

# Math 1554

Milan Capoor

Fall 2021

## 1 Module 1: Linear Equations

### 1.1 TOPIC 1: Systems of Linear Equations

*Linear equation:* an equation in the form  $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$  where  $a_n$  and  $b$  are real or complex coefficients of the variables  $x_n$

*Linear system (system of linear equations):* a collection of linear equations involving the same variables

*Solution:* a list of values  $(s_1, s_2, \dots, s_n)$  for which substitution into the variables  $(x_1, x_2, \dots, x_n)$  makes the equations true - Found where the equations' lines intersect

*Solution set:* the collection of all solutions to a linear system

*Equivalent:* characteristic of two linear systems if they have the same solution sets

*Consistent system:* a linear system which has infinitely many solutions or one solution - Happens when the lines intersect at a single point or when the lines coincide

*Inconsistent system:* system with no solutions - When the lines are parallel

*Matrix:* the essential information of a linear system can be recorded compactly in a rectangular array

Given the system

$$\begin{aligned}x_1 - 2x_2 + x_3 &= 0 \\2x_2 - 8x_3 &= 8 \\5x_1 - 5x_3 &= 10\end{aligned}$$

The *coefficient matrix*

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

encodes the essential information of the system while the *augmented matrix* is

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

An  $m \times n$  matrix is a rectangular array of numbers with  $m$  rows and  $n$  columns.

### 1.1.1 Solving a linear system

The basic strategy is to *replace one system with an equivalent system that is easier to solve*. Three basic operations are used to simplify a linear system:

1. (Replacement) Replace one equation by the sum of itself and a multiple of another equation
2. (Interchange) Swap two equations
3. (Scaling) Multiply all the terms in an equation by a nonzero constant

### 1.1.2 Example 1: Solve the system given above

$$\begin{aligned}x_1 - 2x_2 + x_3 &= 0 \\2x_2 - 8x_3 &= 8 \\5x_1 - 5x_3 &= 10\end{aligned}$$

Row Reduction Procedure:

1. Construct the augmented matrix

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

2.  $R_3 - 5R_1$

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

3.  $\frac{1}{2}R_2$

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

4.  $R_3 - 10R_2$

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 30 & -30 \end{bmatrix}$$

5.  $\frac{1}{30}R_3$

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

6.  $R_1 - R_3$  and  $R_2 + 4R_3$

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

7.  $R_1 + 2R_2$

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

8. Verify the solution  $(1, 0, -1)$  by substituting it for  $(x_1, x_2, x_3)$  in the original system

$$(1) - 2(0) + (-1) = 0$$

$$2(0) - 8(-1) = 8$$

$$5(1) - 5(-1) = 10$$

The equations agree so  $(1, 0, -1)$  is indeed a solution to the system. ■

These row operations can be applied to any matrix —not just those that arise as the augmented matrices of a linear system.

*Row equivalent:* quality of two matrices if there exists a sequence of elementary row operations that transforms one matrix into the other

**All row operations are reversible. Hence, *If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.***

### 1.1.3 Existence and Uniqueness

A large focus of the course and of the analysis of linear systems generally depends on asking two questions:

- Is the system consistent?
- If a solution exists, is it unique?

These questions can often be answered from the triangular form of the matrix (in

Example 1 this was  $\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$ ) by understanding that  $x_3 = -1$  and solving the other equations from there.

If this method creates contradictions, however, such as in the system  $\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & 0 & 1 \end{bmatrix}$ , then the original system is inconsistent.

## 1.2 TOPIC 2: Row Reduction and Echelon Forms

*Nonzero:* any row or column which contains at least one nonzero entry

*Leading entry:* the leftmost nonzero entry of a row

*Row echelon form:* a matrix is in row echelon form if it has the following three properties:

1. All nonzero rows are above any rows of all zeros
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it
3. All entries in a column below a leading entry are zeroes

*Reduced echelon form:* a matrix in row echelon form which satisfies the additional criteria:

4. The leading entry in each nonzero row is 1
5. Each leading 1 is the only nonzero entry in its column

These are the "triangular" matrices of section 1.1.

