

# MATH1331 Linear Algebra with Applications

## Cheatsheet

2021/22

This document collects together the important definitions and results presented throughout the lecture notes. The numbering of the sections will be consistent with that in the lecture notes.

### Contents

<b>2</b>	<b>Linear Equations</b>	<b>2</b>
<b>3</b>	<b>Feasible Sets</b>	<b>4</b>
<b>4</b>	<b>Using Echelon Forms to Solve Equations</b>	<b>5</b>
<b>5</b>	<b>Vectors</b>	<b>6</b>
<b>6</b>	<b>Matrix Algebra</b>	<b>8</b>
<b>7</b>	<b>Determinants</b>	<b>10</b>
<b>8</b>	<b>Eigenvalues and Eigenvectors</b>	<b>12</b>
<b>9</b>	<b>Linear Programming</b>	<b>14</b>
<b>10</b>	<b>Game Theory</b>	<b>16</b>
<b>11</b>	<b>Markov Processes</b>	<b>19</b>

## 2 Linear Equations

### 2.1 Equations and Inequalities

**Definition 2.1** An expression of the form  $x + 3y = 7$  is an **equation/equality**. An expression of the form  $x + 3y \leq 7$  is an **inequality**, whereas  $x + 3y < 7$  is a **strict inequality**.

### 2.3 Matrix Notation for Equations

**Definition 2.3** A **matrix** is a rectangular array of numbers or expressions of the form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The matrix above is an  $m \times n$  matrix, which has  $m$  rows and  $n$  columns.

### 2.4 Operations on a Matrix

**Definition 2.4** Let  $A$  be a matrix. The **leading entries** are the first non-zero elements in each row. It is said that  $A$  is in **row echelon form (REF)** if the following are true:

- (i) Each non-zero row begins with more zeros than the rows above it.
- (ii) Every all-zero row appears at the bottom of the matrix.

**Definition 2.5** A matrix  $A$  is in **reduced row echelon form (RREF)** if the following are true:

- (i)  $A$  is in row echelon form.
- (ii) Each leading entry is 1.
- (iii) Every entry in a column containing a leading one is zero (except for the leading one).

**Note:** There are three types of **row operation** we allow ourselves to perform on a matrix.

- (i) Adding and subtracting one row to/from another, e.g.  $R1 \rightarrow R1 - 3R2$ .
- (ii) Multiplying a row by a constant, e.g.  $R3 \rightarrow 2R3$ .
- (iii) Swapping two rows, e.g.  $R2 \leftrightarrow R4$ .

**Definition 2.7** The process of **Gaussian elimination** transforms a matrix into a REF. The process of **Gauss-Jordan elimination** transforms a matrix into the RREF.

**Method – Gaussian Elimination:** We can find a REF of a matrix  $A$  as follows:

- (i) If the first entry in  $R_1$  is zero, swap the row with one whose first entry is non-zero. Otherwise, if the first entry in  $R_1$  is already non-zero, move to Step (ii).
- (ii) Multiply  $R_1$  by a constant so that the first entry becomes 1.
- (iii) Clear below it by adding/subtracting multiples of  $R_1$  to/from the other rows.
- (iv) Repeat Steps (i)-(iii) with the next rows, until all rows have been covered.

**Method – Gauss-Jordan Elimination:** We can find the RREF of a matrix  $A$  as follows:

- (i) Transform  $A$  into a REF by applying the above method.
- (ii) Starting at the rightmost column with a leading entry, clear above it.
- (iii) Repeat Step (ii) by moving right-to-left through the leading entries.

**Method – Solving Systems of Linear Equations:** To solve a system of equations, we form the matrix of coefficients and augment it by adding one final column; this column contains the numbers that each equation in the system is equal to. We can then do Gaussian/Gauss-Jordan elimination and read-off the solutions.

**Note:** Suppose we write our system of equations as an augmented matrix, in its RREF.

- If there are as many leading ones as there are columns, we have a **unique solution**.
- If there are less leading ones as there are columns, then the variables corresponding to the columns without said leading ones are **free**; this means they can be any real number. This means there are **infinitely-many solutions**.
- If there is a row with zeros to the left of the augmented column but with a non-zero in the augmented column, then there is **no solution**.

## 3 Feasible Sets

### 3.1 Notation

**Definition** The **set of  $n$ -tuples of real numbers** is this:  $\mathbb{R}^n = \{(x_1, \dots, x_n) : \text{each } x_i \in \mathbb{R}\}$ . The familiar ones are  $\mathbb{R}^2$ , which are just pairs of real numbers (these can be plotted in a plane, i.e. on a pair of axes) and  $\mathbb{R}^3$ , which are triples of real numbers (these can be plotted in three-dimensional space, i.e. on a set of three axes).

**Note:** Given a collection of inequalities, the **feasible set** is the region in which **all** inequalities are satisfied. Throughout, we consider maximising/minimising problems sketched in  $\mathbb{R}^2$ .

