THE UNIVERSITY
of ADELAIDE

Mathematics IA

Algebra Outline Notes

School of Mathematical Sciences
The University of Adelaide

Contents

Bibliography

Poole, David (2015), *Linear Algebra: A Modern Introduction*, 4th edition, Cengage Learning .

1 Matrices and Linear Equations

Contents

1.1 Matrices (Poole, §3.0–3.1)

A matrix A of order $m \times n$ is a rectangular array of numbers (usually real numbers), with m rows and n columns:

$$A = \left[a_{ij}\right] = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}$$

where a_{ij} is the element in the ith row and jth column.

◊ **1.0: Space for examples of each type on facing page** . . .

Matrix Operations(Poole, §3.1)

equality : $A = B$ iff $a_{ij} = b_{ij}$ (implies A and B have the same order).

addition : $A + B = [a_{ij} + b_{ij}]$ (implies A and B have the same order). $+$ is associative, commutative; zero matrix: $0 = [0]$.

scalar multiplication : $cA = [ca_{ij}]$, c a scalar (number).

matrix multiplication : A order $m \times r$, B order $r \times n$, then AB is defined to be hte matrix D with i, j-th entry

$$[d_i] = \left[a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj}\right]$$

is of order $m \times n$.

For any matrices A, B, C then the following results hold where the operations are defined:

$$\begin{aligned}(AB)C &= A(BC)\\ A(B+C) &= AB + AC\\ (A+B)C &= AC + BC\\ c(AB) &= (cA)B = A(cB)\end{aligned}$$

but $AB \neq BA$ in general.

◊ **1.1: Example** ...

Identity matrix : Let (the Kronecker symbol) $\delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & \text{otherwise,} \end{cases}$ for $i, j = 1, 2, \ldots, n$. Then the identity matrix of order $n \times n$ is

$$I_n = \left[\delta_{ij}\right].$$

If A is of order $m \times n$, then

$$I_m A = A I_n = A.$$

Matrix powers : $A^p = A \cdot A \cdots A$ for p factors: $A^0 = I$, $A^p A^q = A^{p+q}$.

Diagonal matrix : Square matrix with zeros off the main diagonal.

Matrix transpose :

$$A^t = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}^t = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix}$$

If A is of order $m \times n$, then A^t is of order $n \times m$.

If the element in the ith row and the jth column of A is a_{ij}, then the element in the ith row and jth column of A^t is a_{ji}.

$$(A+B)^t = A^t + B^t, \quad (A^t)^t = A, \quad (cA)^t = cA^t,$$
$$(AB)^t = B^t A^t, \quad (A^p)^t = (A^t)^p,$$

The transpose of a column (row) is a row (column). Thus if $B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, then $B^t = \begin{bmatrix} b_1 & \cdots & b_m \end{bmatrix}$.

Vectors Vectors are written either as $m \times 1$ matrices or as 'm-tuples', that is,

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = (b_1, \ldots, b_m).$$

Whereas 'row vectors' are $1 \times m$ matrices such as $\begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix}$. The equality, addition and scalar multiplication properties listed above for matrices also hold for vectors.

1.2 Linear equations (Poole, Chapter 2)

Linear Equations (Poole, p58) A system of m linear equations in n unknowns (indeterminates) $x_1, x_2, \ldots, x_n$ can be expressed

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ & \vdots & & & \vdots & & \\ a_{i1}x_1 & + & a_{i2}x_2 & + \cdots + & a_{in}x_n & = & b_i \\ & \vdots & & & \vdots & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m \end{array}$$

where a_{ij}, b_i are real numbers $(1 \leq i \leq m,\ 1 \leq j \leq n)$.

Solutions(Poole, p60) A sequence of n real numbers $s_1, s_2, \ldots, s_n$ is called a *solution* of the above system of linear equations if each of the m equations is satisfied when we substitute $x_1 = s_1$, $x_2 = s_2, \ldots, x_n = s_n$. The solution is usually written $(s_1, s_2, \ldots, s_n)$, an element of $\mathbb{R}^n$. The set of all solutions is called the *solution set*.

Two systems of equations are *equivalent* if they have the same solution set.

◊ **1.2: Example** ...

Vectors in $\mathbb{R}^n$ (Poole, p9) We define the set

$$\mathbb{R}^n = \{\mathbf{v} = (v_1, v_2, \ldots, v_n) \mid v_1, \ldots, v_n \in \mathbb{R}\}.$$

$\mathbb{R}^n$ is often called Euclidean n-space.

For $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n)$, $k \in \mathbb{R}$:

$$\begin{aligned}
\mathbf{u} = \mathbf{v} &\text{ if and only if } u_1 = v_1, \ldots, u_n = v_n; \\
\mathbf{u} + \mathbf{v} &= (u_1 + v_1, \ldots, u_n + v_n); \\
k\mathbf{u} &= (ku_1, \ldots, ku_n); \\
\mathbf{0} &= (0, \ldots, 0) \quad \text{the zero vector in } \mathbb{R}^n; \\
-\mathbf{v} &= (-v_1, \ldots, -v_n) \quad \text{the } \textit{negative} \text{ of } \mathbf{v}.
\end{aligned}$$

Algebraic properties of vectors in $\mathbb{R}^n$ (Poole, pp10–12) For each (vector) $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and (scalar) $c, k \in \mathbb{R}$:

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
3. $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
4. $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$
5. $c(k\mathbf{u}) = (ck)\mathbf{u}$
6. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
7. $(c + k)\mathbf{u} = c\mathbf{u} + k\mathbf{u}$
8. $1(\mathbf{u}) = \mathbf{u}$

These properties can be verified using only the definitions of equality, addition and scalar multiplication in $\mathbb{R}^n$.

$\Diamond$ **1.3: Example** ...

Matrix notation for the linear system of equations (Poole, §2.2)

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}.$$

The system of equations is now be represented by the matrix equation

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

A is called the *coefficient matrix* of the system (it is an $m \times n$ matrix). The $m \times (n + 1)$ matrix

$$\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{i1} & \dots & a_{in} & b_i \\ \vdots & & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{bmatrix}$$

is called the *augmented matrix*.

◊ **1.4: Example** . . .

Elementary Operations Elementary operations are performed on a system of linear equations in order to replace it by an equivalent system (one that has the same solution set) which is easier to solve.

The operations are:

1. multiply an equation by a non-zero (real) number;
2. interchange two equations;
3. replace one equation by this equation plus a multiple of another (different) equation.

Elementary Row Operations (Poole, p66) Since the rows of the augmented matrix correspond to the equations in the associated system, these operations correspond to the following *elementary row operations* which are performed on the augmented matrix:

1. multiply a row by a non-zero constant;
2. interchange two rows;
3. add a multiple of one row to another row.

Theorem 1.1. *Any sequence of elementary operations applied to a system of equations produces a new system which is equivalent to the original (that is, has the same solution set).*

◊ **1.5: Outline of Proof** ...

1.3 Gauss–Jordan elimination (Poole, pp68-73)

Given a system of linear equations, our aim is to use elementary row operations to find an equivalent system for which the solutions can easily be found.

Reduced row echelon form A matrix is in reduced echelon form if it has all the following properties:

1. Each row consisting of 0's is at the bottom;
2. The first (left-most) non-zero entry in each row is a 1 (pivot);
3. Entries above and below the pivot are 0's;
4. Each pivot is further to the right than pivots in higher rows.

◊ **1.6: Example** ...

We use elementary row operations to put any matrix into reduced row echelon form using the following algorithm.

Gauss–Jordan Elimination Reduces the augmented matrix using elementary row operations to reduced row echelon form.

1. Locate the left most column which contains a non-zero entry. This is the *pivot* column.
2. Interchange rows (if necessary) so there is a non-zero entry (called the *pivot*) in the pivot column immediately beneath any previous pivot rows.
3. Multiply the pivot row by a suitable number to make the pivot equal to 1.

   Add a suitable multiple of the pivot row to each other row to make all entries above and below the pivot equal to 0. (Do not add multiples of other rows.) Step 3 is called *pivoting*.
4. To choose the next *pivot*, ignore the rows already containing pivots, and locate the left most column that does not consist entirely of zeros. This is the new pivot column.
5. Return to Step 2 and continue until no new pivot columns can be chosen.

Existence of solutions If the rightmost column of the augmented matrix contains a pivot then the system has no solutions.

A system of equations is consistent if there is *a solution*; otherwise it is inconsistent.

Gauss Elimination If in the pivot operation, only the entries *below* the pivot are made zero, when the procedure stops the system is in *row-echelon form*.