***Each matrix is row equivalent to one and only one reduced echelon matrix.***

If a matrix  $A$  is row equivalent to an echelon matrix  $U$ , we call  $U$  an (*reduced*) *echelon form (REF/RREF)* of  $A$

### 1.2.1 Pivot Positions

*Pivot position:* a location in a matrix  $A$  that corresponds to a leading 1 in the reduced echelon form of  $A$ .

*Pivot column:* a column of  $A$  that contains a pivot position

1.2.2 Example 2: Row reduce matrix  $A$  below to echelon form and locate its pivot columns.

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

Solution:

1. Interchange  $R_1$  and  $R_4$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

2.  $R_2 + R_1$  and  $R_3 + 2R_1$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

3.  $R_3 - \frac{5}{2}R_2$  and  $R_4 + \frac{3}{2}R_2$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix}$$

4. Interchange  $R_3$  and  $R_4$

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Matrix  $A$  is thus in echelon form and so columns 1, 2, and 4 are pivot columns.

*Pivot:* a nonzero number in a pivot position which is used during row reduction to create zeros

### 1.2.3 The Row Reduction Algorithm

1. Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.
2. Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

3. Use row replacement operations to create zeros in all positions below the pivot.
4. Cover the row containing the pivot position and all rows above it. Apply steps 1-3 to the remaining submatrix, iterating until there are no more nonzero rows to modify. At this time we have the row echelon form.
5. This step is used to find the reduced row echelon form. Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 via scaling operation.

#### 1.2.4 Solutions of Linear Systems

The row reduction algorithm leads directly to the solution set of a system when the algorithm is applied to the augmented matrix of the system.

*Basic variables:* the variables of a linear system corresponding to pivot columns in the reduced echelon matrix

*Free variables:* the other variables in the system

Whenever a system is consistent, the solution set can be described explicitly by solving the reduced system of equations for the basic variables in terms of the free variables.

Each different choice of a free variable determines a (different) solution of the system and every solution of the system is determined by a choice of the free variables.

1.2.5 Example 4: Find the general solution of the linear system whose augmented matrix has been reduced to

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

Moving from echelon form to reduced echelon form we have:

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 2 & -8 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

Which is equivalent to

$$\begin{aligned}x_1 + 6x_2 + 3x_4 &= 0 \\x_3 - 4x_4 &= 5 \\x_5 &= 7\end{aligned}$$

Because the pivot columns are 1, 3, and 5, the basic variables are  $x_1, x_3$ , and  $x_5$ , leaving  $x_2$  and  $x_4$  as free variables. Thus

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases}$$

#### 1.2.6 Existence and Uniqueness Questions (Redux)

When a system is in echelon form and contains no equation of the form  $0 = b(b \neq 0)$ , *every* nonzero equation contains a basic variable with a nonzero coefficient. Either the basic variables are completely determined (no free variables), signifying a unique solution, or the basic variables can be expressed in terms of one or more free variables, signifying infinitely many solutions. Thus: **A linear system is consistent iff the rightmost column of the augmented matrix is *not* a pivot column (i.e. there are no rows of the form  $[0 \ \dots \ 0 \ b]$ ).**



## 1.3 TOPIC 3: Vector Equations

*Column vector:* a matrix with only one column

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

The set of all vectors with two entries is denoted  $\mathbb{R}^2$  ("r-two"), where  $\mathbb{R}$  refers to the real number entries of the vector and the exponent 2 indicates that each vector contains two entries.

Vectors in  $\mathbb{R}^2$  are ordered pairs which are equal iff their corresponding entries are equal.