**Method – Sketching the Feasible Set:** To find the feasible set, we first sketch the **lines** that define the inequalities (i.e. replace the  $\leq, \geq$  with  $=$  and draw the lines). From here, we can determine which side of each line we need to be on to satisfy the given inequality. The location in which all inequalities are satisfied is then the feasible set.

**Method – Finding Maxima/Minima:** Suppose we wish to maximise/minimise an expression subject to some inequalities. We can do just this by applying the above method.

- (i) Find the feasible set satisfying all the inequalities.
- (ii) Note the vertices of the feasible set, i.e. all intersection points of the lines that form the boundary of the feasible set (these are the vertices of the feasible set).
- (iii) The maximum/minimum will always occur at one of these vertices, so substitute each in and find which gives the highest/lowest value.

## 4 Using Echelon Forms to Solve Equations

### 4.1 Using the Reduced Echelon Form to Solve Equations

**Definition** Suppose  $A$  is an augmented matrix representing a system of equations. Variables which correspond to a leading one are called **leading variables**. Otherwise, they are called **free variables**.

**Method – Solving Systems of Equations:** Let  $A$  be an augmented matrix representing a system of equations in RREF.

- (i) If the final column contains a leading entry, the system is inconsistent (no solution).
- (ii) If the final column **doesn't** contain a leading entry, solutions are obtained by setting the free variables equal to arbitrary parameters and using the equations to determine values of the leading variables in terms of these parameters.
  - If there are free variables, there are infinitely-many solutions.
  - If there are **no** free variables, there is a unique solution.

### 4.2 Homogeneous Equations

**Definition** A **homogeneous equation** is one where the right-hand-side is zero. Therefore, a system of homogeneous equations is one in which every equation is equal to zero.

**Note:** Suppose we have a system of equations, but some of the coefficients involve an unknown constant  $k$ . We can still convert everything into matrix language and apply row operations, but we have to be careful. If our constant  $k = 0$ , then we can **not** divide by it. Additionally, we shouldn't be multiplying rows by zero. Thus, all row operations we do must avoid multiplying/dividing by the constant  $k$ .

## 5 Vectors

$\mathbb{R}^n$  is an example of a **Euclidean vector space**. We introduce two notions of a real vector in  $\mathbb{R}^n$ :

$$\text{row vectors } \mathbf{v} = (v_1, \dots, v_n) \quad \text{and} \quad \text{column vectors } \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

Really, these are  $1 \times n$  matrices and  $n \times 1$  matrices, respectively. We can define addition and scalar multiplication to be component-wise.

**Note:** More generally, a **real vector space** is a set  $V$  which contains elements called **vectors** and has the following structure: we can add together two vectors and we can multiply a vector by any real number. Doing these operations should still give us an element of our set  $V$ .

**Definition 5.1** A **subspace** of  $\mathbb{R}^n$  is a subset  $W \subseteq \mathbb{R}^n$  satisfying these properties:

- (i) The zero vector  $\mathbf{0} \in W$ .
- (ii) For all  $\mathbf{v}, \mathbf{w} \in W$ , we have  $\mathbf{v} + \mathbf{w} \in W$ .
- (iii) For all  $\mathbf{v} \in W$  and  $k \in \mathbb{R}$ , we have  $k\mathbf{v} \in W$ .

**Note:** If we have a system of linear homogeneous equations in  $m$  variables, then the set of solutions is always a subspace of  $\mathbb{R}^m$ .

**Definition 5.3** Let  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$  be a collection of vectors. Then, a **linear combination** is the vector  $a_1\mathbf{v}_1 + \dots + a_k\mathbf{v}_k$ , where each coefficient  $a_1, \dots, a_k \in \mathbb{R}$  is a scalar.

**Definition** Let  $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$  be a collection of vectors. Then, the **span** of these vectors is the subspace  $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$  which contains **all** possible linear combinations of these vectors. This is the smallest subspace containing each of the vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$ .

**Method – Spanning vs Non-Spanning:** Given a collection of vectors, to see if they span  $\mathbb{R}^n$ , we proceed as follows:

- (i) Form a matrix whose columns are the vectors.
- (ii) Find a REF of the matrix.
- (iii) Count the number of leading entries (this is the rank of the matrix; Definition 5.8).

If the number of leading entries is  $n$  (which is the maximum it can be), the vectors are spanning. If the number of leading entries is **less than**  $n$ , the vectors are non-spanning.

**Definition 5.5** A set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subseteq \mathbb{R}^n$  is called **linearly independent** if the only solution to  $a_1\mathbf{v}_1 + \dots + a_k\mathbf{v}_k = \mathbf{0}$  is  $a_1 = \dots = a_k = 0$ . If not, they are **linearly dependent**.

**Method – Linear Independence vs Linear Dependence:** Given a collection of vectors, to see if they are linearly (in)dependent, form a matrix whose columns are the vectors and determine a REF of the matrix. If there are **no** all-zero rows, then they are linearly independent. If there is a zero row, then they are linearly dependent.

**Note:** If we have two vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , they are linearly independent if and only if they are **not** scalar multiples of each other. This is true in **any** of the vector spaces  $\mathbb{R}^n$ .