Now to solve for the basic variables we use back substitution: that is, substitute for the basic variables we have found as they may appear in previous rows.

◊ **1.7: Example** . . .

Possible solutions Reduce the augmented matrix $[A \mid \mathbf{b}]$ to reduced row echelon form.

No solution :

$$\begin{array}{cccccccc}
1 & * & & & & & * & \\
 & & 1 & & & & \vdots & \\
 & & & & \ddots & & & \\
 & & & & & 1 & * & * \\
0 & 0 & 0 & 0 & \dots & & 0 & 1 \quad \leftarrow \text{ last pivot in last column} \\
0 & 0 & 0 & 0 & \dots & & 0 & 0
\end{array}$$

$0x_1 + 0x_2 + \cdots + 0x_n = 1$ has no solution, so the equations are inconsistent.

$\Diamond$ **1.8: Example with no solution.** ...

Unique solution :

$$\begin{array}{ccccccc}
1 & 0 & 0 & & 0 & 0 & * \\
 & 1 & 0 & & 0 & 0 & * \\
 & & 1 & & 0 & 0 & \vdots \\
 & & & \ddots & \dots & \dots & \vdots \\
 & & & & 1 & 0 & \vdots \\
 & & & & & 1 & * \\
0 & 0 & 0 & \dots & \dots & 0 & 0 \\
0 & \dots & \dots & \dots & \dots & 0 & 0
\end{array}$$

Pivots are on the diagonal. *No* free variables.

$\Diamond$ **1.9: Example with a unique solution.** ...

Infinitely many solutions : With free variables $\boxed{*}$,

$$\begin{array}{cccccccc}
1 & 0 & * & & & & & * \\
0 & 1 & \boxed{*} & & & & & * \\
0 & 0 & 0 & 1 & & & & \\
0 & 0 & 0 & & & & & \\
 & & & & \ddots & & & \\
 & & & & & 1 & \boxed{*} & * \\
0 & 0 & & & & 0 & 0 & 0 \\
0 & 0 & \dots & \dots & \dots & 0 & 0 & 0
\end{array}$$

Free variables, so an infinite number of solutions.

Solve for the the basic variables (pivots) in terms of the free variables.

◊ 1.10: Example with infinitely many solutions. . . .

1.4 Linear combinations (Poole, pp9-11, §2.3)

Definition 1.1. A vector $\mathbf{w}$ is a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_r$ if it can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + \cdots + k_r\mathbf{v}_r \ ,$$

where $k_1, \ldots, k_r \in \mathbb{R}$.

◊ 1.11: Examples . . .

In general, the *vector equation*

$$x_1\mathbf{a}_1 + \cdots + x_r\mathbf{a}_r = \mathbf{b}$$

is equivalent to the *matrix equation*

$$A \begin{bmatrix} x_1 \\ \vdots \\ x_r \end{bmatrix} = \mathbf{b}$$

where A is the matrix whose *columns* are $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_r$.

◊ 1.12: Example . . .
Is $(2, 0, -1)$ a linear combination of $(2, -1, 0)$ and $(0, 2, 1)$?

Definition 1.2. Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be vectors in $\mathbb{R}^n$. Define the *span of* $\mathbf{v}_1, \ldots, \mathbf{v}_r$, denoted $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$, to be the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_r$:

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = \{k_1\mathbf{v}_1 + \cdots + k_r\mathbf{v}_r \mid k_1, \ldots, k_r \in \mathbb{R}\}.$$

This set is also called the *linear space spanned by* $\mathbf{v}_1, \ldots, \mathbf{v}_r$. We say that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *span* the set.

The connection between linear combinations and the system of equations $A\mathbf{x} = \mathbf{b}$ (Poole, p88) For any $m \times n$ matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ (that is, A has columns $\mathbf{a}_1$, $\mathbf{a}_2$, ..., $\mathbf{a}_n$) and vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n .$$

Thus the system $A\mathbf{x} = \mathbf{b}$ has exactly the same solution as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} .$$

The question "Does $\mathbf{b}$ belong to $\operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$?" that is, "Can $\mathbf{b}$ be written as a linear combination of $\mathbf{a}_1$, ..., $\mathbf{a}_n$?" is the same as "Is the system $A\mathbf{x} = \mathbf{b}$ consistent?"

Theorem 1.2. *Let A be an $m \times n$ matrix. The following statements are equivalent.*

(a) For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.

(b) The columns of A span $\mathbb{R}^m$ (that is, $\operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} = \mathbb{R}^m$).

(c) the reduced row echelon form of A has a pivot position in every row.

◊ 1.13: Proof ...

1.5 Homogeneous equations (Poole, p76)

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

As the right hand side (RHS) always consists of 0's we *always* get at least one solution,

$$(x_1, x_2, \ldots, x_n) = (0, 0, \ldots, 0)$$

is always a solution (called trivial). Thus for a homogeneous system there are two cases.

Unique solution :

$$\begin{matrix} 1 & & & & & 0 \\ & 1 & & & & \vdots \\ & & 1 & & & \vdots \\ & & & \ddots & & \\ & & & & 1 & 0 \\ 0 & \dots & \dots & & 0 & 0 \\ 0 & \dots & \dots & & \dots & 0 \end{matrix}$$

Unique solution is $(x_1, \dots, x_n) = (0, 0, \dots, 0)$.

Infinitely many solutions : Free variables $\boxed{*}$

$$\begin{matrix} 1 & & & & & & & 0 \\ & 1 & & & & & & \\ & & 1 & \boxed{*} & * & & & \\ & & & & & \ddots & & \\ & & & & & & 1 \ \boxed{*} & 0 \\ 0 & & & & & & & 0 \end{matrix}$$

◊ 1.14: Example . . .

Consider a general system of equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

The associated homogeneous system is the same system of equations with RHS entries zero.

The general solution of the system = general solution of the associated homogeneous system + a particular (that is, one) solution of the system.

◊ 1.15: Example . . .

General system $A\mathbf{x} = \mathbf{b}$ *Suppose* this system has a solution $(x_1, \ldots, x_n) = (s_1, \ldots, s_n)$. The associated homogeneous system is

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

(Note: the homogeneous system always has at least the solution

$$(x_1, \ldots, x_n) = (0, 0, \ldots, 0).)$$

Solution of $A\mathbf{x} = \mathbf{b}$ is

$$(x_1, \ldots, x_n) = \underbrace{(s_1, \ldots, s_n)}_{\text{particular solution}} + (\text{general solution of } A\mathbf{x} = \mathbf{0})$$

◊ **1.16: Why is this true?** ...

If $A\mathbf{x} = \mathbf{0}$ has only one (trivial) solution $(x_1, \ldots, x_n) = (0, \ldots, 0)$, then $A\mathbf{x} = \mathbf{b}$ also has only one (that is, a unique) solution

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}.$$

If $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions, then $A\mathbf{x} = \mathbf{b}$ has general solution

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \underbrace{\begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}}_{\text{particular solution}} \underbrace{+ s \begin{bmatrix} \\ \vdots \\ \\ \end{bmatrix} + t \begin{bmatrix} \\ \vdots \\ \\ \end{bmatrix} + \cdots}_{\text{general solution of } A\mathbf{x}=0}$$

◊ **1.17: Example** ...

1.6 Elementary matrices and the inverse matrix (Poole, §3.3)

Definition 1.3 (Inverse of a matrix)**.** The square $n \times n$ matrix A is *invertible* if there exists a $(n \times n)$ matrix B such that $AB = BA = I$. B is the *inverse* of A and we write $B = A^{-1}$.

Notes

1. Not every square matrix has an inverse.

   ♢ **1.18: example** ...

2. If A has an inverse it is *unique*.

   ♢ **1.19: Proof** ...

Properties of Inverses If A and B have inverses A^{-1} and B^{-1}, respectively, then

1. cA has inverse $\frac{1}{c}A^{-1}$, provided $c \neq 0$;
2. $A^p = \underbrace{A \cdots A}_{p}$ has inverse $\underbrace{A^{-1} \cdots A^{-1}}_{p} = A^{-p}$;
3. A^{-1} has inverse $A = (A^{-1})^{-1}$;
4. A^t has inverse $(A^{-1})^t$;
5. AB has inverse $B^{-1}A^{-1}$.

   ♢ **1.20: proof** ...

6. If (the $n \times n$ matrix) A has an inverse, then the system of equations
   $$A\mathbf{x} = \mathbf{b}$$
   has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$ for all $\mathbf{b} \in \mathbb{R}^n$. In particular, $A\mathbf{x} = \mathbf{0}$ has only the (trivial) zero solution $\mathbf{x} = \mathbf{0}$.