Given two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^2$ , their sum is

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$$

while the scalar multiple of a vector is

$$c\mathbf{u} = c \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} c \cdot u_1 \\ c \cdot u_2 \end{bmatrix}$$

### 1.3.1 Geometric Descriptions of $\mathbb{R}^2$

Because each point in the cartesian plane is determined by an ordered pair of numbers, we can identify a geometric point with the column vector  $\begin{bmatrix} a \\ b \end{bmatrix}$ . So we may regard  $\mathbb{R}^2$  as the set of all points in the plane.

*Parallelogram rule for vector addition:* If  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^2$  are represented as points in the plane, then  $\mathbf{u} + \mathbf{v}$  corresponds to the fourth vertex of the parallelogram whose other vertices are  $\mathbf{u}$ ,  $\mathbf{0}$ , and  $\mathbf{v}$ .

### 1.3.2 Algebraic Properties of $\mathbb{R}^n$

For all  $\mathbf{u}, \mathbf{v}, \mathbf{w}$  in  $\mathbb{R}^n$  and all scalars  $c$  and  $d$ :

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.  $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
6.  $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
7.  $c(d\mathbf{u}) = (cd)\mathbf{u}$
8.  $1\mathbf{u} = \mathbf{u}$

*Linear combination:* a composite formed of vectors in  $\mathbb{R}^n$  and corresponding scalar weights ( $\mathbf{y} = c_1\mathbf{v}_1 + \dots + c_p\mathbf{v}_p$ )

1.3.3 Example 5: Solve  $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$  for  $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$ ,  $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$ , and  $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$ .

We shall begin by constructing an augmented matrix and finding its reduced row echelon form:

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

So  $x_2 = 2$  and  $x_1 = 7 - 2x_2 = 3$ .

Checking:  $3 \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2 \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 + 4 \\ -6 + 10 \\ -15 + 12 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} \blacksquare.$

A vector  $\mathbf{b}$  can be generated by a linear combination of  $\mathbf{a}_1, \dots, \mathbf{a}_n$  iff there exists a solution to the linear system corresponding to the augmented matrix  $[\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_n \ \mathbf{b}]$ . One of the key ideas in linear algebra is to study the set of all vectors that can be written as a linear combination of a fixed set  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  of vectors.

The subset of  $\mathbb{R}^n$  spanned by  $\mathbf{v}_1, \dots, \mathbf{v}_p$  : (written  $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ ) is the collection of all vectors that can be written in the form  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p$  with  $c_1, \dots, c_p$  scalars.

Asking whether a vector is in  $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  amounts to asking whether the linear system with associated augmented matrix has a solution.

Note:  $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  contains every scalar multiple of  $\mathbf{v}_1$  including  $\mathbf{0}$

### 1.3.4 Geometric description of $\text{Span}\{\mathbf{v}\}$

For  $\mathbf{v} \in \mathbb{R}^3$ ,  $\text{Span}\{\mathbf{v}\}$  is the set of points on the line in  $\mathbb{R}^3$  through  $\mathbf{v}$  and  $\mathbf{0}$ .

For  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ ,  $\text{Span}\{\mathbf{u}, \mathbf{v}\}$  is the plane in  $\mathbb{R}^3$  that contains  $\mathbf{u}$ ,  $\mathbf{v}$ , and  $\mathbf{0}$ .

1.3.5 Example 6:  $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$  is a plane through the origin in  $\mathbb{R}^3$ . Is  $\mathbf{b}$  in that plane?

$$(\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix})$$

This question amounts to asking if the equation  $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$  has a solution. From row reduction:

$$\begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

This suggests that  $0 = -2$  but as this is not possible, the matrix is inconsistent and so the vector equation has no solution. Thus,  $\mathbf{b}$  is not in  $\text{Span}\{\mathbf{a}_1, \mathbf{a}_2\}$ . ■

## 1.4 TOPIC 4: The Matrix Equation

- $\in$ : belongs to
- $\mathbb{R}^n$ : the set of vectors with  $n$  real-valued elements
- $\mathbb{R}^{m \times n}$ : the set of real-valued matrices with  $m$  as rows and  $n$  as columns

Example: The notation  $\vec{x} \in \mathbb{R}^5$  means that  $\vec{x}$  is a vector with five real-valued elements.