**Definition 5.6** A **basis** of a subspace  $W \subseteq \mathbb{R}^n$  is a set of linearly independent vectors which spans  $W$ . For  $\mathbb{R}^n$  itself, there is the so-called **standard basis** containing  $n$  vectors, each of which is all-zero bar a 1 in one place (the first position, the second position, etc.):

$$\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^n.$$

**Definition 5.7** The **dimension** of a subspace  $W \subseteq \mathbb{R}^n$  is simply the number of vectors in a basis of  $W$  (this is powerful because this is always the same number for **any** of its bases). We denote the dimension as  $\dim(W)$ .

## 5.1 Testing Spanning, Linear Independence and Basis

**Definition 5.8** The **rank** of a matrix is the dimension of the subspace spanned by its rows.

**Method – Rank of a Matrix:** As mentioned in the previous method, the rank is simply the number of leading entries in a REF of the matrix, so this is really just an application of Gaussian elimination.

**Note:** The non-zero rows of a REF form a basis of the subspace spanned by the rows.

**Method – Basis of a Solution Set of a Homogeneous System:** Solve our homogeneous system of equations (using previous methods), from which we may see that the general solution is of the form  $\mathbf{x} = a_1\mathbf{v}_1 + \dots + a_k\mathbf{v}_k$ . Hence, the solutions are spanned by  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ , which are each non-zero. We then look at linear (in)dependence to either conclude that this is a basis, or that we can make a basis by throwing away any vectors that depend on others.

## 6 Matrix Algebra

### 6.1 Operations on Matrices

We will explain the operations we can do with matrices.

- **Scalar multiplication.** If we have an  $m \times n$  matrix  $A = (a_{ij})$  and a scalar  $k \in \mathbb{R}$ , then we can form the matrix  $kA = (ka_{ij})$ , meaning that the entries of the matrix  $kA$  are found by multiplying each of the entries of  $A$  by the scalar  $k$ .
- **Addition.** If we have two  $m \times n$  matrices  $A = (a_{ij})$  and  $B = (b_{ij})$  of the **same size**, then we can form the matrix  $A + B = (a_{ij} + b_{ij})$ , meaning that the entries of the matrix  $A + B$  are found by adding the corresponding entries of  $A$  and  $B$ .
- **Multiplication.** If we have an  $m \times n$  matrix  $A = (a_{ik})$  and an  $n \times p$  matrix  $B = (b_{kj})$ , then we can form the  $m \times p$  matrix  $AB = (\sum_{k=1}^n a_{ik}b_{kj})$ , meaning that the  $ij^{\text{th}}$  entry of the matrix  $AB$  is found by adding the multiplies of the  $i^{\text{th}}$  row of  $A$  and the  $j^{\text{th}}$  column of  $B$ .

### 6.2 Linear Maps

**Definition 6.1** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a map between two Euclidean vector spaces. It is called a **linear map** (or **linear transformation**) if these are true, for all  $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and  $k \in \mathbb{R}$ :

- $f(\mathbf{v} + \mathbf{w}) = f(\mathbf{v}) + f(\mathbf{w})$ .
- $f(k\mathbf{v}) = kf(\mathbf{v})$ .

**Theorem** Any linear map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is just  $f(\mathbf{x}) = A\mathbf{x}$ , for some  $m \times n$  matrix  $A$ .

**Method – Finding the Matrix of a Linear Map:** To represent a linear map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  by a matrix, we need to see what this map  $f$  does to the standard basis vectors  $\mathbf{e}_1, \dots, \mathbf{e}_n$  of  $\mathbb{R}^n$ . Indeed, if we compute  $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$ , we will get some vectors in  $\mathbb{R}^m$ . If we put these vectors as **columns** of a matrix, we will have constructed  $A$  as in the above theorem.

**Note:** The **identity matrix**  $I_n$  is the  $n \times n$  matrix with ones on the main diagonal and zeros elsewhere.

### 6.3 Inverse Matrices

**Definition** Let  $A$  be an  $n \times n$  matrix. Then, its **inverse** (if it exists) is the matrix  $B$  which is such that  $AB = I_n$  and  $BA = I_n$ . To remind us that the matrix  $B$  is the inverse of  $A$ , we often relabel it as  $A^{-1}$ . In this case, we say that  $A$  is **invertible**.



**Method – Finding the Inverse of a Matrix:** To find the inverse of the  $n \times n$  matrix  $A$ , if it exists, we form the augmented matrix  $(A \mid I_n)$  and apply to it Gauss-Jordan elimination; this will get it into the form  $(I_n \mid A^{-1})$ , from which we can read-off the inverse matrix. In the case that we get a zero row on the left of the vertical line during this process, we conclude that  $A$  is **not** invertible.

## 6.4 An Application of Inverse Matrices

**Method – Solving Systems of Equations with Inverse Matrices:** Suppose we wish to solve a system of equations. If we first convert it into a matrix problem, it will have the form  $A\mathbf{x} = \mathbf{b}$ , where  $A$  is the coefficient matrix,  $\mathbf{x}$  is the vector of variables and  $\mathbf{b}$  is the vector of constants. If we know that  $A$  is invertible, then  $\mathbf{x} = A^{-1}\mathbf{b}$  is the unique solution.