Elementary matrices (Poole, p170) Corresponding to each elementary row operation is an *elementary matrix*.

The elementary matrix is formed by performing the elementary row operation on the identity I.

Premultiplying by the elementary matrix has the same effect as performing the elementary row operation.

Elementary $n \times n$ matrices We list the elementary matrix for the three principal rows operations.

1. $Ri = cRi$

$$\begin{array}{cc}
\begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & & & & \\
\cdots & \cdots & \cdots & c & \cdots & \cdots & \cdots \\
& & & & 1 & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix}
\end{array}$$

2. $Ri, Rj = Rj, Ri$

$$\begin{array}{cc}
& \begin{array}{ccccccccc} & & & i & & j & & & \end{array} \\
\begin{array}{c} \\ \\ \\ i \\ \\ j \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & & & \\
& \ddots & & & & & & & \\
& & 1 & & & & & & \\
0 & \cdots & \cdots & 0 & \cdots & 1 & \cdots & \cdots & 0 \\
\vdots & & & & & & & & \vdots \\
0 & \cdots & \cdots & 1 & \cdots & 0 & \cdots & \cdots & 0 \\
& & & & & & 1 & & \\
& & & & & & & \ddots & \\
& & & & & & & & 1
\end{bmatrix}
\end{array}$$

3. $Rj = Rj + cRi$

$$\begin{array}{cc}
& \begin{array}{ccccccc} & & i & & j & & \end{array} \\
\begin{array}{c} \\ \\ i \\ \\ j \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & \\
& \ddots & & & & & \\
& & 1 & \cdots & 0 & & \\
& & \vdots & \ddots & \vdots & & \\
& & c & \cdots & 1 & & \\
& & & & & \ddots & \\
& & & & & & 1
\end{bmatrix}
\end{array}$$

Elementary matrices are used to derive properties of the inverse of a square matrix. One reason they are particularly useful is that they are straightforwardly invertible.

◊ 1.21: Example ...

Inverses of Elementary Matrices An elementary matrix has an inverse which is also an elementary matrix. The inverse matrix corresponds to the inverse of the elementary row operation.

- Provided $c \neq 0$:

$$\begin{array}{c} \\ \\ i \\ \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & c & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}^{-1} = \begin{array}{c} \\ \\ i \\ \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & \frac{1}{c} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

- Swapping rows is its own inverse:

$$\begin{array}{c} \\ \\ i \\ j \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}^{-1} = \begin{array}{c} \\ \\ i \\ j \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

- For any c:

$$\begin{array}{c} \\ \\ i \\ j \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & c & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}^{-1} = \begin{array}{c} \\ \\ i \\ j \\ \\ \\ \end{array}\begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -c & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

Definition 1.4. If A is a square $n \times n$ matrix, and if A is reduced to I_n by Gauss–Jordan elimination, then A is *row equivalent to the identity*.

Theorem 1.3. (Poole, p172) *If A is an $n \times n$ matrix, then the following statements are equivalent:*

(a) A has an inverse A^{-1};

(b) A is row equivalent to I_n;

(c) The homogeneous equations $A\mathbf{x} = \mathbf{0}$ have only the trivial solution $\mathbf{x} = \mathbf{0}$.

$\Diamond$ **1.22: proof** ...

Algorithm for finding A^{-1} (Poole, pp175–178) The above proof shows us how to compute the inverse of A using elementary row operations.

Begin with the augmented matrix $[A\,I]$ and perform elementary row operations to reduce A to I. The same operations performed on I produce A^{-1}:

If $E_r \cdots E_1 A = I$, then $A^{-1} = E_r \cdots E_1$. But

$$E_r \cdots E_1 I = E_r \cdots E_1 = A^{-1}$$

so $[A\,I] \to [I\,A^{-1}]$ using elementary row operations.

◊ **1.23: Exmaple** ...

Theorem 1.4. *(a) If $AB = I$, then $BA = I$ and A has inverse $A^{-1} = B$.*

(b) If $BA = I$, then $AB = I$ and A has inverse $A^{-1} = B$.

◊ **1.24: Proof** ...

Theorem 1.5. *Let A be an $n \times n$ matrix. The following conditions are equivalent:*

(a) A has inverse A^{-1};

(b) A is row equivalent to I_n;

(c) $A\mathbf{x} = \mathbf{0}$ has only the zero solution $\mathbf{x} = \mathbf{0}$;

(d) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any column matrix $\mathbf{b}$;

(e) The columns of A span $\mathbb{R}^n$.

◊ **1.25: Proof** ...

◊ **1.26: Examples** ...

Further Examples

2 Determinants

Contents

2.1 Basic definitions (Poole, §4.2)

Given matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ you already know

$$\det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc,$$

and that A is invertible iff $\det A \neq 0$. Similar results for 3×3 matrices. We generalise these results.

Basic Definitions Let A be an $n \times n$ matrix, $A = [a_{ij}]$.

- If $n = 1$, so $A = [a_{11}]$, then $\det A = |A| = a_{11}$.
- If $n > 1$, then

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

  where A_{1j} is the determinant of the matrix obtained by omitting the first row and the jth column of A.

◊ **2.0: Calculating determinants** . . .

Generally, $|A|$ is sum of $n!$ terms, each term is $\pm$ product of n terms, one from each row and column of A.

◊ **2.1: Further examples** . . .

Expansion by any Row or Column (Poole, p266) The i, j minor of A, A_{ij}, is the determinant of the matrix obtained by removing the ith row and jth column of A.

Theorem 2.1. • *Expansion by ith row:*

$$\begin{aligned}&\det A\\&= (-1)^{i+1}a_{i1}A_{i1} + (-1)^{i+2}a_{i2}A_{i2} + \cdots + (-1)^{i+n}a_{in}A_{in}\\&= \sum_{j=1}^{n}(-1)^{i+j}a_{ij}A_{ij}\,.\end{aligned}$$

• *Expansion by jth column:*

$$\begin{aligned}&\det A\\&= (-1)^{1+j}a_{1j}A_{1j} + (-1)^{2+j}a_{2j}A_{2j} + \cdots + (-1)^{n+j}a_{nj}A_{nj}\\&= \sum_{i=1}^{n}(-1)^{i+j}a_{ij}A_{ij}\,.\end{aligned}$$

An immediate consequence is that a matrix with a row or column of zeros must have a determinant of 0.

♢ **2.2: Examples** . . .

In general, the determinant of an upper or lower triangular matrix equals the product of its diagonal entries.

2.2 Row operations on determinants (Poole, p271)

Row Properties

1. If B is obtained from A by interchanging two rows, then $\det B = -\det A$.

   ♢ **2.3: Example** . . .

2. If A has a row of zeros or two equal rows, then $\det A = 0$.

   ♢ **2.4: Examples** . . .

3. If B is obtained from A by multiplying one row of A by k, then $\det B = k \det A$.

   ♢ **2.5: Example** . . .

4. If B is obtained from A by adding to one row a multiple of another row, then $\det B = \det A$.
5. If the $n \times n$ matrices A, B and C differ in only one row, say the ith row, and if $c_{ij} = a_{ij} + b_{ij}$ $(j = 1, \ldots, n)$, then $\det C = \det A + \det B$.

   ◊ **2.6: Example** ...

Using row ops to calculate determinants Using these row properties, determinants can be calculated efficiently by row reduction.

◊ **2.7: Examples of using row operations to calculate a determinant** ...

2.3 General properties of determinants (Poole, pp272–274)

Transpose : $\det(A^t) = \det A$.

◊ **2.8: Example** ...

Column Properties : Row properties 1–5 above also hold with "row" replaced by "column".

Product : $\det(AB) = (\det A)(\det B)$.

◊ **2.9: Example** ...

Inverse :

- If A is invertible, then $\det A^{-1} = (\det A)^{-1}$.

  Proof. Since $AA^{-1} = I$, $\det A \cdot \det A^{-1} = \det I$. □

- A is invertible iff $\det A \neq 0$.

  ◊ **2.10: Proof** ...

2.4 The Adjoint (Poole, pp274–275)

Inverse of a matrix Let $A = [a_{ij}]$ be an $n \times n$ matrix. The i, j *cofactor of* A is

$$C_{ij} = (-1)^{i+j} A_{ij}$$

From the ith row expansion

$$\begin{aligned} \det A &= a_{i1}(-1)^{i+1}A_{i1} + a_{i2}(-1)^{i+2}A_{i2} + \cdots + a_{in}(-1)^{i+n}A_{in} \\ &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \, . \end{aligned}$$

From the jth column expansion

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \, .$$

The *adjoint* of A, $\operatorname{adj} A$, is the matrix whose i, j element is C_{ji}.

