as the envelope function, in mean-variance space or mean-standard deviation space ($\Re_+ \times \Re$), of the variance minimisation problem.

Definition 6.4.2 $\mathbf{w}$ *is a* frontier portfolio
$\Longleftrightarrow$
its return has the minimum variance among all portfolios that have the same cost, $\mathbf{w}^\top \mathbf{1}$*, and the same expected payoff,* $\mathbf{w}^\top \mathbf{e}$*.*

We will begin by supposing that all assets are risky. Formally, the frontier portfolio corresponding to initial wealth W_0 and expected return μ (expected terminal wealth μW_0) is the solution to the quadratic programming problem:

$$\min_{\mathbf{w}} \mathbf{w}^\top \mathbf{V} \mathbf{w} \tag{6.4.1}$$

subject to the linear constraints:

$$\mathbf{w}^\top \mathbf{1} = W_0 \tag{6.4.2}$$

and

$$\mathbf{w}^\top \mathbf{e} \geq W_1 = \mu W_0. \tag{6.4.3}$$

The first constraint is just the budget constraint, while the second constraint states that the expected rate of return on the portfolio is at least the desired mean return μ.

The frontier in this case is the set of solutions for all values of W_0 and W_1 (or μ) to this variance minimisation problem, or to the equivalent maximisation problem:

$$\max_{\mathbf{w}} -\mathbf{w}^\top \mathbf{V} \mathbf{w} \tag{6.4.4}$$

subject to the same linear constraints (6.4.2) and (6.4.3).

The properties of this two-moment frontier are well known, and can be found, for example, in **?** or **?**. The notation here follows **?**. The derivation of the mean-variance frontier is generally presented in the literature in terms of portfolio *weight* vectors or, equivalently, assuming that initial wealth, W_0, equals 1. This assumption is not essential and will be avoided.

The solution

The inequality constrained maximisation problem (6.4.4) is just a special case of the canonical quadratic programming problem considered at the end of Section 3.5, except that it has explicitly one equality constraint and one inequality constraint.
To avoid degeneracies, we require:

1. that not every portfolio has the same expected return, *i.e.*

$$\mathbf{e} \neq E[\tilde{r}_1]\mathbf{1}, \tag{6.4.5}$$

   and in particular that $N > 1$.

2. that the variance-covariance matrix, $\mathbf{V}$, is (strictly) positive definite. We already know from (1.12.4) that $\mathbf{V}$ must be positive semi-definite, but we require this slightly stronger condition. To see why, suppose

$$\exists \mathbf{w} \neq \mathbf{0}_N \; s.t. \; \mathbf{w}^\top \mathbf{V} \mathbf{w} = 0 \tag{6.4.6}$$

   Then $\exists$ a portfolio whose return $\mathbf{w}^\top \tilde{\mathbf{r}} = \tilde{r}_{\mathbf{w}}$ has zero variance.

   This implies that $\tilde{r}_{\mathbf{w}} = r_0$ (say) w.p.1 or, essentially, that this portfolio is riskless.

   Arbitrage will force the returns on all riskless assets to be equal in equilibrium, so this situation is equivalent economically to the introduction of a riskless asset later.

In the portfolio problem, the place of the matrix $\mathbf{A}$ in the canonical quadratic programming problem is taken by the (symmetric) negative definite matrix, $-\mathbf{V}$, which is just the negative of the variance-covariance matrix of asset returns; $\mathbf{g}^1 = \mathbf{1}^\top$ and $\alpha_1 = W_0$; and $\mathbf{g}^2 = \mathbf{e}^\top$ and $\alpha_2 = W_1$. (6.4.5) guarantees that the $2 \times N$ matrix $\mathbf{G}$ is of full rank 2.
The parallels are a little fuzzy in the case of the budget constraint since it is really an equality constraint.
(3.5.39) says that the optimal $\mathbf{w}$ is a linear combination of the two columns of the $N \times 2$ matrix

$$\mathbf{V}^{-1}\mathbf{G}^\top \left(\mathbf{G}\mathbf{V}^{-1}\mathbf{G}^\top\right)^{-1},$$

with columns weighted by initial wealth W_0 and expected final wealth, W_1.
We will call these columns $\mathbf{g}$ and $\mathbf{h}$ and write the solution as

$$\mathbf{w} = W_0\mathbf{g} + W_1\mathbf{h} = W_0\left(\mathbf{g} + \mu\mathbf{h}\right). \tag{6.4.7}$$

The components of $\mathbf{g}$ and $\mathbf{h}$ are functions of the means and variances of security returns. Thus the vector of optimal portfolio proportions,

$$\frac{1}{W_0}\mathbf{w} = \mathbf{g} + \mu\mathbf{h}, \tag{6.4.8}$$

is independent of the initial wealth W_0.
It is easy to see the economic interpretation of $\mathbf{g}$ and $\mathbf{h}$:

- $\mathbf{g}$ is the frontier portfolio corresponding to $W_0 = 1$ and $W_1 = 0$. In other words, it is the normal portfolio which would be held by an investor whose objective was to (just) go bankrupt with minimum variance.

- Similarly, $\mathbf{h}$ is the frontier portfolio corresponding to $W_0 = 0$ and $W_1 = 1$. In other words, it is the hedge portfolio which would be purchased by a variance-minimising investor in order to increase his expected final wealth by one unit.

Alternatively, (3.5.35) says that the optimal $\mathbf{w}$ is a linear combination of the two columns of the $N \times 2$ matrix

$$\frac{1}{2}\mathbf{V}^{-1}\mathbf{G}^\top = \left(\begin{array}{cc} \frac{1}{2}\mathbf{V}^{-1}\mathbf{1} & \frac{1}{2}\mathbf{V}^{-1}\mathbf{e} \end{array} \right),$$

with columns weighted by the Lagrange multipliers corresponding to the two constraints. We will call the Lagrange multipliers $2\gamma/C$ and $2\lambda/A$ respectively, where we define:

$$A \equiv \mathbf{1}^\top\mathbf{V}^{-1}\mathbf{e} = \mathbf{e}^\top\mathbf{V}^{-1}\mathbf{1} \tag{6.4.9}$$

$$B \equiv \mathbf{e}^\top\mathbf{V}^{-1}\mathbf{e} > 0 \tag{6.4.10}$$

$$C \equiv \mathbf{1}^\top\mathbf{V}^{-1}\mathbf{1} > 0 \tag{6.4.11}$$

and

$$D \equiv BC - A^2 \tag{6.4.12}$$

and the inequalities follow from the fact that $\mathbf{V}^{-1}$ (like $\mathbf{V}$) is positive definite.
This allows the solution to be written as:

$$\mathbf{w} = \frac{\gamma}{C}(\mathbf{V}^{-1}\mathbf{1}) + \frac{\lambda}{A}(\mathbf{V}^{-1}\mathbf{e}). \tag{6.4.13}$$