### 1.4.1 Matrix-Vector Product as a Linear Combination

If  $A \in \mathbb{R}^{m \times n}$  has columns  $\vec{a}_1, \dots, \vec{a}_n$ , then the matrix vector product  $A\vec{x}$  is a linear combination of the columns of  $A$ .

$$A\vec{x} = \sum_{n=1}^n x_n \vec{a}_n$$

Note that  $A\vec{x}$  is in the span of the columns of  $A$ .

This means that the solution sets for

$$A\vec{x} = \vec{b}$$

is the same as

$$x_1 \vec{a}_1 + \dots + x_n \vec{a}_n = \vec{b}$$

which is again equivalent to the set of linear equations with the augmented matrix

$$\begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n & \vec{b} \end{bmatrix}$$

### 1.4.2 Example:

Suppose that  $A = \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$  and  $\vec{x} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$

1. The following product can be written as a linear combination of vectors:

$$A\vec{x} = 2\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 0 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ -9 \end{pmatrix}$$

2. Is  $\vec{b} = \begin{pmatrix} 2 \\ 9 \end{pmatrix}$  in the span of the columns of A?

If  $\vec{b} \in \text{Span}\{A\}$ , then  $\vec{b} = c_1\begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2\begin{pmatrix} 0 \\ -3 \end{pmatrix}$ . This is true for  $\vec{c} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}$  so  $\vec{b} \in \text{Span}\{A_{col}\}$

**The equation  $A\vec{x} = \vec{b}$  has a solution iff  $\vec{b}$  is a linear combination of the columns of  $A$**

1.4.3 Example: For what vectors  $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$  does the equation have a solution?

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 8 & 4 \\ 0 & 1 & -2 \end{bmatrix} \vec{x} = \vec{b}$$

Solution:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ 2 & 8 & 4 & b_2 \\ 0 & 1 & -2 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 2 & -4 & b_2 - 2b_1 \\ 0 & 1 & -2 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 2 & -4 & b_2 - 2b_1 \\ 0 & 0 & 0 & b_3 - \frac{1}{2}b_2 + b_1 \end{bmatrix}$$

So

$$\vec{b} = \begin{bmatrix} -\frac{1}{2}b_2 + b_3 \\ b_2 \\ b_3 \end{bmatrix}$$

Essential concept: If  $A$  is an  $m \times n$  matrix, the following statements are logically equivalent – for a particular  $A$ , *all* are true or *all* are false:

- For each  $\mathbf{b}$  in  $\mathbb{R}^m$ , the equation  $A\mathbf{x} = \mathbf{b}$  has a solution.
- Each  $\mathbf{b}$  in  $\mathbb{R}^m$  is a linear combination of the columns of  $A$
- The columns of  $A$  span  $\mathbb{R}^m$
- $A$  has a pivot position in every row

#### 1.4.4 Summary: Ways of representing Linear Systems

1. A list of equations
2. An augmented matrix
3. A vector equation
4. A matrix equation

#### 1.4.5 Matrix-vector products

If  $A$  is an  $m \times n$  matrix,  $\mathbf{u}$  and  $\mathbf{b}$  are vectors in  $\mathbb{R}^n$ , and  $c$  is a scalar, then:

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$
- $A(c\mathbf{u}) = c(A\mathbf{u})$

Moreover, the product  $A\mathbf{x}$  can be easily calculated by taking advantage of the nature of a Matrix-vector product:

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1x_1 + a_2x_2 + a_3x_3 \\ a_4x_1 + a_5x_2 + a_6x_3 \\ a_7x_1 + a_8x_2 + a_9x_3 \end{bmatrix}$$

*The identity matrix:* denoted  $\mathbf{I}$ , this  $n \times n$  matrix contains 1's on the diagonal and 0's elsewhere, creating the universal property that  $\mathbf{I}_n\mathbf{x} = \mathbf{x}$  for every  $\mathbf{x} \in \mathbb{R}^n$

## 1.5 TOPIC 5: Solution sets of Linear Systems

*Homogenous:* characteristic of linear systems of the form  $A\vec{x} = \vec{b}$ ,  $\vec{b} = \vec{0}$

*Inhomogeneous:* systems of the form  $A\vec{x} = \vec{b}$ ,  $\vec{b} \neq \vec{0}$

Because homogenous systems always have trivial solutions, the interesting question comes in asking whether they have non-trivial solutions.