◊ 2.11: Example . . .

Theorem 2.2. $A \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I$. *If A is invertible, then* $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$.

◊ 2.12: Proof (if time) . . .

3 Optimisation and Convex Sets

Contents

Optimisation Example A cereal manufacturer makes two kinds of muesli, *nutty special* and *fruity extra*. Each muesli consists of raisins and nuts.

- 0.2 boxes of raisins and 0.4 boxes of nuts make 1 box of *nutty special*.
- 0.4 boxes of raisins and 0.2 boxes of nuts make 1 box of *fruity extra*.
- 14 boxes of raisins and 10 boxes of nuts are available each day.
- The profit on each box of *nutty special* is $8.
- The profit on each box of *fruity extra* is $10.

Assuming that all the muesli that is made will be sold, how many boxes of *nutty special* and *fruity extra* should be made each day, in order to maximise the profit?

◊ **3.0: Mathematical Formulation** . . .

3.1 Convex sets

Optimisation problems typically involve a set of *linear constraints* which restrict the possible solutions.

For example, $x_1, x_2, \ldots, x_n$ measure the quantities of inputs (perhaps to a manufacturing process). These are subject to constraints (such as availability). Hence

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &\leq b_2 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &\leq b_m \end{aligned}$$

or

$$A\mathbf{x} \leq \mathbf{v}.$$

If, in addition, each $x_i \geq 0$, then we write $\mathbf{x} \geq 0$.

Subject to these constraints we may wish to *maximise* profit, given by

$$f(x_1, \ldots, x_n) = c_1x_1 + \cdots + c_nx_n.$$

Alternatively $f(x_1, \ldots, x_n)$ may represent a cost which we would want to *minimise*. $f(x_1, \ldots, x_n)$ is the *objective function*.

The constraints define the *feasible region* which is a *convex set*. The maximum or minimum value of the objective function occurs at a vertex. Hence we need only find the value of the objective function at the vertices of the feasible region.

Definition 3.1. A convex set C is a set of points (or region) in $\mathbb{R}^n$ such that the line segment joining any two points in C lies completely in C.

◊ **3.1: Examples** . . .

Often we consider sets defined by systems of inequalities.

◊ **3.2: Examples** . . .

We now give a more precise (mathematical) formulation of convex sets and vertices, and also show that a set defined by linear inequalities is a convex set. We use vectors in $\mathbb{R}^n$. Recall that we extended vectors to n-dimensional space $\mathbb{R}^n$:

$$\mathbb{R}^n = \{(v_1, \dots, v_n) \mid v_i \in \mathbb{R}\}$$

In addition we extend the dot (scalar) product by the definition that if $\mathbf{u} = (u_1, \dots, u_n)$ and $\mathbf{v} = (v_1, \dots, v_n)$ then

$$\mathbf{u} \cdot \mathbf{v} = (u_1, \dots, u_n) \cdot (v_1, \dots, v_n) = u_1 v_1 + u_2 v_2 + \dots + u_n v_n \,.$$

All the properties of the dot product carry over from $\mathbb{R}^2$ or $\mathbb{R}^3$ to $\mathbb{R}^n$: for three examples,

$$\begin{aligned}(\mathbf{u}+\mathbf{v}) \cdot \mathbf{w} &= \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \\ (t\mathbf{u}) \cdot \mathbf{v} &= t(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (t\mathbf{v}) \\ \mathbf{u} \cdot \mathbf{v} &= \mathbf{v} \cdot \mathbf{u} \,.\end{aligned}$$

Recall that a set C is convex if for any two points P and Q in C, the line joining P and Q lies in C.

We look for a mathematical expression for the line segment PQ. This is done with vectors.

◊ 3.3: Derivation of formula for line segment . . .

Thus a set C (of $\mathbb{R}^n$) is *convex* if for any two points P and Q in C, with position vectors $\mathbf{u}$ and $\mathbf{v}$ respectively ($\overrightarrow{OP} = \mathbf{u}, \overrightarrow{OQ} = \mathbf{v}$), then for all t, $0 \le t \le 1$,

$$(1-t)\mathbf{u} + t\mathbf{v} \in C.$$

We now show the set of points (vectors) satisfying a set of inequalities such as $A\mathbf{x} \le \mathbf{b}$ is convex.

Consider the system of inequalities $A\mathbf{x} \le \mathbf{b}$; that is,

$$\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n \le b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n \le b_2 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \le b_m \end{array}$$

The ith inequality ($i = 1, \dots, m$) $a_{i1}x_1 + \dots + a_{in}x_n \le b_i$ can be written in the form $\mathbf{a}_i \cdot \mathbf{x} \le b_i$ where $\mathbf{a}_i = (a_{i1}, \dots, a_{in})$ and $\mathbf{x} = (x_i, \dots, x_n)$.

Consider the set $C_i = \{\mathbf{x} \mid \mathbf{a}_i \cdot \mathbf{x} \leq b_i\} = \{\mathbf{x} = (x_1, \ldots, x_n) \mid (a_{i1}, a_{i2}, \ldots, a_{in}) \cdot \mathbf{x} \leq b_i\}$. That is, C_i is the set consisting of all vectors $\mathbf{x} = (x_i, \ldots, x_n)$ in $\mathbb{R}^n$ satisfying the ith inequality $a_{i1}x_1 + \cdots + a_{in}x_n \leq b_i$ or $\mathbf{a}_i \cdot \mathbf{x} \leq b_i$. We show that C_i is a convex set in $\mathbb{R}^n$.

♢ 3.4: Showing that C_i is convex ...

Theorem 3.1. *The intersection of two convex sets is convex.*

♢ 3.5: proof ...

We have shown that each C_i is convex, and the theorem just proved shows that the intersection of convex sets is convex.

Thus $A\mathbf{x} \leq \mathbf{b}$ defines a convex set in $\mathbb{R}^n$, as does the combination $A\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.

Finally, as a necessary preliminary to introducing our general optimisation theorem, we give precise (mathematical) definitions of a *vertex* of a convex set and a bounded convex set.

Vertices A vertex $\mathbf{v}$ of C is point of C that does not lie on a straight line between two other points in C.

Definition 3.2. A point P with position vector $\overrightarrow{OP} = \mathbf{v}$ is a vertex of the convex set C if $\mathbf{v}$ cannot be written in the form

$$(1-t)\mathbf{u} + t\mathbf{w}, \quad 0 \leq t \leq 1,$$

for any two distinct points $\mathbf{u}, \mathbf{w} \in C$ except when $\mathbf{v} = \mathbf{u}$ (and $t = 0$) or $\mathbf{v} = \mathbf{w}$ (and $t = 1$).

Definition 3.3. A convex set C is *bounded* if there is some positive number M such that for any point P in C, the length

$$\|\overrightarrow{OP}\| < M.$$

(That is, the length of $\overrightarrow{OP}$ cannot be arbitrarily large for points P in C.)

We now state our main result concerning optimisation problems.

Theorem 3.2. *Let C be a* convex set *defined by a set of linear inequalities $A\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$. If C is bounded, then the linear function*

$$f(x_1, \ldots, x_n) = c_1 x_1 + \cdots + c_n x_n$$

takes its maximum (or minimum) value on C at a vertex of C. If C is an unbounded convex set, and if f takes a maximum (or minimum) value on C, then this maximum (or minimum) occurs at a vertex of C.

$\Diamond$ **3.6: Proof** ...

3.2 Methods for solving optimisation problems

Theorem 3.2 tells us how to solve these optimisation problems. Determine the vertices, evaluate the linear function f at each vertex, and check which is the maximum (or minimum) value. The only problem that remains to be solved is to find the vertices. In two dimensions we can sketch the region and determine the vertices, which are the intersections of certain boundary lines.

Cereal example A cereal manufacturer makes two kinds of muesli, Nutty Special and Fruity Extra. Each muesli consists of raisons and nuts. 0.2 boxes of raisins and 0.4 boxes of nuts make 1 box of Nutty Special. 0.4 boxes of raisins and 0.2 boxes of nuts make 1 box of Fruity Extra. Suppose that 14 boxes of raisins and 10 boxes of nuts are available each day. The profit on each box of Nutty Special is \$8 and on each box of Fruity Extra is \$10. Assuming all the muesli that is made is sold how many boxes of Nutty Special and Fruity Extra should be made each day in order to maximise the profit?