$\frac{1}{C}(\mathbf{V}^{-1}\mathbf{1})$ and $\frac{1}{A}(\mathbf{V}^{-1}\mathbf{e})$ are both unit portfolios, so $\gamma + \lambda = W_0$. We know that for the portfolio which minimises variance for a given initial wealth, regardless of expected final wealth, the corresponding Lagrange multiplier, $\lambda = 0$. Thus $\frac{\gamma}{C}(\mathbf{V}^{-1}\mathbf{1})$ is the global minimum variance portfolio with cost W_0 (which in fact

equals γ in this case) and $\frac{1}{C}(\mathbf{V}^{-1}\mathbf{1})$ is the global minimum variance unit cost portfolio, which we will denote $\mathbf{w}_{\mathrm{MVP}}$.
In fact, we can combine (6.4.7) and (6.4.13) and write the solution as:

$$\mathbf{w} = W_0 \left(\mathbf{w}_{\mathrm{MVP}} + \left(\mu - \frac{A}{C} \right) \mathbf{h} \right). \tag{6.4.14}$$

The details are left as an exercise.[4]
The set of solutions to this quadratic programming problem for all possible (W_0, W_1) combinations (including negative W_0) is the vector subspace of the portfolio space, which is generated either by the vectors $\mathbf{g}$ and $\mathbf{h}$ or by the vectors $\mathbf{w}_{\mathrm{MVP}}$ and $\mathbf{h}$ (or by any pair of linearly independent frontier portfolios).
In $\Re^N$, the set of unit cost frontier portfolios is the line passing thru $\mathbf{g}$, parallel to $\mathbf{h}$.
It follows immediately that the frontier (like any straight line in $\Re^N$) is a convex set, and can be generated by linear combinations of any pair of frontier portfolios with weights of the form α and $(1 - \alpha)$.
An important **exercise** at this stage is to work out the means, variances and co-variances of the returns on $\mathbf{w}_{\mathrm{MVP}}$, $\mathbf{g}$ and $\mathbf{h}$. They will drop out of the portfolio decomposition below.
The portfolio weight vectors $\mathbf{g}$ and $\mathbf{h}$ are

$$\mathbf{g} = \frac{1}{D}[B(\mathbf{V}^{-1}\mathbf{1}) - A(\mathbf{V}^{-1}\mathbf{e})] \tag{6.4.15}$$

$$\mathbf{h} = \frac{1}{D}[C(\mathbf{V}^{-1}\mathbf{e}) - A(\mathbf{V}^{-1}\mathbf{1})] \tag{6.4.16}$$

We have

$$\mathrm{Var}[\tilde{r}_{\mathbf{g}}] = \mathbf{g}^\top \mathbf{V} \mathbf{g} = \frac{B}{D} \tag{6.4.17}$$

$$\mathrm{Var}[\tilde{r}_{\mathbf{h}}] = \mathbf{h}^\top \mathbf{V} \mathbf{h} = \frac{C}{D} \tag{6.4.18}$$

from which it follows that $D > 0$.

Orthogonal decomposition of portfolios

At this stage, we must introduce a scalar product on the portfolio space, namely that based on the variance-covariance matrix $\mathbf{V}$. Since $\mathbf{V}$ is a non-singular, positive definite matrix, it defines a well behaved scalar product and all the standard results on orthogonal projection (*&c.*) from linear algebra are valid.

[4] At least for now.

Two portfolios $\mathbf{w}_1$ and $\mathbf{w}_2$ are orthogonal with respect to this scalar product
$\Longleftrightarrow$
$\mathbf{w}_1^\top \mathbf{V}\mathbf{w}_2 = 0$
$\Longleftrightarrow$
$\mathrm{Cov}\left[\mathbf{w}_1^\top \tilde{\mathbf{r}}, \mathbf{w}_2^\top \tilde{\mathbf{r}}\right] = 0$
$\Longleftrightarrow$
the random variables representing the returns on the portfolios are uncorrelated.
Thus, the terms 'orthogonal' and 'uncorrelated' may legitimately, and shall, be applied interchangeably to pairs of portfolios. Furthermore, the squared length of a weight vector corresponds to the variance of its returns.
Note that $\mathbf{w}_{\mathrm{MVP}}$ and $\mathbf{h}$ are orthogonal vectors in this sense. In fact, we have the following theorem:

Theorem 6.4.1 *If* $\mathbf{w}$ *is a frontier portfolio and* $\mathbf{u}$ *is a zero mean hedge portfolio, then* $\mathbf{w}$ *and* $\mathbf{u}$ *are uncorrelated.*

Proof There is probably a full version of this proof lost somewhere but the following can be sorted out.
Since $\mathbf{w}_{\mathrm{MVP}}$ is collinear with $\mathbf{V}^{-1}\mathbf{1}$, it is orthogonal to all portfolios $\mathbf{w}$ for which $\mathbf{w}^\top \mathbf{V}\mathbf{V}^{-1}\mathbf{1} = 0$ or in other words to all portfolios for which $\mathbf{w}^\top \mathbf{1} = 0$. But these are precisely all hedge portfolios, including $\mathbf{h}$.
Similarly, any portfolio collinear with $\mathbf{V}^{-1}\mathbf{e}$ is orthogonal to all portfolios with zero expected return, since $\mathbf{w}^\top \mathbf{V}\mathbf{V}^{-1}\mathbf{e} = 0$ or in other words $\mathbf{w}^\top \mathbf{e} = 0$.

Q.E.D.

Some pictures are in order at this stage.
For $N = 3$, in the set of portfolios costing W_0 (the W_0 simplex), the iso-variance curves are concentric ellipses, the iso-mean curves are parallel lines, and the solutions for different μs (or W_1s) are the tangency points between these ellipses and lines, which themselves lie on a line orthogonal (in the sense defined above) to the iso-mean lines. The centre of the concentric ellipses is at the global minimum variance portfolio corresponding to W_0, $W_0\mathbf{w}_{\mathrm{MVP}}$. A similar geometric interpretation can be applied in higher dimensions.
? has some nice pictures of the frontier in portfolio space, as opposed to mean-variance space.
At this stage, recall the definition of β in (5.2.2).
We will now derive an orthogonal decomposition of a portfolio $\mathbf{q}$ into two frontier portfolios and a zero-mean zero-cost portfolio and prove that the coefficients on the two frontier portfolios are the βs of $\mathbf{q}$ with respect to those portfolios and sum to unity.
We can always choose an orthogonal basis for the portfolio frontier.

For any frontier portfolio $\mathbf{p} \neq \mathbf{w}_{\mathrm{MVP}}$, there is a unique unit cost frontier portfolio $\mathbf{z_p}$ which is orthogonal to $\mathbf{p}$.
Another important **exercise** is to figure out the relationship between $E\left[\tilde{r}_{\mathbf{p}}\right]$ and $E\left[\tilde{r}_{\mathbf{z_p}}\right]$.
Any two frontier portfolios span the frontier, in particular any unit cost $\mathbf{p} \neq \mathbf{w}_{\mathrm{MVP}}$ and $\mathbf{z_p}$ (or the original basis, $\mathbf{w}_{\mathrm{MVP}}$ and $\mathbf{h}$).
Any (frontier or non-frontier) portfolio $\mathbf{q}$ with non-zero cost W_0 can be written in the form $\mathbf{f_q} + \mathbf{u_q}$ where

$$\mathbf{f_q} \equiv W_0\left(\mathbf{g} + E[\tilde{r}_{\mathbf{q}}]\mathbf{h}\right) \tag{6.4.19}$$

$$= W_0\left(\beta_{\mathbf{qp}}\mathbf{p} + (1-\beta_{\mathbf{qp}})\mathbf{z_p}\right) \text{ (say)} \tag{6.4.20}$$