## 6.5 Error Analysis

**Definition** We perform **error analysis** if the constants in a system of equations are measured to within a given degree of accuracy/error.

**Method – Error Analysis:** Suppose we wish to solve the system of  $m$  equations described by  $A\mathbf{x} = \mathbf{b}$ , but the elements of  $\mathbf{b} \in \mathbb{R}^m$  have errors  $e_1, \dots, e_m$ . Then, we can compute the errors in the solution as follows:

$$\varepsilon_i = \sum_{j=1}^m a_{1j}e_j.$$

## 6.6 Input-Output Analysis (The Leontief Model)

**Definition** The **Leontief model** describes how output from one industrial sector may become an input for another industrial sector. This is described using a matrix  $A = (a_{ij})$  which tells us that the  $j^{\text{th}}$  sector requires  $a_{ij}$ -units from the  $i^{\text{th}}$  sector to produce one unit. The **final demand** in the  $i^{\text{th}}$  sector is the number of units sold to customers, denoted  $d_i$ .

**Note:** Let  $A$  be the input-output matrix from the Leontief model,  $\mathbf{x}$  the vector representing the total output and  $\mathbf{d}$  be the final demand vector (with entries  $d_i$ ). Then,  $\mathbf{x} = A\mathbf{x} + \mathbf{d}$ .

**Method – Production to Meet Demand:** Suppose we have a Leontief equation  $\mathbf{x} = A\mathbf{x} + \mathbf{d}$  and we are told to calculate how much each sector must produce to meet a final demand.

- (i) Rearrange the Leontief equation to  $(I - A)\mathbf{x} = \mathbf{d}$ .
- (ii) Compute the inverse  $(I - A)^{-1}$ .
- (iii) Use Step (ii) to rearrange the equation in Step (i) to  $\mathbf{x} = (I - A)^{-1}\mathbf{d}$ .

## 7 Determinants

**Definition** The **determinant** of a square matrix is a scalar value (number) which is found by considering the entries of the matrix. If  $A$  is a matrix, then we either write the determinant as  $\det(A)$  or by writing out the matrix with vertical lines  $|\cdot|$  instead of brackets  $(\cdot)$ .

**Note:** The determinant of a general  $2 \times 2$  matrix  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$  is the number  $ad - bc$ .

**Method – Inverse of a  $2 \times 2$  Matrix:** Suppose we wish to find the inverse of  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ .

- (i) Compute the determinant  $\det(A) = ad - bc$ .
- (ii) Swap the main diagonal entries.
- (iii) Multiply the off-diagonal entries by  $-1$ .
- (iv) Divide everything by the determinant.

This provides us with the following formula for the inverse:  $A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ .

### 7.1 An Application of Determinants

**Theorem (Cramer's Formula)** Let  $A\mathbf{x} = \mathbf{b}$  represent a system of  $n$  equations in  $n$  variables. If the matrix  $A_{i|\mathbf{b}}$  denotes the matrix  $A$  except with the  $i^{\text{th}}$  column replaced by the vector  $\mathbf{b}$ , then the system has solutions  $x_i = \det(A_{i|\mathbf{b}}) / \det(A)$ .

### 7.3 General Determinants

**Definition** Let  $A$  be a matrix. The  $ij^{\text{th}}$  **minor**  $A_{ij}$  is the determinant of the sub-matrix formed by deleting the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column. The  $ij^{\text{th}}$  **cofactor** is then  $(-1)^{i+j} A_{ij}$ .

**Note:** The determinant of an  $n \times n$  matrix  $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$  is this:  $\sum_{j=1}^n (-1)^{1+j} a_{1j} A_{1j}$ .

**Remark** We just need to pick a row and move along it, computing the minors at each step. We then multiply each minor by the number in that position of the matrix and then do an alternating sum (“plus, minus, plus, minus, etc.” **or** “minus, plus, minus, plus, etc.”). In fact, the formula above is the expansion along the first row, but really we could do it along **any** row, or indeed down **any** column.

## 7.4 Properties of Determinants

**Definition** The **transpose** of a matrix  $A = (a_{ij})$  is the matrix  $A^T = (a_{ji})$ , that is the matrix where we have swapped the columns and rows.

**Proposition** *These are some important properties of the determinant:*

- (i) *The transpose doesn't change the determinant, that is  $\det(A^T) = \det(A)$ .*
- (ii) *If  $A$  has two identical rows, then  $\det(A) = 0$ .*
- (iii) *If  $A$  has one row which is a multiple of the other, then  $\det(A) = 0$ .*
- (iv) *Applying the row operation  $R_i \leftrightarrow R_j$  changes the determinant to  $-\det(A)$ .*
- (v) *Applying the row operation  $R_i \rightarrow kR_i$  changes the determinant to  $k\det(A)$ .*
- (vi) *Applying the row operation  $R_i \rightarrow R_i + kR_j$  does **not** change the determinant.*

**Method – Determinant of an  $n \times n$  Matrix:** To find the determinant of an  $n \times n$  matrix in an 'easy' way, we do the following:

- (i) Apply Gaussian elimination to get a REF.
- (ii) Use (iv) and (v) from the above proposition to see how our determinant changes.
- (iii) The determinant of a REF is then the product of the main diagonal entries.
- (iv) The determinant of the original matrix is then the determinant of the REF, combined with undoing any sign changes or row multiplications done in Step (i).