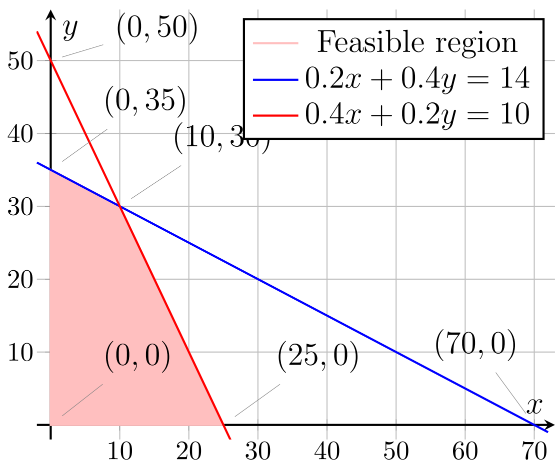
Graphical Solution Recall the mathematical formulation. We wish to *maximise* the profit

$$P(x, y) = 8x + 10y,$$

subject to the four *constraints*

$$0.2x + 0.4y \leq 14 \tag{3.1}$$
$$0.4x + 0.2y \leq 10 \tag{3.2}$$
$$x \geq 0 \tag{3.3}$$
$$y \geq 0 \tag{3.4}$$

The set of points $(x, y) \in \mathbb{R}^2$ satisfying the constraints is the *feasible region*. We plot this region and find the solution.



◊ 3.7: Finding the solutions ...

This graphical method is only practical if there are just two variables in the problem.

We now give an algebraic method of solution of optimisation problems which is applicable in general.

First note that if we have an inequality of the form

$$a_{j_1}x_1 + \cdots + a_{j_n}x_n \geq b_j$$

we can multiply by (-1) and put our inequality in the form

$$-a_{j_1}x_1 + (-a_{j_2}x_2) + \cdots + (-a_{j_n}x_n) \leq (-b_j).$$

Thus we may suppose all our inequalities are of the form

$$a_{j_1}x_1 + \cdots + a_{j_n}x_n \leq b_j.$$

To solve an Optimisation Problem we have to find the vertices of the convex set (which is the feasible region). The values of the function f are then be computed at each vertex and the maximum or minimum of f determined.

Method For each inequality, add a *slack variable* to produce a linear equation (equality):

$$\left.\begin{array}{l} a_{11}x_1 + \cdots + a_{1n}x_n \leq b_1 \\ \quad\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \leq b_m \end{array}\right\} m \text{ inequalities}$$

$$\left.\begin{array}{llllll} a_{11}x_1 + \cdots + a_{1n}x_n & + x_{n+1} & & & & = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n & & + x_{n+2} & & & = b_2 \\ \quad\vdots & & & & & \\ a_{m1}x_1 + \cdots + a_{mn}x_n & & & & + x_{n+m} & = b_m \end{array}\right\} \begin{array}{l} m \text{ different} \\ \text{slack} \\ \text{variables} \\ \text{added} \end{array}$$

Note each variable $x_j \geq 0$ (including the slack variables). (If one of the constraints is actually an equality, then there is no need to add a slack variable.)

As there are m equations there are m possible pivots or basic variables; here by a *pivot column* we mean a column with a pivot, in other words one of the columns of the $m \times m$ identity matrix I_m. There are $\binom{m+n}{m}$ possible choices for m pivots or basic variables. For *each choice*, use Gauss–Jordan elimination to obtain a row equivalent matrix with the chosen columns as pivot columns.

Put the free variables $= 0$ and solve for the basic variables. This solution is called a basic solution.

There at most $\binom{m+n}{m}$ basic solutions; however for some of the choices of basic variables it may not be possible to get a basic solution; for the last (or second to last or ...) choice of pivot it may not be possible to pivot as there may be a zero in the pivot position.

The *vertices* of the convex set are the non-negative basic solutions. (Non-negative as each $x_j \geq 0$.)

In the previous example the basic solutions correspond to the intersections of the defining (bounding) lines. The negative basic solutions correspond to those intersections which are not vertices (that is, lie outside the feasible region).

Example Find the maximum value of the function

$$f(x_1, x_2, x_3) = 6x_1 + 4x_2 - 7x_3$$

subject to the constraints

$$\begin{aligned} 2x_1 + x_2 - 2x_3 &\leq 10, \\ x_1 + 4x_2 - x_3 &\leq 12, \end{aligned}$$

and $x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0.$

Solution Add *slack variables* x_4 and x_5 so that the first two inequalities become equalities:

$$\begin{aligned} 2x_1 + x_2 - 2x_3 + x_4 &= 10, \\ x_1 + 4x_2 - x_3 + x_5 &= 12 \end{aligned}$$

Note that $x_4 \geq 0$ and $x_5 \geq 0$ also.

The vertices of the feasible region are the non-negative basic solutions. To find the basic solutions we consider all $\binom{5}{2} = 10$ pairs of variables. The free variables in each case are put equal to zero and the solution is then a basic solution. Take the non-negative basic solutions and determine the maximum value of f for each of these solutions.

Note

- It is usually possible to go from one basic solution to another by a single pivot operation.
- Not every pair of variables may give rise to a basic solution. The second pivot operation is not be possible when there is a zero in the position in which you must perform this pivoting operation.

	Pivots (basic var's)	Equations						Basic solution
1.	x_4, x_5	2	1	-2	1	0	10	
		1	4	-1	0	1	12	$(0, 0, 0, 10, 12)$
2.	pivot (1,3)							
	x_3, x_5	-1	$-1/2$	1	$-1/2$	0	-5	
		0	$7/2$	0	$-1/2$	1	7	$(0, 0, -5, 0, 7)$
3.	pivot (1,2)							
	x_2, x_5	2	1	-2	1	0	10	
		-7	0	7	-4	1	-28	$(0, 10, 0, 0, -28)$
4.	pivot (1,1)							
	x_1, x_5	1	$1/2$	-1	$1/2$	0	5	
		0	$7/2$	0	$-1/2$	1	7	$(5, 0, 0, 0, 7)$
5.	pivot (2,2)							
	x_1, x_2	1	0	-1	$4/7$	$-1/7$	4	
		0	1	0	$-1/7$	$2/7$	2	$(4, 2, 0, 0, 0)$
6.	pivot (2,3)							
	x_1, x_3	Is not possible as the (2,3) entry is zero. Thus x_1, x_3 *cannot* be a pair of basic variables.						
7.	pivot (2,4)							
	x_1, x_4	1	4	-1	0	1	12	
		0	-7	0	1	-2	-14	$(12, 0, 0, -14, 0)$
8.	pivot (1,2)							
	x_2, x_4	$1/4$	1	$-1/4$	0	$1/4$	3	
		$7/4$	0	$-7/4$	1	$-1/4$	7	$(0, 3, 0, 7, 0)$
9.	pivot (2,3)							
	x_2, x_3	0	1	0	$-1/7$	$2/7$	2	
		-1	0	1	$-4/7$	$1/7$	-4	$(0, 2, -4, 0, 0)$
10.	pivot (1,4): $R1, R2 = R2, R1$							
	x_3, x_4	-1	-4	1	0	-1	-12	
		0	-7	0	1	-2	-14	$(0, 0, -12, -14, 0)$

$\Diamond$ **3.8: Finding the solution** ...

3.3 Formulation of optimisation problems

There are two parts to any optimisation problem. The first is to formulate the problem as a mathematical problem and the second is to solve the mathematical problem. In the next examples we particularly consider the first part of the problem.

Example 1: A typical optimisation problem

A fruit dealer can transport up to 800 boxes of fruit from Renmark to Adelaide on a truck. He must transport *at least* 200 boxes of oranges, at *at least* 100 boxes of grapefruit and *at most* 200 boxes of tangerines. The profit per box is \$2 for oranges, \$1 for grapefruit, and \$3 for tangerines. *How many boxes* of *each* kind of fruit should be loaded onto the truck in order to maximise profit?

Before we can solve this problem we need to *formulate* it mathematically. This means doing three things:

1. Identify the unknowns in the problem, that is, define the variables;
2. Write down what is to be maximised or minimised (objective function);
3. Write down any constraints, including non-negativity constraints.