is the frontier portfolio with expected return $E[\tilde{r}_{\mathbf{q}}]$ and cost W_0 and $\mathbf{u_q}$ is a hedge portfolio with zero expected return. Geometrically, this decomposition is equivalent to the orthogonal projection of $\mathbf{q}$ onto the frontier.
Theorem 6.4.1 shown that any portfolio sharing these properties of $\mathbf{u_q}$ is uncorrelated with all frontier portfolios.[5]
If $\mathbf{p}$ is a unit cost frontier portfolio (*i.e.* the vector of portfolio proportions) and $\mathbf{q}$ is an arbitrary unit cost portfolio, then the following decomposition therefore holds:

$$\mathbf{q} = \mathbf{f_q} + \mathbf{u_q} = \beta_{\mathbf{qp}}\mathbf{p} + \left(1-\beta_{\mathbf{qp}}\right)\mathbf{z_p} + \mathbf{u_q} \tag{6.4.21}$$

where the three components (*i.e.* the vectors $\mathbf{p}$, $\mathbf{z_p}$ and $\mathbf{u_q}$) are mutually orthogonal.
We can extend this decomposition to cover

1. portfolio proportions (orthogonal vectors)
2. portfolio proportions (scalars/components)
3. returns (uncorrelated random variables)
4. expected returns (numbers)

Note again the parallel between orthogonal portfolio vectors and uncorrelated portfolio returns/payoffs.
We will now derive the relation:

$$E[\tilde{r}_{\mathbf{q}}] - E[\tilde{r}_{\mathbf{z_p}}] = \beta_{\mathbf{qp}}(E[\tilde{r}_{\mathbf{p}}] - E[\tilde{r}_{\mathbf{z_p}}]) \tag{6.4.22}$$

[5] Aside: For the frontier portfolio $\mathbf{f_q}$ to second degree stochastically dominate the arbitrary portfolio q, we will need zero conditional expected return on $\mathbf{u_q}$, and will have to show that

$$\mathrm{Cov}\left[\tilde{r}_{\mathbf{u_q}}, \tilde{r}_{\mathbf{f_q}}\right] = 0 \implies E[\tilde{r}_{\mathbf{u_q}}|\tilde{r}_{\mathbf{f_q}}] = 0$$

The normal distribution is the only case where this is true.

which may be familiar from earlier courses in financial economics and which is quite general and neither requires asset returns to be normally distributed nor any assumptions about preferences.

Since $\mathrm{Cov}\left[\tilde{r}_{\mathbf{u_q}}, \tilde{r}_{\mathbf{p}}\right] = \mathrm{Cov}\left[\tilde{r}_{\mathbf{z_p}}, \tilde{r}_{\mathbf{p}}\right] = 0$, taking covariances with $\tilde{r}_{\mathbf{p}}$ in (6.4.21) gives:

$$\mathrm{Cov}\left[\tilde{r}_{\mathbf{q}}, \tilde{r}_{\mathbf{p}}\right] = \mathrm{Cov}\left[\tilde{r}_{\mathbf{f_q}}, \tilde{r}_{\mathbf{p}}\right] = \beta_{\mathbf{qp}}\mathrm{Var}[\tilde{r}_{\mathbf{p}}] \tag{6.4.23}$$

or

$$\beta_{\mathbf{qp}} = \frac{\mathrm{Cov}\left[\tilde{r}_{\mathbf{q}}, \tilde{r}_{\mathbf{p}}\right]}{\mathrm{Var}[\tilde{r}_{\mathbf{p}}]} \tag{6.4.24}$$

Thus β in (6.4.21) has its usual definition from probability theory, given by (5.2.2).[6]

Reversing the roles of $\mathbf{p}$ and $\mathbf{z_p}$, it can be seen that

$$\beta_{\mathbf{qz_p}} = 1 - \beta_{\mathbf{qp}} \tag{6.4.25}$$

Taking expected returns in (6.4.21) yields again:

$$E[\tilde{r}_{\mathbf{q}}] = \beta_{\mathbf{qp}}E[\tilde{r}_{\mathbf{p}}] + (1 - \beta_{\mathbf{qp}})E[\tilde{r}_{\mathbf{z_p}}], \tag{6.4.26}$$

which can be rearranged to obtain (6.4.22).

The Global Minimum Variance Portfolio

$$\mathrm{Var}[\tilde{r}_{\mathbf{g}+\mu\mathbf{h}}] = \mathbf{g}^\top\mathbf{V}\mathbf{g} + 2\mu(\mathbf{g}^\top\mathbf{V}\mathbf{h}) + \mu^2(\mathbf{h}^\top\mathbf{V}\mathbf{h}) \tag{6.4.27}$$

which has its minimum at

$$\mu = -\frac{\mathbf{g}^\top\mathbf{V}\mathbf{h}}{\mathbf{h}^\top\mathbf{V}\mathbf{h}} \tag{6.4.28}$$

The latter expression reduces to A/C and the minimum value of the variance is $1/C$. The global minimum variance portfolio is denoted MVP.

$$\mathrm{Cov}\left[\tilde{r}_{\mathbf{h}}, \tilde{r}_{\mathrm{MVP}}\right] = \mathbf{h}^\top\mathbf{V}\left(\mathbf{g} - \frac{\mathbf{g}^\top\mathbf{V}\mathbf{h}}{\mathbf{h}^\top\mathbf{V}\mathbf{h}}\mathbf{h}\right) \tag{6.4.29}$$

$$= \mathbf{h}^\top\mathbf{V}\mathbf{g} - \frac{\mathbf{g}^\top\mathbf{V}\mathbf{h}.\mathbf{h}^\top\mathbf{V}\mathbf{h}}{\mathbf{h}^\top\mathbf{V}\mathbf{h}} = 0 \tag{6.4.30}$$

i.e. the returns on the portfolio with weights $\mathbf{h}$ and the minimum variance portfolio are uncorrelated.

[6] Assign some problems involving the construction of portfolio proportions for various desired βs. Also problems working from prices for state contingent claims to returns on assets and portfolios in both single period and multi-period worlds.

Further, if p is any portfolio, the MVP is the minimum variance combination of itself and p, *i.e.* $a = 0$ solves:

$$\min_a \frac{1}{2}\mathrm{Var}[\tilde{r}_{ap+(1-a)\mathrm{MVP}}] \tag{6.4.31}$$

which has necessary and sufficient first order condition:

$$a\mathrm{Var}[\tilde{r}_{\mathbf{p}}] + (1-2a)\mathrm{Cov}\,[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathrm{MVP}}] - (1-a)\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] = 0 \tag{6.4.32}$$

Hence, setting $a = 0$:

$$\mathrm{Cov}\,[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathrm{MVP}}] - \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] = 0 \tag{6.4.33}$$

and the covariance of any portfolio with MVP is $1/C$.