## 7.6 Further Facts about Determinants

**Theorem** *Let  $A$  be a matrix. Then, the following are equivalent:*

- (i) *The determinant  $\det(A) = 0$ .*
- (ii) *The matrix  $A$  is **not** invertible.*
- (iii) *The rank  $\text{rank}(A) < n$ .*
- (iv) *The row reduction will contain an all-zero row.*
- (v) *The rows are **not** linearly independent.*

## 8 Eigenvalues and Eigenvectors

**Definition 8.1** Let  $A$  be an  $n \times n$  matrix. Then, an **eigenvalue** of  $A$  is a real number  $\lambda \in \mathbb{R}$  such that there exists a non-zero vector  $\mathbf{x} \in \mathbb{R}^n$ , called an **eigenvector**, such that  $A\mathbf{x} = \lambda\mathbf{x}$ .

**Theorem 8.2** The number  $\lambda$  is an eigenvalue of  $A$  if and only if  $\det(A - \lambda I) = 0$ .

*Proof:* Now,  $\lambda$  is an eigenvalue of  $A$  if and only if the system  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  has a solution, just by rearranging the equation in Definition 8.1. However, this is true if and only if the matrix  $(A - \lambda I)$  is **not** invertible, which means  $\det(A - \lambda I) = 0$  by the last theorem in Chapter 7.  $\square$

**Note:** The **characteristic polynomial** for a matrix  $A$  is  $p_A(t) = \det(A - tI)$ . Hence, Theorem 8.2 tells us that we can find the eigenvalues by equating the characteristic polynomial to zero (this is called the **characteristic equation**) and solving for  $t$ .

**Method – Finding the Eigenvectors of a Matrix:** Let  $A$  be a square matrix.

- (i) Find the eigenvalues  $\lambda_1, \dots, \lambda_k$  of  $A$  by solving  $\det(A - \lambda I) = 0$ .
- (ii) Choose one of the eigenvalues, say  $\lambda_1$ .
- (iii) Solve the system  $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$ ; we should get a family of solutions.
- (iv) Repeat this for each of the other eigenvalues  $\lambda_2, \dots, \lambda_k$  in Step (i).

**Definition** A matrix  $A$  is **diagonalisable** if there exists a diagonal matrix  $D$  (i.e. one where the only non-zero entries are on the main diagonal) and invertible matrix  $P$  such that  $P^{-1}AP = D$ .

**Method – Diagonalising a Matrix:** We want to diagonalise  $A$  (assuming it possible).

- (i) Find the eigenvalues  $\lambda_1, \dots, \lambda_k$  of  $A$ .
- (ii) Find the corresponding eigenvectors  $\mathbf{x}_1, \dots, \mathbf{x}_k$  using the previous method.
- (iii) The diagonal matrix  $D$  is that with the eigenvalues on the diagonal.
- (iv) The invertible matrix  $P$  is that with the eigenvectors as its columns.

We need to have the eigenvalues and corresponding eigenvectors in the same order, so

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{pmatrix} \quad \text{means that} \quad P = \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}.$$

**Method – Powers of a Matrix:** Suppose we wish to find  $A^m$  for a huge integer  $m \in \mathbb{Z}^+$ .

- (i) Diagonalise the matrix, so that  $P^{-1}AP = D$ .
- (ii) Rearrange the equation in Step (i) so that we have  $A = PDP^{-1}$ .
- (iii) Notice that  $A^m = (PDP^{-1})^m = PD^mP^{-1}$ .
- (iv) Finally, the entries of  $D^m$  are just the entries of  $D$  raised to the power of  $m$ .

## 8.1 What Can Go Wrong in Computing Eigenvalues and Eigenvectors?

**Note:** There are a number of problems that can arise if we want to diagonalise/when finding eigenvalues and eigenvectors.

- It may be that the characteristic equation has no real roots, e.g.  $\lambda^2 + 1 = 0$ .
- It may be that an eigenvalue is repeated (we call the number of times it appears as a root of the characteristic polynomial the **algebraic multiplicity**), so there may not be enough eigenvectors to form a basis (i.e to make the matrix  $P$  from them).

## 9 Linear Programming

### 9.1 Simplex Algorithm

**Definition** A **maximising linear programming problem** is a problem where we wish to maximise a function  $c_1x_1 + \dots + c_mx_m$  subject to the following inequalities:

$$\begin{array}{rcl} a_{11}x_1 + \dots + a_{1m}x_m & \leq & b_1, & x_1 \geq 0, \\ a_{21}x_1 + \dots + a_{2m}x_m & \leq & b_2, & x_2 \geq 0, \\ & & \vdots & \vdots \\ a_{n1}x_1 + \dots + a_{nm}x_m & \leq & b_n, & x_m \geq 0, \end{array}$$

where the numbers  $a_{ij}, b_i, c_j \in \mathbb{Z}$  are all integers.