Thus a typical set-up (and the sort we want to see!) is

In general

1. Let $x_1 = \cdots$, $x_2 = \cdots$, ...
2. Maximise $z = \cdots$ (or minimise)
3. subject to

$$a_1x_1 + a_2x_2 + \cdots \le b_1$$
etc.
$$x_i \ge 0$$

This problem

1. Let x_1 = the no. of boxes of oranges, x_2 = the no. of boxes of grapefruit, x_3 = the no. of boxes of tangerines loaded onto the truck.
2. Maximise (profit) $z = 2x_1 + x_2 + 3x_3$ (in dollars)
3. subject to

$$\begin{aligned}
&x_1 + x_2 + x_3 \le 800,\\
&x_1 \ge 200,\\
&x_2 \ge 100,\\
&x_3 \le 200,\\
&x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0.
\end{aligned}$$

Now, how should *you* approach the problem to get it in this form?

1. Read the problem right through before you start.

2. Focus on the question asked in the problem.

   Clue: It normally begins "How" (How many, how much, etc.) and ends with "?". What follows after the "how" should indicate to you the quantities you wish to find, that is, the unknowns. Thus "How many boxes of each kind of fruit" suggests (since there are three kinds of fruit in our problem) that you need three variables, each one representing the number of boxes of a particular type of fruit. Consequently, you can use this to define your variables.

3. To get the objective function, look for the *task* or *objective* the question sets.

   Clue: This normally includes the words 'maximise' or 'minimise'. Thus in the fruit dealer problem the phrase 'maximise profit' gives the objective, and hence the objective function.

   Note that profits, costs, etc., are given on a unit basis, for example, 2 dollars per box of oranges. Thus if we use x_1 of these units, that is x_1 boxes, we get a profit from oranges of $2x_1$ dollars. Similarly for grapefruits and tangerines. Thus total profit will be a sum of the three individual contributions: from oranges, grapefruit, and tangerines.

4. Look for *constraints* on the variables in the problem. These might be recorded in sentence form, or tabulated. A correct choice of variables in Step 2 would generally lead to an easy identification of constraints in Step 4. The fruit dealer problem has four constraints, not including the non-negativity conditions. These are based on the total number of boxes the truck can carry, and the individual quotas on the three different kinds of fruit. Don't forget non-negativity constraints, if your variables in Step 2 must be greater than or equal to zero.

If you follow these steps, you should be well on the way to formulating an optimisation problem mathematically and remember, the more practice you do, the better you'll get at it.

◊ **3.9: Solving this problem** ...

Example 2: chemical plant

A manufacturer of a certain chemical product has two plants where the product is made. Plant X can make at most 30 tons per week and plant Y can make at most 40 tons per week. The manufacturer wants to make a total of at least 50 tons per week. The amount of

particulate matter found weekly in the atmosphere over a nearby town is measured and found to be 20 pounds for each ton of the product made by plant X and 30 pounds for each ton of the product made at plant Y. How many tons should be made weekly at each plant to minimise the total amount of particulate matter in the atmosphere?

◊ **3.10: Mathematical Formulation** ...

Example 3: nutrition

A nutritionist is planning a meal that includes foods A and B as its main staples. Suppose that each gram of food A contains 60 units of protein, 30 units of iron, and 30 units of thiamine; each gram of food B contains 30 units of protein, 30 units of iron, and 90 units of thiamine. Suppose that each gram of A costs 2 cents, while each gram of B costs 3 cents. The nutritionist wants the meal to provide at least 360 units of protein, at least 270 units of iron, and at least 450 units of thiamine. How many grams of each of the foods should be used to minimise the cost of the meal?

◊ **3.11: Mathematical Formulation** ...

4 The Vector Space $\mathbb{R}^n$

Contents

Recall that $\mathbb{R}^n = \{\mathbf{v} = (v_1, v_2, \dots, v_n) \mid v_1, \dots, v_n \in \mathbb{R}\}$. We say $\mathbb{R}^n$ is a *vector space* under the usual addition and scalar multiplication of vectors, because it satisfies the eight properties previously discussed and listed on page 12.

4.1 Linear dependence and independence (Poole, pp92–97)

Recall that a vector $\mathbf{b}$ is a linear combination of vectors $\mathbf{v}_1,\dots,\mathbf{v}_r$ if

$$\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r$$

for some scalars c_1, c_2,..., c_r. This statement is equivalent to saying the matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution $(x_1 = c_1,\dots,x_r = c_r)$ where A has columns $\mathbf{v}_1,\dots,\mathbf{v}_r$. The concepts of linear dependence or independence are related to whether the homogeneous equation

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

has more than one solution. We know we can write $\mathbf{0}$ as a linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_r$ by

$$0\mathbf{v}_1 + \cdots + 0\mathbf{v}_r = \mathbf{0}$$

Is there any other way? Can we write

$$x_1\mathbf{v}_1 + \cdots + x_r\mathbf{v}_r = \mathbf{0}$$

for some scalars $x_1, \dots, x_r$ not all zero?

Definition 4.1. The list of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is $\mathbb{R}^n$ is *linearly independent* if the equation

$$x_1\mathbf{v}_1 + \cdots + x_r\mathbf{v}_r = \mathbf{0}$$

has only the trivial solution

$$x_1 = \cdots = x_r = 0.$$

Otherwise the list is *linearly dependent*.

Thus $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is dependent if there exist scalars $x_1, \ldots, x_r$, not all zero, such that $x_1\mathbf{v}_1 + \cdots + x_r\mathbf{v}_r = \mathbf{0}$.

Note If the matrix A has columns $\mathbf{v}_1, \ldots, \mathbf{v}_r$, then this definition is equivalent to:

1. if the homogeneous equations $A\mathbf{x} = \mathbf{0}$ have only the solution $\mathbf{x} = \mathbf{0}$, then $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent
2. if the homogeneous equations $A\mathbf{x} = \mathbf{0}$ have non-zero solutions, then $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly dependent.

$\lozenge$ **4.0: Examples** ...

Linear dependence in $\mathbb{R}^2$

When are vectors in $\mathbb{R}^2$ linearly dependent? Two vectors in $\mathbb{R}^2$ are linearly dependent iff one is a multiple of the other.

$\lozenge$ **4.1: Why is this true?** ...

Any list of three or more vectors in $\mathbb{R}^2$ is linearly dependent. For let $\mathbf{v}_1, \ldots, \mathbf{v}_r \in \mathbb{R}^2$, $r \geq 3$. Consider

$$x_1\mathbf{v}_1 + \cdots + x_r\mathbf{v}_r = \mathbf{0}.$$

This leads to two linear equations in r unknowns. There must be at least one free variable, as $r > 2$, so there are infinitely many solutions. Hence there is a non-trivial solution, and $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly dependent. *Any list of more than n vectors in $\mathbb{R}^n$ is linearly dependent.*

$\lozenge$ **4.2: example** ...

Linear Dependence in $\mathbb{R}^3$ We already know:

- any list of four or more vectors in $\mathbb{R}^3$ is linearly dependent (by the above argument);
- two vectors in $\mathbb{R}^3$ are linearly dependent iff one of them is a multiple of the other.

◊ 4.3: When are three vectors in $\mathbb{R}^3$ linearly dependent? ...

Thus three vectors in $\mathbb{R}^3$ are linearly dependent iff they lie in the same plane.

Theorem 4.1. (Poole, p93) *The list of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly dependent iff one $\mathbf{v}_i$ is a linear combination of the others. Equivalently, if $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_j$ ($j > 1$) is a linear combination of the preceding vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.*

◊ 4.4: proof ...

4.2 Subspaces (Poole, pp433-438)

Definition 4.2. A *subspace* of $\mathbb{R}^n$ is a subset $\mathbb{W}$ of $\mathbb{R}^n$ such that

- the zero vector of $\mathbb{R}^n$ is in $\mathbb{W}$, $\mathbf{0} \in \mathbb{W}$, and
- for all $\mathbf{u}, \mathbf{v} \in \mathbb{W}$, $\mathbf{u} + \mathbf{v} \in \mathbb{W}$, and
- for all $c \in \mathbb{R}$ and $\mathbf{u} \in \mathbb{W}$, $c\mathbf{u} \in \mathbb{W}$.

That is, $\mathbb{W}$ is *closed under addition* and *scalar multiplication*.

◊ **4.5: Examples** . . .

Theorem 4.2. $\mathbb{W} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ *is a subspace of* $\mathbb{R}^n$*. It is the smallest subspace of* $\mathbb{R}^n$ *containing* $\mathbf{v}_1, \ldots, \mathbf{v}_r$*.*

◊ **4.6: proof** . . .

Subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$

Every subspace of $\mathbb{R}^2$ is either

- $\{\mathbf{0}\}$, or
- a line through $\mathbf{0}$, or
- $\mathbb{R}^2$.

◊ **4.7: proof** . . .