6.4.2 The portfolio frontier in mean-variance space: risky assets only

The portfolio frontier in mean-variance and in mean-standard deviation space

We now move on to consider the mean-variance relationship along the portfolio frontier.
The mean, μ, and variance, σ^2, of the rate of return associated with each point on the frontier are related by the quadratic equation:

$$(\sigma^2 - \mathrm{Var}[\mathbf{w}_{\mathrm{MVP}}^{\top}\tilde{\mathbf{r}}]) = \phi(\mu - \mathrm{E}[\mathbf{w}_{\mathrm{MVP}}^{\top}\tilde{\mathbf{r}}])^2, \tag{6.4.34}$$

where the shape parameter $\phi = C/D$ represents the variance of the (gross) return on the hedge portfolio, $\mathbf{h}$. The two-moment frontier is generally presented as the graph in mean-variance space of this parabola, showing the most desirable distributions attainable, but the frontier can also be thought of as a plane in portfolio space or as a line in portfolio weight space. The latter interpretations are far more useful when it comes to extending the analysis to higher moments.
The equations of the frontier in mean-variance and mean-standard deviation space can be derived heuristically using the following stylized diagram illustrating the portfolio decomposition.

Figure 3A goes here.

Applying Pythagoras' theorem to the triangle with vertices at 0, p and MVP yields:

$$\sigma^2 = \mathrm{Var}[\tilde{r}_{\mathbf{p}}] = \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] + \left(\mu - \frac{A}{C}\right)^2 \mathrm{Var}[\tilde{r}_{\mathbf{h}}] \tag{6.4.35}$$

Recall from the coordinate geometry of conic sections that

$$\mathrm{Var}[\tilde{r}_{\mathbf{p}}] = \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] + (\mu - E[\tilde{r}_{\mathrm{MVP}}])^2 \, \mathrm{Var}[\tilde{r}_{\mathbf{h}}] \tag{6.4.36}$$

or

$$V(\mu) = \frac{1}{C} + \frac{C}{D}\left(\mu - \frac{A}{C}\right)^2 \tag{6.4.37}$$

is a quadratic equation in μ.
i.e. the equation of the parabola with vertex at

$$\mathrm{Var}[\tilde{r}_{\mathbf{p}}] = \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] = \frac{1}{C} \tag{6.4.38}$$

$$\mu = E[\tilde{r}_{\mathrm{MVP}}] = \frac{A}{C} \tag{6.4.39}$$

Thus in mean-variance space, the frontier is a parabola.

Figure 3.11.2 goes here: indicate position of **g** on figure.

Similarly, in mean-standard deviation space, the frontier is a hyperbola. To see this, recall that:

$$\sigma^2 = \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] + \left(\mu - \frac{A}{C}\right)^2 \mathrm{Var}[\tilde{r}_{\mathbf{h}}] \tag{6.4.40}$$

is the equation of the hyperbola with vertex at

$$\sigma = \sqrt{\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}]} = \sqrt{\frac{1}{C}} \tag{6.4.41}$$

$$\mu = \frac{A}{C} \tag{6.4.42}$$

centre at $\sigma = 0$, $\mu = A/C$ and asymptotes as indicated.

Figure 3.11.1 goes here: indicate position of **g** on figure.

The other half of the hyperbola ($\sigma < 0$) has no economic meaning.
Recall two other types of conic sections:
$\mathrm{Var}[\tilde{r}_{\mathbf{h}}] < 0$ (impossible) gives a circle with center $(1/C, A/C)$.
$\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] = 0$ (the presence of a riskless asset) allows the square root to be taken on both sides:

$$\sigma = \pm\left(\mu - \frac{A}{C}\right)\sqrt{\mathrm{Var}[\tilde{r}_{\mathbf{h}}]} \tag{6.4.43}$$

i.e. the conic section becomes the pair of lines which are its asymptotes otherwise.

Portfolios on which the expected return, μ, exceeds $\mathbf{w}_{\mathrm{MVP}}^{\top}\mathbf{e}$ are termed *efficient*, since they maximise expected return given variance; other frontier portfolios minimise expected return given variance and are *inefficient*.
A frontier portfolio is said to be an efficient portfolio iff
its expected return exceeds the minimum variance expected return $A/C = E[\tilde{r}_{\mathrm{MVP}}]$.
The set of efficient portfolios in $\Re^N$ (or **efficient frontier**) is the half-line emanating from MVP in the direction of $\mathbf{h}$, and hence is also a convex set.
Convex combinations (but not all linear combinations with weights summing to 1) of efficient portfolios are efficient.
We now consider zero-covariance (zero-beta) portfolios. In portfolio weight space, can easily construct a frontier portfolio having zero covariance with any given frontier portfolio:

Figure 3B goes here.

Algebraically, the expected return μ_0 on the zero-covariance frontier portfolio of a frontier portfolio with expected return μ solves:

$$\mathrm{Cov}\left[\tilde{r}_{\mathrm{MVP}} + (\mu - E[\tilde{r}_{\mathrm{MVP}}])\tilde{r}_{\mathbf{h}}, \tilde{r}_{\mathrm{MVP}} + (\mu_0 - E[\tilde{r}_{\mathrm{MVP}}])\tilde{r}_{\mathbf{h}}\right] = 0 \tag{6.4.44}$$

or, since $\tilde{r}_{\mathbf{h}}$ and $\tilde{r}_{\mathrm{MVP}}$ are uncorrelated:

$$\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}] + (\mu - E[\tilde{r}_{\mathrm{MVP}}])(\mu_0 - E[\tilde{r}_{\mathrm{MVP}}])\mathrm{Var}[\tilde{r}_{\mathbf{h}}] = 0 \tag{6.4.45}$$

To make this true, we must have

$$(\mu - E[\tilde{r}_{\mathrm{MVP}}])(\mu_0 - E[\tilde{r}_{\mathrm{MVP}}]) < 0 \tag{6.4.46}$$

or μ and μ_0 on opposite sides of $E[\tilde{r}_{\mathrm{MVP}}]$ as shown.
There is a neat trick which allows zero-covariance portfolios to be plotted in mean-standard deviation space.
Implicit differentiation of the $\mu - \sigma$ relationship $(6.4.35)$ along the frontier yields:

$$\frac{d\mu}{d\sigma} = \frac{\sigma}{(\mu - E[\tilde{r}_{\mathrm{MVP}}])\mathrm{Var}[\tilde{r}_{\mathbf{h}}]} \tag{6.4.47}$$

so the tangent at (σ, μ) intercepts the μ axis at

$$\begin{aligned}
\mu - \sigma\frac{d\mu}{d\sigma} &= \mu - \frac{\sigma^2}{(\mu - E[\tilde{r}_{\mathrm{MVP}}])\mathrm{Var}[\tilde{r}_{\mathbf{h}}]} && (6.4.48)\\
&= \mu - \frac{\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}]}{(\mu - E[\tilde{r}_{\mathrm{MVP}}])\mathrm{Var}[\tilde{r}_{\mathbf{h}}]} - (\mu - E[\tilde{r}_{\mathrm{MVP}}]) && (6.4.49)\\
&= E[\tilde{r}_{\mathrm{MVP}}] - \frac{\mathrm{Var}[\tilde{r}_{\mathrm{MVP}}]}{(\mu - E[\tilde{r}_{\mathrm{MVP}}])\mathrm{Var}[\tilde{r}_{\mathbf{h}}]} && (6.4.50)
\end{aligned}$$

where we substituted for σ^2 from the definition of the frontier.
A little rearrangement shows that expression $(6.4.50)$ satisfies the equation $(6.4.45)$ defining the return on the zero-covariance portfolio.
In mean-standard deviation space the picture is like this:

Figure 3.15.1 goes here.