**Definition** A **slack variable** is a variable  $u_i \geq 0$  such that, when added to an inequality, it gives equality, e.g. the inequalities in the above definition become

$$\begin{array}{rcl} a_{11}x_1 + \dots + a_{1m}x_m + u_1 & = & b_1, \\ a_{21}x_1 + \dots + a_{2m}x_m + u_2 & = & b_2, \\ & & \vdots \\ a_{n1}x_1 + \dots + a_{nm}x_m + u_n & = & b_n. \end{array}$$

**Definition** The **maximising variable**  $M$  is the variable defined as  $M = c_1x_1 + \dots + c_mx_m$  (this is what we are wanting to maximise).

**Definition** The **simplex tableau** is an augmented matrix which encodes information about a linear programming problem. Indeed, the first  $n$  rows represent each of the  $n$  inequalities in the problem (where we have introduced the slack variables to convert them to equalities) and the bottom row represents the maximising variable:

$$\left( \begin{array}{cccccccc|c} a_{11} & \cdots & a_{1m} & 1 & 0 & 0 & \cdots & 0 & b_1 \\ a_{21} & \cdots & a_{2m} & 0 & 1 & 0 & \cdots & 0 & b_2 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \cdots & 0 & \vdots \\ a_{n1} & \cdots & a_{nm} & 0 & \cdots & 0 & 1 & 0 & b_n \\ \hline -c_1 & \cdots & -c_m & 0 & \cdots & 0 & 0 & 1 & 0 \end{array} \right)$$

**Note:** A simplex tableau is in **standard form** if the last column is non-negative; all  $b_i \geq 0$ .

**Definition** For a simplex tableau, a **Type II variable** is one whose columns consist of a single 1 and the rest zeros. Every other variable is called a **Type I variable**.

**Method – The Simplex Algorithm:** We consider the maximisation in the first definition.

- (i) Construct the simplex tableau.
- (ii) Find the most negative entry in the final row to the left of the vertical line (**not** the final column); if it is repeated, choose the **leftmost** one. This is the **pivot column**.
- (iii) For every **positive** entry in the pivot column above the horizontal line (**not** in the bottom row), we compute the ratio [entry in last column]  $\div$  [entry in pivot column]. The row corresponding to the smallest ratio is the **pivot row**.
- (iv) We now look at the entry which is in both the pivot column and pivot row; we clear the column (get zeros above and below it).
- (v) Repeat Steps (ii)–(iv) so all entries in the final row to the left of the line are positive.
- (vi) Set all Type I variables in the final tableau to zero and solve for the Type II variables.

**Note:** Let's change the inequality  $a_{k1}x_1 + \dots + a_{km}x_m \leq b_k$  to  $a_{k1}x_1 + \dots + a_{km}x_m \leq \beta_k$ . Instead of repeating the Simplex Algorithm all over again, we can apply a column operation to the final column (the one with  $b_i$ 's in it) to change the entry  $b_k$  to the new value  $\beta_k$ . Indeed, we just need to add to the final column a multiple of the column representing the slack variable  $u_k$  for this equation. Explicitly, we do  $C_{m+n+2} \rightarrow C_{m+n+2} + (\beta_k - b_k)C_{m+k}$ .

### 9.3 Problems Not in Standard Form (Broken Tableaux)

**Definition** A simplex tableau is called **broken** if at least one of the  $b_i < 0$ .

**Method – 'Fixing' a Broken Simplex Tableau:** Suppose we have a broken simplex tableau.

- (i) Find the **uppermost** negative entry in the final column and go along to the **leftmost** negative entry in that row. This is the **pivot column**.
- (ii) Compute the ratios as in Step (iii) in the Simplex Algorithm and find the row corresponding to the smallest **positive** ratio; this is the **pivot row**.
- (iii) Find the entry in both the pivot column and pivot row; clear the column.
- (iv) Repeat Steps (i)–(iii) until the simplex tableau is in standard form.

**Method – Solving a Broken Simplex Tableau:** 'Fix' it and apply the Simplex Algorithm.

**Note:** If we are asked to **minimise** a function  $c_1x_1 + \dots + c_mx_m$ , then we can simply **maximise** the function  $-c_1x_1 - \dots - c_mx_m$  using the usual Simplex Algorithm. We must remember to **negate** the answer to make sure that we indeed have a minimum. Moreover, if  $a_{k1}x_1 + \dots + a_{km}x_m \geq b_k$ , we convert it to  $-a_{k1}x_1 - \dots - a_{km}x_m \leq -b_k$ . It may be that the tableau is broken, so we 'fix' it and then use the Simplex Algorithm.

## 10 Game Theory

### 10.3 Games – Introduction

A **pay-off matrix** is an array whose rows and columns both represent outcomes of a game involving two participants. As such, any game will have two pay-off matrices, one each from the perspectives of the people playing the game.