$A\vec{x} = 0$  has a nontrivial solution  $\iff$  there is a free variable  $\iff A$  has a column with no pivot

### 1.5.1 Example: Identify the free variables and the solution set for

$$\begin{aligned}x_1 + 3x_2 + x_3 &= 0 \\2x_1 - x_2 - 5x_3 &= 0 \\x_1 - 2x_3 &= 0\end{aligned}$$

$$\begin{bmatrix} 1 & 3 & 1 & 0 \\ 2 & -1 & -5 & 0 \\ 1 & 0 & -2 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & -7 & -7 & 0 \\ 0 & -3 & -3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Row 3 has no pivot ( $x_3$  is free) so there is a non-trivial solution.

$$\begin{cases} x_1 = 2x_3 \\ x_2 = -x_3 \\ x_3 \text{ is free} \end{cases} \implies x_3 \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

### 1.5.2 Parametric vector forms

*Parametric vector form:* a more convenient way of expressing the solutions of a linear system, taking advantage of the geometric interpretation of a linear system. In general, for free variables  $x_k, \dots, x_n$  of  $A\vec{x} = 0$ , the solutions can all be written as

$$\vec{x} = \sum_{n=k}^n x_n \vec{v}_n$$

In other words, solving an equation amounts to finding an explicit description of the solution plane as a set spanned by  $\mathbf{u}$  and  $\mathbf{v}$ . Thus, the equation from earlier describing the solution set can also be written as

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \quad (s, t \in \mathbb{R})$$

emphasising the role of the free variables as arbitrary scalar multiples of the vectors forming a plane.

### 1.5.3 Example: Describe all solutions of $A\mathbf{x} = \mathbf{b}$

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus  $x_1 = -1 + \frac{4}{3}x_3$ ,  $x_2 = 2$ , and  $x_3$  is free. The general form vector is

$$\mathbf{x} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} \implies \mathbf{x} = \mathbf{p} + x_3\mathbf{v}$$

By replacing  $x_3$  with a general parameter  $t \in \mathbb{R}$ , we have a universal solution for nonhomogenous systems:

**$\mathbf{x} = \mathbf{p} + t\mathbf{v}$  is a solution for  $A\mathbf{x} = \mathbf{b}$  which is parallel to the line of the solution set  $A\mathbf{x} = 0$  because it is a translation of  $\mathbf{v}$  by the particular solution  $\mathbf{p}$**



#### 1.5.4 Theorem:

Suppose the equation  $Ax = b$  is consistent for some given  $\mathbf{b}$  and let  $\mathbf{p}$  be a solution. The solution set of  $Ax = b$  is the set of all vectors of the form  $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$  where  $\mathbf{v}_h$  is any solution of the homogenous equation  $Ax = 0$

## 1.6 TOPIC 6: Linear Independence

A set of vectors  $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$  are linearly independent if  $\sum_{n=1}^k c_n \vec{v}_n = \vec{0}$  has only the trivial solution ( $\vec{c} = \vec{0}$ ). It is linearly dependent otherwise.

Establishing linear independence is thus equivalent to asking whether the equation  $V\vec{c} = \vec{0}$  ( $V \in \mathbb{R}^k$  := the matrix corresponding to the linear combinations of the vectors) is only true for  $\vec{c} = \vec{0}$

Example: For what values of  $h$  is the set of vectors linearly independent?

$$\begin{bmatrix} 1 \\ 1 \\ h \end{bmatrix}, \begin{bmatrix} 1 \\ h \\ 1 \end{bmatrix}, \begin{bmatrix} h \\ 1 \\ 1 \end{bmatrix}$$

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

If  $2-h-h^2 = 0$  then we have a free variable and the vectors will be linearly dependent (because there will be more solutions than the trivial case).