To find $\mathbf{z_p}$ in mean-variance space, note that the line joining (σ^2, μ) to the MVP intercepts the μ axis at:

$$\mu - \sigma^2 \frac{\mu - E[\tilde{r}_{\mathrm{MVP}}]}{\sigma^2 - \mathrm{Var}[\tilde{r}_{\mathrm{MVP}}]} = \mu - \sigma^2 \frac{\mu - E[\tilde{r}_{\mathrm{MVP}}]}{(\mu - E[\tilde{r}_{\mathrm{MVP}})^2 \mathrm{Var}[\tilde{r}_{\mathbf{h}}]} \tag{6.4.51}$$

After cancellation, this is exactly the first expression $(6.4.48)$ for the zero-covariance return we had on the previous page.

Figure 3.15.2 goes here.

Alternative derivations

The treatment of the portfolio frontier with risky assets only concludes with some alternative derivations following closely **?**. They should probably be omitted altogether at this stage.

1. The variance minimisation solution from first principles.

   It can be seen that $\mathbf{w}$ is the solution to:

$$\min_{\{\mathbf{w}, \ , \ \}} L = \frac{1}{2}\mathbf{w}^\top \mathbf{V}\mathbf{w} + \lambda(\mu - \mathbf{w}^\top \mathbf{e}) + \gamma(W_0 - \mathbf{w}^\top \mathbf{1}) \tag{6.4.52}$$

   which has necessary and sufficient first order conditions:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{V}\mathbf{w} - \lambda \mathbf{e} - \gamma \mathbf{1} = \mathbf{0} \tag{6.4.53}$$

$$\frac{\partial L}{\partial \lambda} = \mu - \mathbf{w}^\top \mathbf{e} = 0 \tag{6.4.54}$$

$$\frac{\partial L}{\partial \gamma} = W_0 - \mathbf{w}^\top \mathbf{1} = 0 \tag{6.4.55}$$

   The solution can be found by premultiplying the FOC (6.4.13) in turn by $\mathbf{e}^\top$ and $\mathbf{1}^\top$ and using the constraints yields:

$$\mu = \lambda(\mathbf{e}^\top \mathbf{V}^{-1}\mathbf{e}) + \gamma(\mathbf{e}^\top \mathbf{V}^{-1}\mathbf{1}) \tag{6.4.56}$$

$$1 = \lambda(\mathbf{1}^\top \mathbf{V}^{-1}\mathbf{e}) + \gamma(\mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1}) \tag{6.4.57}$$

The solutions for λ and γ are:

$$\lambda = \frac{C\mu - A}{D} \tag{6.4.58}$$

$$\gamma = \frac{B - A\mu}{D} \tag{6.4.59}$$

2. Derivation of (6.4.22).

   If we only have frontier portfolio p and interior portfolio q, we get a frontier (in μ-σ space) entirely within the previous frontier and tangent to it at p.

   The frontiers must have the same slope at p:

   Figure 3C goes here.

   We already saw that the outer frontier has slope $\frac{E[\tilde{r}_{\mathbf{p}} - \tilde{r}_{\mathbf{z_p}}]}{\sqrt{\mathrm{Var}[\tilde{r}_{\mathbf{p}}]}}$.

   At the point on the inner frontier with $w_{\mathbf{q}}$ invested in q and $(1 - w_{\mathbf{q}})$ in p,

$$\mu = E[\tilde{r}_{\mathbf{p}}] + w_{\mathbf{q}}(E[\tilde{r}_{\mathbf{q}} - \tilde{r}_{\mathbf{p}}]) \tag{6.4.60}$$

$$\sigma^2 = w_{\mathbf{q}}^2 \mathrm{Var}[\tilde{r}_{\mathbf{q}}] + 2w_{\mathbf{q}}(1 - w_{\mathbf{q}})\mathrm{Cov}\,[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathbf{q}}] + (1 - w_{\mathbf{q}})^2 \mathrm{Var}[\tilde{r}_{\mathbf{p}}] \tag{6.4.61}$$

   Differentiating these w.r.t. $w_{\mathbf{q}}$:

$$\frac{d\mu}{dw_{\mathbf{q}}} = E[\tilde{r}_{\mathbf{q}} - \tilde{r}_{\mathbf{p}}] \tag{6.4.62}$$

$$2\sigma\frac{d\sigma}{dw_{\mathbf{q}}} = 2w_{\mathbf{q}}\mathrm{Var}[\tilde{r}_{\mathbf{q}}] + 2(1 - 2w_{\mathbf{q}})\mathrm{Cov}\,[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathbf{q}}] - 2(1 - w_{\mathbf{q}})\mathrm{Var}[\tilde{r}_{\mathbf{p}}] \tag{6.4.63}$$

   Taking the ratio and setting $w_{\mathbf{q}} = 0$ gives the slope of the inner frontier at p:

$$\frac{d\mu}{d\sigma} = \frac{E[\tilde{r}_{\mathbf{q}} - \tilde{r}_{\mathbf{p}}]}{\frac{2\mathrm{Cov}[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathbf{q}}] - 2\mathrm{Var}[\tilde{r}_{\mathbf{p}}]}{2\sqrt{\mathrm{Var}[\tilde{r}_{\mathbf{p}}]}}} \tag{6.4.64}$$

   Equating this to the slope of the outer frontier, setting

$$\beta_{\mathbf{qp}} = \frac{\mathrm{Cov}\,[\tilde{r}_{\mathbf{p}}, \tilde{r}_{\mathbf{q}}]}{\mathrm{Var}[\tilde{r}_{\mathbf{p}}]} \tag{6.4.65}$$

   and rearranging yields:

$$E[\tilde{r}_{\mathbf{q}}] - E[\tilde{r}_{\mathbf{z_p}}] = \beta_{\mathbf{qp}}(E[\tilde{r}_{\mathbf{p}}] - E[\tilde{r}_{\mathbf{z_p}}]) \tag{6.4.66}$$

6.4.3 The portfolio frontier in $\Re^N$: riskfree and risky assets

We now consider the mathematics of the portfolio frontier when there is a riskfree asset.
In this case, the frontier portfolio solves:

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top \mathbf{V}\mathbf{w} \tag{6.4.67}$$

$$\text{s.t.} \qquad \mathbf{w}^\top \mathbf{e} + (1 - \mathbf{w}^\top \mathbf{1})r_f = \mu \tag{6.4.68}$$

There is no longer a restriction on portfolio weights, and whatever is not invested in the N risky assets is assumed to be invested in the riskless asset.
The solution (which can be left as an exercise) is by a similar method to the case where all assets were risky:

$$\mathbf{w}_p = \mathbf{V}^{-1}(\mathbf{e} - r_f\mathbf{1})\frac{\mu - r_f}{H} \tag{6.4.69}$$

where

$$H = (\mathbf{e} - \mathbf{1}r_f)^\top \mathbf{V}^{-1}(\mathbf{e} - \mathbf{1}r_f) = B - 2Ar_f + C{r_f}^2 > 0\ \forall r_f \tag{6.4.70}$$

Along the frontier, we have:

$$\sigma = \begin{cases} \frac{\mu - r_f}{\sqrt{H}} & \text{if } \mu \geq r_f, \\ -\frac{\mu - r_f}{\sqrt{H}} & \text{if } \mu < r_f, \end{cases} \tag{6.4.71}$$

6.4.4 The portfolio frontier in mean-variance space: riskfree and risky assets

We can now establish the shape of the mean-standard deviation frontier with a riskless asset.
Graphically, in mean-standard deviation space, combining any portfolio $\mathbf{p}$ with the riskless asset in proportions a and $(1-a)$ gives a portfolio with expected return

$$aE[\tilde{r}_{\mathbf{p}}] + (1-a)r_f = r_f + a(E[\tilde{r}_{\mathbf{p}}] - r_f)$$

and standard deviation of returns $a\sqrt{\text{Var}[\tilde{r}_{\mathbf{p}}]}$.
i.e. these portfolios trace out the ray in σ-μ space emanating from $(0, r_f)$ and passing through p.
For each σ the highest return attainable is along the ray from r_f which is tangent to the frontier generated by the risky assets.