**Definition** A **saddle point** in a pay-off matrix is an entry which is the minimum in its row but also the maximum in its column. If a pay-off matrix contains at least one saddle point, then the game is called **strictly determined**.

**Note:** If a pay-off matrix has a saddle point, the corresponding row is the **optimal pure strategy** for the player represented by the rows. Similarly, the corresponding column is the **optimal pure strategy** for the player represented by the columns. The **pay-off** (or **value**) of the game is then that saddle number.

### 10.4 Mixed Strategies

**Definition** If a pay-off matrix contains **no** saddle points, then there are **no** optimal pure strategies, so we must turn to a **mixed strategy**. This is the strategy that one player should adopt based on what the other player has chosen. For instance, if  $A$  is an  $m \times n$  pay-off matrix, the row vector  $\mathbf{r} = (r_i)$  is the strategy played by the row-player and, similarly, the column vector  $\mathbf{c} = (c_j)$  is the strategy played by the column-player, then the **expected value of the strategies** is

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \in \mathbb{R}.$$

### 10.6 What Should Holmes Do?

**Lemma** For  $a, b \in \mathbb{R}$ , the **minimum** of  $c_1a + c_2b$  subject to the conditions that  $c_1 \geq 0$  and  $c_2 \geq 0$  and  $c_1 + c_2 = 1$  is precisely  $\min(a, b)$ .

*Proof:* (i) If  $a \leq b$ , we can write  $c_1a + c_2b = (c_1 + c_2)a + c_2(b - a)$ . If we impose the condition that  $c_1 + c_2 = 1$ , this becomes  $a + c_2(b - a)$ . Because  $a \leq b$ , we know that  $b - a$  is non-negative, so this achieves a minimum of  $a$  precisely when  $c_2 = 0$ , and thus  $c_1 = 1$ .

(ii) If  $a > b$ , we can write  $c_1a + c_2b = c_1(a - b) + (c_1 + c_2)b$ . If we impose the condition that  $c_1 + c_2 = 1$ , this becomes  $c_1(a - b) + b$ . Again, because  $a > b$ , we know that  $a - b$  is positive, so this achieves a minimum of  $b$  precisely when  $c_1 = 0$ , and thus  $c_2 = 1$ .  $\square$



**Lemma** For  $a, b \in \mathbb{R}$ , the **maximum** of  $c_1a + c_2b$  subject to the conditions that  $c_1 \geq 0$  and  $c_2 \geq 0$  and  $c_1 + c_2 = 1$  is precisely  $\max(a, b)$ .

*Sketch of Proof:* The proof is similar to the proof of the previous lemma. □

**Note:** The above lemmas are mainly used when solving a mixed strategy problem.

## 10.8 Optimal Mixed Strategy: General Case

**Definition** The **optimal mixed strategy** for a player is that where the expected value with the best counter-strategy for the opposing player is as large as possible.

**Method – The Row Optimal Mixed Strategy:** Let the matrix  $A$  have **no** saddle points.

(i) Find the **minimum** of  $y_1 + \cdots + y_m$  subject to the following inequalities:

$$\begin{array}{rcl} a_{11}y_1 + a_{21}y_2 + \cdots + a_{m1}y_m & \geq & 1, & y_1 & \geq & 0, \\ a_{12}y_1 + a_{22}y_2 + \cdots + a_{m2}y_m & \geq & 1, & y_2 & \geq & 0, \\ & & \vdots & & & \vdots \\ a_{1n}y_1 + a_{2n}y_2 + \cdots + a_{mn}y_m & \geq & 1, & y_m & \geq & 0. \end{array}$$

(ii) Compute  $v = 1/(y_1 + \cdots + y_m)$ , the reciprocal of the minimum from Step (i).

(iii) The optimal mixed strategy is the row  $(r_1, r_2, \dots, r_m) = v(y_1, y_2, \dots, y_m)$ .

**Remark** The left-hand side of each of the inequalities are just the entries of this vector:

$$\begin{pmatrix} y_1 & y_2 & \cdots & y_m \end{pmatrix} A.$$

**Method – The Column Optimal Mixed Strategy:** Let the matrix  $A$  have **no** saddle points.

(i) Find the **maximum** of  $z_1 + \cdots + z_n$  subject to the following inequalities:

$$\begin{array}{rcl} a_{11}z_1 + a_{12}z_2 + \cdots + a_{1n}z_n & \leq & 1, & z_1 & \geq & 0, \\ a_{21}z_1 + a_{22}z_2 + \cdots + a_{2n}z_n & \leq & 1, & z_2 & \geq & 0, \\ & & \vdots & & & \vdots \\ a_{m1}z_1 + a_{m2}z_2 + \cdots + a_{mn}z_n & \leq & 1, & z_n & \geq & 0. \end{array}$$

(ii) Compute  $w = 1/(z_1 + \cdots + z_n)$ , the reciprocal of the maximum from Step (i).

(iii) The optimal mixed strategy is the column  $(c_1, c_2, \dots, c_n)^T = w(z_1, z_2, \dots, z_n)^T$ .