Factoring, we get

$$0 = -(h+2)(h-1)$$

so for the vectors to be independent,

$$h \neq \{-2, 1\}$$

### 1.6.1 Linear Independence Theorems

1. More Vectors Than Elements For vectors  $\vec{v}_1, \dots, \vec{v}_k \in \mathbb{R}^n$ , if  $k > n$ , then  $\{\vec{v}_1, \dots, \vec{v}_k\}$  is linearly dependent (because not every column of the matrix  $A = (\vec{v}_1, \dots, \vec{v}_k)$  would be pivotal).
2. Set Contains Zero Vector If any one or more of  $\vec{v}_1, \dots, \vec{v}_k$  is  $\vec{0}$ , then  $\{\vec{v}_1, \dots, \vec{v}_k\}$  is linearly dependent (again because there would be non-pivotal columns of the corresponding matrix).
3. A set containing only one vector is linearly independent iff  $\mathbf{v}$  is not the zero vector
4. Sets of two vectors can be determined to be linearly dependent by inspection if one is a multiple of the other.

5. In geometric terms, two vectors are linearly dependent iff they lie on the same line through the origin.

## 1.7 TOPIC 7: Introduction to Linear Transformations

Instead of as a linear combination of variables, matrices can also be viewed as a *function* from one set of vectors to another. Thus, solving the equation  $Ax = b$  amounts to finding all vectors  $\mathbf{x}$  in  $\mathbb{R}^4$  that are transformed into the vector  $\mathbf{b}$  in  $\mathbb{R}^2$  under the "action" of multiplication by  $A$ .

*Transformation (function, mapping):* a rule that assigns to each vector  $\mathbf{x} \in \mathbb{R}^n$  a vector  $T(\mathbf{x}) \in \mathbb{R}^m$ .

*Domain:* The set  $\mathbb{R}^n$  is the domain of  $T$  ( $T \in \mathbb{R}^n$ )

*Codomain:* The set  $\mathbb{R}^m$  is the codomain of  $T$  ( $T \in \mathbb{R}^n$ )

This domain-codomain relationship is expressed in the notation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$

*Image:* this is the "resultant" vector of the transformation

$$T(\mathbf{x}) \in \mathbb{R}^m \text{ for } \mathbf{x} \in \mathbb{R}^n \implies$$

"The vector  $T(\mathbf{x})$  is the image of  $\mathbf{x}$  under the action of  $T$ "

*Range:* The set of all images of  $T(\mathbf{x})$  is the range of  $T$ . This is equivalent to the Span of the columns of a matrix.

For example,  $f(x) = \sin x : \mathbb{R} \mapsto \mathbb{R}$ , so the domain is  $\mathbb{R}$ , the codomain is  $\mathbb{R}$ , and the range is  $[-1, 1]$

A function  $T : \mathbb{R}^n \mapsto \mathbb{R}^m$  is linear if

- $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) \quad \forall \vec{u}, \vec{v} \in \mathbb{R}^n$
- $T(c\vec{v}) = cT(\vec{v}) \quad \forall \vec{v} \in \mathbb{R}^n, c \in \mathbb{R}$

*Principle of superposition:*

$$T(c_1\vec{v}_1 + \dots + c_k\vec{v}_k) = c_1T(\vec{v}_1) + \dots + c_kT(\vec{v}_k)$$

**Every matrix transformation  $T_A$  is linear.**

### 1.7.1 Matrix Transformations

A matrix transformation (a mapping where  $\forall \mathbf{x} \in \mathbb{R}^n, T(\mathbf{x}) = A\mathbf{x}$  ( $A \in \mathbb{R}^{m \times n}$ )) can be written as  $\mathbf{x} \mapsto A\mathbf{x}$

Similarly, every matrix transform is completely determined by what it does to the columns of the  $n \times n$  identity matrix.