On this ray, the riskless asset is held in combination with the tangency portfolio $\mathbf{t}$.
This only makes sense for $r_f < A/C = E[\tilde{r}_{mvp}]$.
Above $\mathbf{t}$, there is a negative weight on the riskless asset — *i.e.* borrowing.

Figure 3D goes here.

Limited borrowing

Unlimited borrowing as allowed in the preceding analysis is unrealistic.
Consider what happens

1. with margin constraints on borrowing:

   Figure 3E goes here.

   The frontier is the envelope of all the finite rays through risky portfolios, extending as far as the borrowing constraint allows.

2. with differential borrowing and lending rates:

   Figure 3F goes here.

   There is a range of expected returns over which a pure risky strategy provides minimum variance;

   lower expected returns are achieved by riskless lending;

   and higher expected returns are achieved by riskless borrowing.

6.5 Market Equilibrium and the Capital Asset Pricing Model

6.5.1 Pricing assets and predicting security returns

Need more waffle here about prediction and the difficulties thereof and the properties of equilibrium prices and returns.
We are looking for assumptions concerning probability distributions that lead to useful and parsimonious asset pricing models. The CAPM restrictions are the best known. At a very basic level, they can be expressed by saying that every

investor has mean-variance preferences. This can be achieved either by restricting preferences to be quadratic or the probability distribution of asset returns to be normal. CAPM is basically a single-period model, but can be extended by assuming that return distributions are stable over time. **?** and **?** have generalised the distributional conditions.
Recall also the limiting behaviour of the variance of the return on an equally weighted portfolio as the number of securities included goes to infinity. If securities are added in such a way that the average of the variance terms and the average of the covariance terms are stable, then the portfolio variance approaches the average covariance as a lower bound.

6.5.2 Properties of the market portfolio

Let

m_j	=	weight of security j in the market portfolio $\mathbf{m}$
$W_0^i(>0)$	=	individual i's initial wealth
w_{ij}	=	proportion of individual i's initial wealth invested in j-th security

Then total wealth is defined by

$$W_{m0} \equiv \sum_{i=1}^{I} W_0^i \tag{6.5.1}$$

and in equilibrium the relation

$$\sum_{i=1}^{I} w_{ij} W_0^i = m_j W_{m0} \quad \forall j \tag{6.5.2}$$

must hold. Dividing by W_{m0} yields:

$$\sum_{i=1}^{I} w_{ij} \frac{W_0^i}{W_{m0}} = m_j \quad \forall j \tag{6.5.3}$$

and thus in equilibrium the market portfolio is a convex combination of individual portfolios.

6.5.3 The zero-beta CAPM

Theorem 6.5.1 (Zero-beta CAPM theorem) *If every investor holds a mean-variance frontier portfolio, then the market portfolio, $\mathbf{m}$, is a mean-variance frontier portfolio, and hence, $\forall \mathbf{q}$, the* CAPM equation

$$E\left[\tilde{r}_{\mathbf{q}}\right] = \left(1 - \beta_{\mathbf{qm}}\right) E\left[\tilde{r}_{\mathbf{z_m}}\right] + \beta_{\mathbf{qm}} E\left[\tilde{r}_{\mathbf{m}}\right] \tag{6.5.4}$$

holds.

Theorem 6.5.2 *All strictly risk-averse investors hold frontier portfolios if and only if*

$$E\left[\tilde{r}_{\mathbf{u_q}}|\tilde{r}_{\mathbf{f_q}}\right] = 0 \quad \forall \mathbf{q} \tag{6.5.5}$$

Note the subtle distinction between uncorrelated returns (in the definition of the decomposition) and independent returns (in this theorem). They are the same only for the normal distribution and related distributions.
We can view the market portfolio as a frontier portfolio under two fund separation.
If $\mathbf{p}$ is a **frontier** portfolio, then we showed earlier that for purely mathematical reasons in the definition of a frontier portfolio:

$$E[\tilde{r}_{\mathbf{q}}] = (1 - \beta_{\mathbf{qp}})E[\tilde{r}_{\mathbf{z_p}}] + \beta_{\mathbf{qp}}E[\tilde{r}_{\mathbf{p}}] \tag{6.5.6}$$

If two fund separation holds, then individuals hold frontier portfolios.
Since the market portfolio is then on the frontier, it follows that:

$$E[\tilde{r}_{\mathbf{q}}] = (1 - \beta_{\mathbf{qm}})E[\tilde{r}_{\mathbf{z_m}}] + \beta_{\mathbf{qm}}E[\tilde{r}_{\mathbf{m}}] \tag{6.5.7}$$

where

$$\tilde{r}_{\mathbf{m}} = \sum_{j=1}^{N} m_j \tilde{r}_j \tag{6.5.8}$$

$$\beta_{\mathbf{qm}} = \frac{\mathrm{Cov}\left[\tilde{r}_{\mathbf{q}}, \tilde{r}_{\mathbf{m}}\right]}{\mathrm{Var}[\tilde{r}_{\mathbf{m}}]} \tag{6.5.9}$$

This implies for any particular security, from the economic assumptions of equilibrium and two fund separation:

$$E[\tilde{r}_j] = (1 - \beta_{j\mathbf{m}})E[\tilde{r}_{\mathbf{z_m}}] + \beta_{j\mathbf{m}}E[\tilde{r}_{\mathbf{m}}] \tag{6.5.10}$$

This relation is the **?** Zero-Beta version of the Capital Asset Pricing Model (CAPM).

6.5.4 The traditional CAPM

Now we add the risk free asset, which will allow us to determine the tangency portfolio, $\mathbf{t}$, and to talk about Capital Market Line (return *v.* standard deviation) and the Security Market Line (return *v.* β).
Normally in equilibrium there is zero aggregate supply of the riskfree asset.
Recommended reading for this part of the course is **?**, **?**, **?** and **?**.
Now we can derive the traditional CAPM. Note that by construction

$$r_f = E\left[\tilde{r}_{\mathbf{z_t}}\right]. \tag{6.5.11}$$

Theorem 6.5.3 (Separation Theorem) *The risky asset holdings of all investors who hold mean-variance frontier portfolios are in the proportions given by the tangency portfolio,* $\mathbf{t}$.