**Remark** The left-hand side of each of the inequalities are just the entries of this vector:

$$A \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$

## 10.10 Symmetric Games

**Definition** A matrix  $A = (a_{ij})$  is called **anti-symmetric** if  $a_{ij} = -a_{ji}$  for all  $i$  and  $j$ ; this means that there is a near-symmetry about the main diagonal, where all entries are the same **except** corresponding entries on either side of the diagonal have opposite signs.

**Definition** A **symmetric game** is a game where the pay-off depends only on the strategies employed in the game, and **not** on who has played them.

# 11 Markov Processes

## 11.1 Probability

The **probability** of an event  $E$  is a number  $P(E)$  lying in the interval  $[0, 1]$ .

- If  $E$  and  $F$  are **disjoint**, meaning they can't both happen, then the probability of either  $E$  or  $F$  happening is  $P(E \cup F) = P(E) + P(F)$ .
- If  $E$  and  $F$  are **independent**, meaning they don't affect each other, then the probability of both  $E$  and  $F$  happening is  $P(E \cap F) = P(E)P(F)$ .

## 11.2 Introduction to Markov Processes

**Definition** A **Markov process** is a sequence of events where the outcome at each stage depends on the outcome of the previous stage, but is independent of the outcomes prior to the previous stage. These can be represented by **transition matrices**, where the columns acting as the events in the **current** state and the rows acting as the events in the **next** state.

**Definition** A matrix  $A = (a_{ij})$  is called a **stochastic matrix** if the following are true:

- (i) All entries are non-negative, meaning  $a_{ij} \geq 0$  for every  $i$  and  $j$ .
- (ii) All columns sum to one, meaning  $a_{1j} + a_{2j} + \dots + a_{mj} = 1$  for every  $j$ .

## 11.3 Iteration

**Method – Iterating a Markov Process:** If we have a transition matrix  $A$  representing a Markov process, we can compute the probability of an event occurring, given a starting state, after  $n$  stages by looking at the entries in the matrix  $A^n$ .

## 11.5 Regular Stochastic Matrices

**Definition** A stochastic matrix  $A = (a_{ij})$  is **regular** if a power of it contains **strictly** positive entries, i.e. there exists some  $k \in \mathbb{Z}^+$  where the entries of  $A^k$  are greater than zero.

## 11.6 The Stable Matrix

**Proposition** Let  $A$  be a regular stochastic matrix. Then, the sequence of powers  $A^n$  has a limit as  $n \rightarrow \infty$ , which we call the **stable matrix**  $S$ . In particular, for a starting distribution  $\mathbf{x}_0$ , the sequence  $\mathbf{x}_n = A^n \mathbf{x}_0$  has a limit as  $n \rightarrow \infty$ , which we call the **stable distribution**  $\mathbf{x}$ .

**Method – Finding the Stable Distribution and Stable Matrix:** If we have a regular stochastic matrix  $A$ , the previous proposition implies  $\lim_{n \rightarrow \infty} A^n = S$  exists. In fact, the columns of  $S$  are all identical, and all equal the stable distribution  $\mathbf{x}$ .

- (i) Solve the equation  $A\mathbf{x} = \mathbf{x}$ , which can be done by solving  $(A - I)\mathbf{x} = \mathbf{0}$ .
- (ii) Step (i) gives infinitely-many solutions; choose one where the entries of  $\mathbf{x}$  sum to 1.
- (iii) Finally,  $S$  is just the matrix where every column is  $\mathbf{x}$ .

## 11.7 Absorbing Stochastic Matrices

**Definition** Let  $A$  be a stochastic matrix. An **absorbing state** is one from which “there is no escape”, that is it’s a state whose column consists of zeros everywhere **except** for a one on the main diagonal.

**Definition** An **absorbing stochastic matrix** is a stochastic matrix  $A$  where these occur:

- (i) There is at least one absorbing state.
- (ii) It is possible to get to an absorbing state when we start at **any** state.

## 11.8 Formula for the Stable Matrix

**Method – Stable Matrix of an Absorbing Matrix:** Let  $A$  be an absorbing stochastic matrix. Then, we can always write it in the form

$$A = \left( \begin{array}{c|c} I & Q \\ \hline O & R \end{array} \right),$$

where  $I$  is the identity matrix and  $O$  is the matrix of all-zeros.

1. Compute the inverse matrix  $(I - R)^{-1}$ .
2. Compute the matrix  $Q(I - R)^{-1}$ .
3. The stable matrix of  $A$  is as follows:

$$S = \left( \begin{array}{c|c} I & Q(I - R)^{-1} \\ \hline O & O \end{array} \right).$$

**Warning:** I have used different notation compared to the lecture notes. This is because we use  $S$  to represent the stable matrix, but for some reason the notes use  $S$  (only in this one place) to mean a sub-matrix of  $A$ . Consequently, I have replaced it with  $Q$  instead.

**Note:** The matrix  $(I - R)^{-1}$  above is called the **fundamental matrix**. Its  $ij^{\text{th}}$  entry tells us the average number of times one can expect to visit the non-absorbing state  $i$  if we start at the non-absorbing state  $j$ .