Geometric interpretations of transforms in  $\mathbb{R}^2$ :

1.  $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  is a reflection through  $x_1 = x_2$
2.  $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$  is a projection onto the x-axis ( $\begin{bmatrix} x_1 \\ 0 \end{bmatrix}$ )
3.  $A = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$  is a scaling by k

Geometric interpretations of transforms in  $\mathbb{R}^3$ :

1.  $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$  is a projection onto the  $x_1, x_2$ -plane ( $T(\vec{x}) = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$ )
2.  $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$  is a reflection through the  $x_1, x_3$ -plane ( $T(\vec{x}) = \begin{bmatrix} a \\ -b \\ c \end{bmatrix}$ )

*Standard vectors in  $\mathbb{R}^n$ :*

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Multiplying a matrix by  $\vec{e}_i$  gives column  $i$  of  $A$

### 1.7.2 Standard transform matrices

Counterclockwise rotation by angle  $\theta$  about  $(0, 0)$ :

$$T(\mathbf{x}) = A\vec{x} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Reflection through  $x_1$ -axis:

$$T(x) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Reflection through  $x_2$ -axis:

$$T(x) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Reflection through  $x_2 = x_1$ :

$$T(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Reflection through  $x_2 = -x_1$

$$T(x) = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Horizontal contraction:

$$T(x) = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \quad |k| < 1$$

Horizontal expansion:

$$T(x) = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \quad |k| > 1$$

Vertical contraction:

$$T(x) = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}, \quad |k| < 1$$

Vertical expansion:

$$T(x) = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}, \quad |k| > 1$$

Horizontal shear (left):

$$T(x) = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad k < 0$$

Horizontal shear (right):

$$T(x) = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \quad k > 0$$

Vertical shear (down):

$$T(x) = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \quad k < 0$$

Vertical shear (up):

$$T(x) = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \quad k > 0$$

Projection onto the  $x_1$ -axis:

$$T(x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Projection onto the  $x_2$ -axis:

$$T(x) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

*Onto transform:* a transform for which  $A\vec{x} = \vec{b}$

- This is an existence property, for any  $\vec{b} \in \mathbb{R}^m$ ,  $A\vec{x} = \vec{b}$  has a solution
- T is onto iff its standard matrix has a pivot in every row

*One-to-one transform:* a transform for which there is at most one solution to  $A\vec{x} = \vec{b}$

- This uniqueness property does not assert existence for all  $\vec{b}$
- T is one-to-one iff the only solution to  $T(\vec{x}) = \vec{0}$  is  $\vec{x} = \vec{0}$
- T is one-to-one iff every column of A is pivotal

Onto means that the columns of A span  $\mathbb{R}^n$

One-to-one means that the only solution is the trivial one and A has linearly independent columns

### 1.7.3 Determining the image of a transform

Example: Suppose T is a transform  $T : \mathbb{R}^2 \mapsto \mathbb{R}^3$  such that

$$T(\vec{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} \quad T(\vec{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$$

As any matrix can be written as a linear combination

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1 \vec{e}_1 + x_2 \vec{e}_2$$

And since  $T$  is a linear transformation

$$\begin{aligned} T(\vec{x}) &= x_1 T(\vec{e}_1) + x_2 T(\vec{e}_2) \\ &= x_1 \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5x_1 - 3x_2 \\ -7x_1 + 8x_2 \\ 2x_1 + 0 \end{bmatrix} \end{aligned}$$

#### 1.7.4 Example: Constructing a standard matrix

Define a linear transformation by

$$T(x_1, x_2) = (3x_1 + x_2, 5x_1 + 7x_2, x_1 + 3x_2)$$

$$A = [T(\vec{e}_1) \quad T(\vec{e}_2)]$$

$$T(e_1) = T(1, 0) = \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix}$$

$$T(e_2) = T(0, 1) = \begin{bmatrix} 1 \\ 7 \\ 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix}$$

Because there are more columns than rows, it cannot be onto. Because the two columns are not multiples of each other by inspection, the columns are linearly independent so the matrix is one-to-one.