Similarly, every subspace of $\mathbb{R}^3$ is either

- $\{\mathbf{0}\}$,
- a line through $\mathbf{0}$,
- a plane through $\mathbf{0}$, or
- $\mathbb{R}^3$.

4.3 Basis (Poole, pp446-448)

Definition 4.3. Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$. A list of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ in $\mathbb{W}$ is a *basis* for $\mathbb{W}$ if

- $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent, and
- $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ span $\mathbb{W}$.

◊ 4.8: Example ...

In general, $\mathbf{e}_1 = (1, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{e}_n = (0, \ldots, 0, 1)$ form a basis for $\mathbb{R}^n$. This basis is called the *standard basis* of $\mathbb{R}^n$.

Theorem 4.3 (Basic basis property)**.** *If $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis for $\mathbb{W}$, then every vector in $\mathbb{W}$ can be written uniquely as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$.*

◊ 4.9: proof ...

Example Every vector $(x_1, \ldots, x_n) \in \mathbb{R}^n$ can be written uniquely in terms of $\mathbf{e}_1, \ldots, \mathbf{e}_n$:

$$(x_1, \ldots, x_n) = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$$

◊ 4.10: Examples of bases ...

Dimension (Poole, pp452-456) You will see in Maths IB a proof that although a subspace has (infinitely) many different bases, the *number* of vectors in a basis is always the same. This number is called the *dimension* of the subspace. The dimension of a subspace in $\mathbb{R}^3$ agrees with the common usage of the word dimension. For example, the subspace $\mathbb{W} = \{t(1, 1, 1) \mid t \in R\}$ is a line through (the origin) $\mathbf{0}$ and has dimension one (it has one basis vector)—"lines are 1-dimensional"; and the subspace $\mathbb{V} = \{t(0, 1, 0) + s(0, 0, 1) \mid s, t \in R\}$ has two basis vectors and has dimension two; $\mathbb{V}$ represents a plane through $\mathbf{0}$ and "planes are two-dimensional".

Note: the dimension of the subspace $\mathbb{W} = \{\mathbf{0}\}$ is defined to be zero.

5 Eigenvalues and Eigenvectors

Contents

5.1 Eigenvalue problem (Poole, §4.1,4.3)

In general, for an 2×2 matrix A and a vector $\mathbf{x}$ in $\mathbb{R}^2$, there is no obvious geometric relationship between the vectors $\mathbf{x}$ and $A\mathbf{x}$.

However for certain matrices A and/or vectors $\mathbf{x}$ there is such a relationship.

◊ **5.0: example** . . .

Definition 5.1. Let A be an $n \times n$ matrix. Then a non-zero vector $\mathbf{x}$ in $\mathbb{R}^n$ is called an *eigenvector* of A if $A\mathbf{x}$ is scalar multiple of $\mathbf{x}$, that is

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

for some scalar λ. The scalar λ is called an *eigenvalue* of A, and $\mathbf{x}$ is said to be *an* eigenvector of A corresponding to λ.

- As $A\mathbf{0} = \mathbf{0} = \lambda\mathbf{0}$ for any scalar λ, it is essential that $\mathbf{x} \neq \mathbf{0}$.
- Eigenvalues and eigenvectors are sometimes called characteristic values and characteristic vectors.
- The word "eigen" is German for "proper".

◊ **5.1: How do we find eigenvalues and eigenvectors?** . . .

Method

1. Find all values of λ for which $|\lambda I - A| = 0$.
2. For each eigenvalue λ, solve $(\lambda I - A)\mathbf{x} = \mathbf{0}$ for eigenvectors $\mathbf{x}$.

Definition 5.2. The equation $|\lambda I - A| = 0$ is called the *characteristic equation* of A. When expanded, $|\lambda I - A|$ is a polynomial in λ called the *characteristic polynomial* of A.

Theorem 5.1 (Fundamental Theorem of Algebra)**.** *Let $p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$ be a polynomial with complex coefficients, $a_i \in \mathbb{C}$. Then we can always factor the polynomial as*

$$p(z) = a_n(z - z_1)(z - z_2)\cdots(z - z_n)$$

for some complex numbers $z_1, z_2, \ldots, z_n$ called the roots of the polynomial.

◊ 5.2: Examples ...

In general, for any eigenvalue λ, the set of corresponding eigenvectors, together with $\mathbf{0}$, is a subspace (of $\mathbb{R}^n$) as these vectors are solutions of a system of homogeneous equations

$$(\lambda I - A)\mathbf{x} = \mathbf{0}.$$

Definition 5.3. The set of eigenvectors corresponding to an eigenvalue λ, together with $\mathbf{0}$, is called the *eigenspace* corresponding to λ.

Theorem 5.2. *An eigenspace is a subspace.*

In the above example the eigenspace corresponding to

$$\begin{aligned} \lambda = 1 \quad &\text{is} \quad 1D \quad \text{(one dimensional)} \\ \lambda = 5 \quad &\text{is} \quad 2D. \end{aligned}$$

When solving the characteristic equation $|\lambda I - A| = 0$, it is possible that some of the roots are complex. Corresponding to these complex eigenvalues we would obtain complex eigenvectors; that is, our vectors would be in $\mathbb{C}^n$ = the n-tuples of complex numbers (and not in $\mathbb{R}^n$).

◊ 5.3: Example ...

If we extend our definition for eigenvectors such that $\mathbf{x} \in \mathbb{C}^n$, then the above two vectors could be considered to be eigenvectors corresponding to complex eigenvalues. In this course we restrict attention (almost always) to *real* eigenvalues and eigenvectors.

5.2 Some properties of eigenvalues

The characteristic polynomial of an $n \times n$ matrix A, being of degree n, can be written as the product of n linear factors, by the Fundamental Theorem of Algebra, in the following manner.

$$\begin{aligned} |\lambda I - A| &= \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 && (5.1) \\ &= (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) && (5.2) \end{aligned}$$

the scalars λ_i are the eigenvalues of A.

◊ 5.4: example ...

From the characteristic polynomial we observe the following.

1. Setting $\lambda = 0$ in (5.1) and (5.2)

$$\begin{aligned} c_0 &= (-\lambda_1)(-\lambda_2)(-\lambda_3)\cdots(-\lambda_n) \\ &= |0I - A| = |-A| \\ \Longrightarrow (-1)^n\lambda_1\lambda_2\cdots\lambda_n &= c_0 = (-1)^n|A| \\ \Longrightarrow |A| &= \lambda_1\lambda_2\cdots\lambda_n = (-1)^n c_0\,. \end{aligned}$$

   Thus, the product of eigenvalues equals det A.

2. Multiplying (5.2) out and obtaining the coefficient of λ^{n-1} yields

$$\begin{aligned} c_{n-1} &= -\lambda_1 - \lambda_2 - \cdots - \lambda_n \\ &= -(\lambda_1 + \lambda_2 + \cdots + \lambda_n). \end{aligned}$$

   Also considering the characteristic equation explicitly,

$$\begin{aligned} |\lambda I - A| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix} \\ &= (\lambda - a_{11}) \begin{vmatrix} \lambda - a_{22} & -a_{2n} \\ \vdots & \vdots \\ -a_{n2} & \lambda - a_{nn} \end{vmatrix} \\ &\qquad + \text{polynomial of degree} \leq n-2 \text{ in } \lambda \\ &= (\lambda - a_{11})(\lambda - a_{22})\cdots(\lambda - a_{nn}) \\ &\qquad + \text{polynomial of degree} \leq n-2 \text{ in } \lambda \\ &= \lambda^n - (a_{11} + a_{22} + \cdots + a_{nn})\lambda^{n-1} \\ &\qquad + \text{polynomial of degree} \leq n-2 \text{ in } \lambda \\ \Longrightarrow \lambda_1 + \lambda_2 + \cdots + \lambda_n &= -c_{n-1} \\ &= a_{11} + a_{22} + \cdots + a_{nn} \end{aligned}$$

That is, the sum of all eigenvalues is equivalent to the *trace* of A, $\operatorname{Tr} A$, where the trace of A is the sum of the diagonal elements of A. These properties are summarised in a theorem.

Theorem 5.3. *For an $n \times n$ matrix A,*

- $|A| = \lambda_1 \lambda_2 \cdots \lambda_n = (-1)^n c_0$,
- $\lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn} = -c_{n-1}$.

◊ **5.5: Example** ...

Definition 5.4. An eigenvalue λ_0 of matrix A is said to have *multiplicity* m if precisely m of the λ_i's in the factorisation of the characteristic polynomial are the same as λ_0, that is,

$$|\lambda I - A| = (\lambda - \lambda_0)^m \, g(\lambda)$$

where $g(\lambda)$ is a polynomial of degree $n - m$.