Theorem 6.5.4 (Traditional CAPM Theorem) *If every investor holds a mean-variance frontier portfolio, then the market portfolio of risky assets,* $\mathbf{m}$, *is the tangency portfolio,* $\mathbf{t}$, *and hence,* $\forall \mathbf{q}$, *the* traditional CAPM equation

$$E\left[\tilde{r}_{\mathbf{q}}\right] = \left(1 - \beta_{\mathbf{qm}}\right) r_f + \beta_{\mathbf{qm}} E\left[\tilde{r}_{\mathbf{m}}\right] \tag{6.5.12}$$

holds.

Theorem 6.5.4 is sometimes known as the Sharpe-Lintner Theorem.
The riskless rate is unique by the No Arbitrage Principle, since otherwise a greedy investor would borrow an infinite amount at the lower rate and invest it at the higher rate, which is impossible in equilibrium.
We can also think about what happens the CAPM if there are different riskless borrowing and lending rates (see **?**). If all individuals face this situation in equilibrium, realism demands that both riskless assets are in zero aggregate supply and hence that all investors hold risky assets only.
Note that the No Arbitrage Principle also allows us to rule out correlation matrices for risky assets which permit the construction of portfolios with zero return variance, *i.e.* synthetic riskless assets.
Assume that a riskless asset exists, with return $r_f < E[\tilde{r}_{\mathtt{mvp}}]$.
If the distributional conditions for two fund separation are satisfied, then the tangency portfolio, $\mathbf{t}$, must be the market portfolio of risky assets in equilibrium. We know then that for any portfolio $\mathbf{q}$ (with or without a riskless component):

$$E[\tilde{r}_{\mathbf{q}}] - r_f = \beta_{\mathbf{qm}}(E[\tilde{r}_{\mathbf{m}}] - r_f) \tag{6.5.13}$$

This is the traditional Sharpe-Lintner version of the CAPM.
Figure 4A goes here.
The next theorem relates to the mean-variance efficiency of the market portfolio.

Theorem 6.5.5 *If*

1. *the distributional conditions for two fund separation are satisfied;*
2. *risky assets are in strictly positive supply; and*
3. *investors have strictly increasing (concave) utility functions*

then the market/tangency portfolio is efficient.

Proof By Jensen's inequality and monotonicity, the riskless asset dominates any portfolio with

$$E[\tilde{r}] < r_f \tag{6.5.14}$$

for

$$\begin{aligned} E[u(W_0(1+\tilde{r}))] &\leq u(E[W_0(1+\tilde{r})]) && (6.5.15) \\ &< u(W_0(1+r_f)) && (6.5.16) \end{aligned}$$

Hence the expected returns on all individuals' portfolios exceed r_f.
It follows that the expected return on the market portfolio must exceed r_f.
Q.E.D.

Now we can calculate the risk premium of the market portfolio.
CAPM gives a relation between the risk premia on individual assets and the risk premium on the market portfolio.
The risk premium on the market portfolio must adjust in equilibrium to give market-clearing.
In some situations, the risk premium on the market portfolio can be written in terms of investors' utility functions.
Assume there is a riskless asset and returns are multivariate normal (MVN). Recall the first order conditions for the canonical portfolio choice problem:

$$\begin{aligned} 0 &= E[u_i'(\tilde{W}_i)(\tilde{r}_j - r_f)] \qquad \forall\, i, j && (6.5.17) \\ &= E[u_i'(\tilde{W}_i)]E[\tilde{r}_j - r_f] + \text{Cov}\left[u_i'(\tilde{W}_i), \tilde{r}_j\right] && (6.5.18) \\ &= E[u_i'(\tilde{W}_i)]E[\tilde{r}_j - r_f] + E[u_i''(\tilde{W}_i)]\text{Cov}\left[\tilde{W}_i, \tilde{r}_j\right] && (6.5.19) \end{aligned}$$

using the definition of covariance and Stein's lemma for MVN distributions. Rearranging:

$$\frac{E[\tilde{r}_j - r_f]}{\theta_i} = \text{Cov}\left[\tilde{W}_i, \tilde{r}_j\right] \tag{6.5.20}$$

where

$$\theta_i \equiv \frac{-E[u_i''(\tilde{W}_i)]}{E[u_i'(\tilde{W}_i)]} \tag{6.5.21}$$

is the i-th investor's **global absolute risk aversion**. Since

$$\tilde{W}_i = W_0^i(1 + r_f + \sum_{k=1}^{N} w_{ik}(\tilde{r}_k - r_f)) \tag{6.5.22}$$

we have, dropping non-stochastic terms,

$$\text{Cov}\left[\tilde{W}_i, \tilde{r}_j\right] = \text{Cov}\left[W_0^i \sum_{k=1}^{N} w_{ik}\tilde{r}_k, \tilde{r}_j\right] \tag{6.5.23}$$

Hence,

$$\frac{E[\tilde{r}_j - r_f]}{\theta_i} = \text{Cov}\left[W_0^i \sum_{k=1}^{N} w_{ik}\tilde{r}_k, \tilde{r}_j\right] \tag{6.5.24}$$

Summing over i, this gives (since we have $\sum_i W_0^i w_{ik} = W_0^m w_{mk}$ by market-clearing and $\sum_k w_{mk}\tilde{r}_k = \tilde{r}_{\mathbf{m}}$ by definition):

$$E[\tilde{r}_j - r_f](\sum_{i=1}^{I} \theta_i^{-1}) = W_{m0}\text{Cov}\left[\tilde{r}_{\mathbf{m}}, \tilde{r}_j\right] \tag{6.5.25}$$

$$\text{or} \qquad E[\tilde{r}_j - r_f] = (\sum_{i=1}^{I} \theta_i^{-1})^{-1} W_{m0}\text{Cov}\left[\tilde{r}_{\mathbf{m}}, \tilde{r}_j\right] \tag{6.5.26}$$

i.e., in equilibrium, the risk premium on the j-th asset is the product of the **aggregate relative risk aversion** of the economy and the covariance between the return on the j-th asset and the return on the market.

Now take the average over j weighted by market portfolio weights:

$$E[\tilde{r}_{\mathbf{m}} - r_f] = (\sum_{i=1}^{I} \theta_i^{-1})^{-1} W_{m0}\text{Var}[\tilde{(}]\tilde{r}_{\mathbf{m}}) \tag{6.5.27}$$

i.e., in equilibrium, the risk premium on the market is the product of the **aggregate relative risk aversion** of the economy and the variance of the return on the market. Equivalently, the return to variability of the market equals the aggregate relative risk aversion.

We conclude with some examples.

1. Negative exponential utility:

$$u_i(z) = -\frac{1}{a_i}\exp\{-a_i z\} \qquad a_i > 0 \tag{6.5.28}$$

   implies:

$$(\sum_{i=1}^{I} \theta_i^{-1})^{-1} = (\sum_{i=1}^{I} a_i^{-1})^{-1} > 0 \tag{6.5.29}$$

   and hence the market portfolio is efficient.

2. Quadratic utility:

$$u_i(z) = a_i z - \frac{b_i}{2} z^2 \quad a_i, b_i > 0 \tag{6.5.30}$$

   implies:

$$(\sum_{i=1}^{I} \theta_i^{-1})^{-1} = \left(\sum_{i=1}^{I}\left(\frac{a_i}{b_i} - E[\tilde{W}_i]\right)\right)^{-1} \tag{6.5.31}$$

   This result can also be derived without assuming MVN and using Stein's lemma.