◊ **5.6: Examples** ...

Consequently, counted according to their multiplicity there are precisely n eigenvalues of an $n \times n$ matrix (although some may be complex).

The above properties are for eigenvalues of the matrix A. Let us turn our attention to some properties of the eigenvectors for matrix A. We already know that an eigenspace is a subspace.

We are interested in the dimension of this subspace called an eigenspace.

Theorem 5.4. *The dimension of the eigenspace* $\mathbb{E}_{\lambda_0}$ *corresponding to a given eigenvalue* λ_0 *is less than or equal to the multiplicity of the eigenvalue* λ_0*; that is,*

$$1 \leq \dim \mathbb{E}_{\lambda_0} \leq \textit{ multiplicity of } \lambda_0 \,.$$

◊ **5.7: Examples** ...

Theorem 5.5. *Eigenvectors corresponding to distinct eigenvalues of an* $n \times n$ *matrix are linearly independent.*

◊ **5.8: Proof** ...

◊ **5.9: Examples** ...

5.3 Diagonalisation (Poole, pp303–309)

Definition 5.5. A matrix A is *diagonalisable* if there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix. If so, we say that P *diagonalises* A.

Notation: the diagonal matrix

$$\begin{bmatrix} d_1 & & 0 & 0 \\ & d_2 & & 0 \\ 0 & & \ddots & \\ 0 & 0 & & d_n \end{bmatrix}$$

is often written $\operatorname{diag}(d_1, d_2, \ldots, d_n)$.

◊ **5.10: Example** . . .

Theorem 5.6. *An $n \times n$ matrix A is diagonalisable if and only if it has n linearly independent eigenvectors.*

◊ **5.11: Proof** . . .

Theorem 5.7. *(a) A matrix P which diagonalises A has n linearly independent eigenvectors of A as its columns.*

(b) $D = P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ where λ_i is the eigenvalue corresponding to the eigenvector $\mathbf{v}_i$, the ith column of P.

◊ **5.12: Examples** . . .

Theorem 5.8. *If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalisable.*

◊ **5.13: proof** . . .

5.4 Cayley–Hamilton theorem

(Poole, p300)

Theorem 5.9. *Every $n \times n$ matrix A satisfies its own characteristic equation; that is, if*

$$|\lambda I - A| = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = 0$$

then

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I_n = O.$$

◊ **5.14: Example** . . .

5.5 Applications

Calculating Powers of a Matrix Given a diagonalisable $n \times n$ matrix A, what is A^{100}? The first method is to just multiply it out but the following method is much quicker for large powers.

Let P diagonalise A and set $D = P^{-1}AP = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Thus

$$\begin{aligned} A &= PDP^{-1} \\ \Longrightarrow A^k &= (PDP^{-1})^k \\ &= \underbrace{(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{k \text{ factors}} \\ \Longrightarrow A^k &= PD^kP^{-1}. \end{aligned}$$

We calculate D^k to obtain

$$\begin{aligned} D^k &= \text{diag}(\lambda_1, \ldots, \lambda_n)^k \\ &= \text{diag}(\lambda_1^k, \ldots, \lambda_n^k). \end{aligned}$$

◊ **5.15: Example** ...

5.5.1 Dynamical systems (Poole, pp348–355)

Example 1: the spotted owl

In 1990, environmentalists convinced the U.S. government that continued logging in the old-growth forests on the Pacific coast (with trees over 200 years old) would lead to the extinction of the spotted owl. The timber industry argued that the owl was not a "threatened" species. Mathematical ecologists developed a model to study the population dynamics of the spotted owl, based on data collected from field studies. Most population models concentrate on the female of the species and for the spotted owl the life cycle is divided into three stages: juvenile (up to 1 year old), subadult (1 to 2 years) and adult (over 2 years). The owl mates for life during the subadult and adult stages and lives up to 20 years. Each owl pair requires about 1000 hectares for its own home territory. Critical to the life cycle is when juveniles leave the nest—to survive and become a subadult they must find a mate and a new home range.

The model Consider time intervals of one year denoted by $k = 0, 1, 2, \ldots$ and let $\mathbf{x}_k = (j_k, s_k, a_k)$ where j_k, s_k, a_k are the numbers of females in the juvenile, subadult and adult stages respectively. The transition from year k to year $k+1$ is

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad \text{where } A = \begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix}.$$

The entries in A are determined by the data collected. Row 1 says that the birth rate for an adult pair is 0.33, while the 0's say that neither juveniles nor subadults give birth. In Row 2, the 0.18 entry indicates the number of juveniles who survive to become subadults (actually 60% survive to leave the nest but only 30% of these find home ranges—this is the crucial figure!).

Question For large values of k, what happens to the entries in $\mathbf{x}_k$?

$\Diamond$ **5.16: Solution** ...

Example 2

Consider a simplified predator-prey system in which wood rats are the main diet for owls. At time k, measured in months, $\mathbf{x}_k = (O_k, R_k)$ measures the population of owls O_k and R_k the number of rats in thousands.

Suppose that the relationship between the owls and rats is

$$\begin{aligned} O_{k+1} &= (0.5)O_k + (0.4)R_k, \\ R_{k+1} &= (-0.104)O_k + 1.1R_k. \end{aligned}$$

The diagonal entries can be interpreted as: 0.5 indicates if there were no rats, the owl population would halve each month; whereas 1.1 says the rat population would increase by 10% without owls as predators.

This dynamical system is written as

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \quad A = \begin{bmatrix} 0.5 & 0.4 \\ -0.104 & 1.1 \end{bmatrix}.$$

We use eigenvalues and eigenvectors to analyse the long term behaviour of the system.

$\Diamond$ **5.17: Calculations** ...

5.5.2 Markov chains (Poole, pp230–235,325–330)

Example 1: population movement

Suppose that at a certain time 20% of the inhabitants of a city and its suburbs live in the city with 80% in the suburbs. If each year 5% of the population in the suburbs move to the city and 3% of those in the city move to the suburbs. Assuming that other factors such as births, deaths, migration do not affect these proportions, what will be the long term distribution of the population?

Let $\mathbf{x}_0 = (0.2, 0.8)$ which represents the proportions of the population in the city and suburbs at time $k = 0$. The transition from one year to the next can be given as a 2×2 matrix:

$$A = \begin{array}{cl} \text{From:} & \text{To:} \\ \begin{array}{cc} \text{City} & \text{Suburbs} \end{array} & \\ \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix} & \begin{array}{l} \text{City} \\ \text{Suburbs} \end{array} \end{array}$$

For $k = 0, 1, 2, 3, \ldots$ we have $\mathbf{x}_{k+1} = A\mathbf{x}_k$.

For example, $\mathbf{x}_1 = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.8 \end{bmatrix} = \begin{bmatrix} .234 \\ .766 \end{bmatrix}$ which gives the proportions of the population in the city and suburbs one year later.

This is an example of a Markov Chain. In a Markov chain, the vectors $\mathbf{x}_k$ are called *state* vectors and have the property that each entry lies between 0 and 1 and their sum is one. Vectors with this property are called *probability* vectors. The matrix A is a square matrix all of whose columns are probability vectors; such a matrix A is called a *stochastic* matrix.

In most cases of Markov chains, determining the long-term behaviour is equivalent to finding the steady state-vector. For a Markov Chain with stochastic matrix A, the *steady state vector* $\mathbf{q}$ is a probability vector which satisfies $A\mathbf{q} = \mathbf{q}$.

The vector $\mathbf{q}$ is an eigenvector for the matrix A (with corresponding eigenvalue 1).

◊ **5.18: Solution** . . .

Example 2: Adelaide's weather

Suppose that there are three states for the local weather: sunny, cloudy and raining. If today is sunny, the probability tomorrow is sunny is 0.7, cloudy 0.2 and raining 0.1; while if today is cloudy the

probability tomorrow is sunny is 0.5, cloudy 0.3 and raining 0.2; and if today it is raining then tomorrow will be sunny with a probability of 0.2, cloudy with a probability of 0.5 and still raining with a probability of 0.3.

Let $\mathbf{x}_k = (s_k, c_k, r_k)$ is the probability vector which gives the probabilities in k days time the weather is sunny (s_n), cloudy (c_k) and raining (r_k) then this Markov Chain is

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \quad A = \begin{bmatrix} 0.7 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.1 & 0.2 & 0.3 \end{bmatrix}.$$

◊ **5.19: Calculations** ...