Chapter 7

INVESTMENT ANALYSIS

7.1 Introduction

[To be written.]

7.2 Arbitrage and the Pricing of Derivative Securities

7.2.1 The binomial option pricing model

This still has to be typed up. It follows very naturally from the stuff in Section 5.4.

7.2.2 The Black-Scholes option pricing model

Fischer Black died in 1995. In 1997, Myron Scholes and Robert Merton were awarded the Nobel Prize in Economics 'for a new method to determine the value of derivatives.' See
`http://www.nobel.se/announcement-97/economy97.html`
Black and Scholes considered a world in which there are three assets: a stock, whose price, $\tilde{S}_t$, follows the stochastic differential equation:

$$d\tilde{S}_t = \mu\tilde{S}_t dt + \sigma\tilde{S}_t d\tilde{z}_t,$$

where $\{\tilde{z}_t\}_{t=0}^{T}$ is a Brownian motion process; a bond, whose price, B_t, follows the differential equation:

$$dB_t = rB_t dt;$$

and a call option on the stock with strike price X and maturity date T.

They showed how to construct an instantaneously riskless portfolio of stocks and options, and hence, assuming that the principle of no arbitrage holds, derived the Black-Scholes partial differential equation which must be satisfied by the option price.
The option pays $(S_T - X)^+ \equiv \max\{S_T - X, 0\}$ at maturity.
Let the price of the call at time t be $\tilde{C}_t$. Guess that $\tilde{C}_t = C(\tilde{S}_t, t)$. By Ito's lemma:

$$d\tilde{C}_t = \left(\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}\mu\tilde{S}_t + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2\tilde{S}_t^2\right)dt + \frac{\partial C}{\partial S}\sigma\tilde{S}_t d\tilde{z}_t$$

The no arbitrage principle yields the partial differential equation:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2 + \frac{\partial C}{\partial S}rS - rC = 0$$

subject to the boundary condition

$$C(S,T) = (S - X)^+.$$

Let $\tau = T - t$ be the time to maturity.
Then we claim that the solution to the Black-Scholes equation is:

$$C(S,t) = SN\left(d\left(S,\tau\right)\right) - Xe^{-r\tau}N\left(d\left(S,\tau\right) - \sigma\sqrt{\tau}\right),$$

where $N(\cdot)$ is the cumulative distribution function of the standard normal distribution and

$$d\left(S,\tau\right) = \frac{\ln\frac{S}{X} + \left(r - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}} + \sigma\sqrt{\tau} \quad (7.2.1)$$

$$= \frac{\ln\frac{S}{X} + \left(r + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}. \quad (7.2.2)$$

We can check that this is indeed the solution by calculating the various partial derivatives and substituting them in the original equation.
Note first that

$$N\left(z\right) \equiv \int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt$$

and hence by the fundamental theorem of calculus

$$N'\left(z\right) \equiv \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2},$$

which of course is the corresponding probability density function. For the last step in this proof, we will need the partials of $d(S,\tau)$ with respect to S and t, which are:

$$\frac{\partial d(S,\tau)}{\partial t} = -\frac{\partial d(S,\tau)}{\partial \tau} = -\frac{\left(r+\frac{1}{2}\sigma^2\right)}{2\sigma\sqrt{\tau}} \tag{7.2.3}$$

and

$$\frac{\partial d(S,\tau)}{\partial S} = \frac{1}{S\sigma\sqrt{\tau}}. \tag{7.2.4}$$

Note also that

$$N'\left(d(S,\tau)-\sigma\sqrt{\tau}\right) = e^{-\frac{1}{2}\sigma^2\tau}e^{d(S,\tau)\sigma\sqrt{\tau}}N'(d(S,\tau)) \tag{7.2.5}$$

$$= e^{-\frac{1}{2}\sigma^2\tau}\left(\frac{S}{X}e^{\left(r+\frac{1}{2}\sigma^2\right)\tau}\right)N'(d(S,\tau)) \tag{7.2.6}$$

$$= \frac{S}{X}e^{r\tau}N'(d(S,\tau)). \tag{7.2.7}$$

Using these facts and the chain rule, we have:

$$\frac{\partial C}{\partial t} = SN'(d(S,\tau))\frac{\partial d(S,\tau)}{\partial t} - Xe^{-r\tau}\times \left(N'\left(d(S,\tau)-\sigma\sqrt{\tau}\right)\left(\frac{\partial d(S,\tau)}{\partial t}-\frac{\sigma}{2\sqrt{\tau}}\right)+rN\left(d(S,\tau)-\sigma\sqrt{\tau}\right)\right) \tag{7.2.8}$$

$$= -SN'(d(S,\tau))\frac{\sigma}{2\sqrt{\tau}} - Xe^{-r\tau}rN\left(d(S,\tau)-\sigma\sqrt{\tau}\right) \tag{7.2.9}$$

$$\frac{\partial C}{\partial S} = SN'(d(S,\tau))\frac{\partial d(S,\tau)}{\partial S} + N(d(S,\tau)) - Xe^{-r\tau}N'\left(d(S,\tau)-\sigma\sqrt{\tau}\right)\frac{\partial d(S,\tau)}{\partial S} \tag{7.2.10}$$

$$= N(d(S,\tau)) \tag{7.2.11}$$

$$\frac{\partial^2 C}{\partial S^2} = N'(d(S,\tau))\frac{\partial d(S,\tau)}{\partial S}. \tag{7.2.12}$$

Substituting these expressions in the original partial differential equation yields:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2 + \frac{\partial C}{\partial S}rS - rC$$

$$\begin{aligned} &= N\left(d\left(S,\tau\right)\right)\left(rS-rS\right)+N'\left(d\left(S,\tau\right)\right)\left(\frac{-S\sigma}{2\sqrt{\tau}}+\frac{1}{2}\sigma^2 S^2\frac{\partial d\left(S,\tau\right)}{\partial S}\right) \\ &\quad +N\left(d\left(S,\tau\right)-\sigma\sqrt{\tau}\right)\left(-Xe^{-r\tau}r+rXe^{-r\tau}\right) \qquad (7.2.13) \\ &= 0. \qquad (7.2.14) \end{aligned}$$

The boundary condition should also be checked. As $\tau \to 0$, $d\left(S,\tau\right) \to \pm\infty$ according as $S > X$ or $S < X$. In the former case, $C\left(S,T\right) = S - X$; and in the latter case, $C\left(S,T\right) = 0$, so the boundary condition is indeed satisfied.

7.3 Multi-period Investment Problems

In Section 4.2, it was pointed out that the objects of choice can be differentiated not only by their physical characteristics, but also both by the time at which they are consumed and by the state of nature in which they are consumed. These distinctions were suppressed in the intervening sections but are considered again in this section and in Section 5.4 respectively.

The multi-period model should probably be introduced at the end of Chapter 4 but could also be left until Chapter 7. For the moment this brief introduction is duplicated in both chapters.

Discrete time multi-period investment problems serve as a stepping stone from the single period case to the continuous time case.

The main point to be gotten across is the derivation of interest rates from equilibrium prices: spot rates, forward rates, term structure, etc.

This is covered in one of the problems, which illustrates the link between prices and interest rates in a multiperiod model.

7.4 Continuous Time Investment Problems

?