## 2 Module 2: Matrix Algebra

### 2.1 TOPIC 1: Matrix Operations

*Main diagonal:* the diagonal entries  $a_{11}, a_{22}, a_{33}, \dots$  of  $A = [a_{ij}]$ ,  $A \in \mathbb{R}^{m \times n}$

*Diagonal matrix:* a square matrix  $n \times n$  whose nondiagonal entries are zero (e.g.  $I_n$ )

*Zero matrix:* a matrix whose entries are all zero, written 0

*Equal:* property of two matrices of the same size with corresponding entries equal

Properties of Matrix addition:

1.  $A + B = B + A$
2.  $(A + B) + C = A + (B + C)$
3.  $A + 0 = A$
4.  $r(A + B) = rA + rB$
5.  $(r + s)A = rA + sB$
6.  $r(sA) = (rs)A$

If a matrix  $B$  multiplies a vector  $\mathbf{x}$ , it is a transform  $B\mathbf{x}$ . If this vector is then multiplied by  $A$ , the composite transformed vector is  $A(Bx)$ . Thus:

$$A(Bx) = (AB)x$$

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \dots & A\mathbf{b}_p \end{bmatrix}$$

Multiplication of matrices corresponds to composition of linear transformations.

Row column rule for matrix multiplication: If  $A \in \mathbb{R}^{m \times n}$  has rows  $\vec{a}_i$ , and  $B \in \mathbb{R}^{n \times p}$  has columns  $\vec{b}_j$ , each element of the product  $C = AB$  is the dot product  $c_{ij} = \vec{a}_i \cdot \vec{b}_j$

Each column of  $AB$  is a linear combination of the columns of  $A$  using weights from the corresponding column of  $B$ .

**The number of columns of  $A$  must match the number of rows in  $B$  in order for a linear combination such as  $A\mathbf{b}_1$  to be defined.**

Properties of Matrix multiplication:

1.  $A(BC) = (AB)C$
2.  $A(B + C) = AB + AC$
3.  $(B + C)A = BA + CA$
4.  $r(AB) = (rA)B = A(rB)$
5.  $I_m A = A = A I_n$
6. If  $AB = AC$  it is usually NOT true that  $B = C$
7. If  $AB$  is the zero matrix, you CANNOT conclude that either  $A =$  or  $B = 0$

If  $A \neq 0$  and  $\mathbf{x} \in \mathbb{R}^n$ , then  $A^k \mathbf{x}$  is the result of left-multiplying  $\mathbf{x}$  by  $A$  repeatedly  $k$  times. ( $A^0 = I_n$ )

*Transpose:* the transpose of an  $m \times n$  matrix  $A$  is the  $n \times m$  matrix  $A^T$  whose columns are the corresponding rows of  $A$ .

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Transposition properties:

1.  $(A^T)^T = A$
2.  $(A + B)^T = A^T + B^T$
3.  $(rA)^T = rA^T$
4.  $(AB)^T = B^T A^T$

Hence, *the transpose of a product of matrices equals the product of their transposes in the reverse order.*

## 2.2 TOPIC 2: The Inverse of a Matrix

*Invertible:* a square matrix is invertible if there is a matrix  $C \in \mathbb{R}^{n \times n}$  such that

$$CA = I \quad \text{and} \quad AC = I$$

Here,  $C$  is the unique inverse of  $A$  often denoted  $A^{-1}$

*Singular matrix:* a noninvertible matrix

*Nonsingular matrix:* an invertible matrix

### 2.2.1 Finding the inverse

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

If  $ad - bc \neq 0$ , then  $A$  is invertible and

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

*Determinant:*  $\det A = ad - bc$ . If  $\det A = 0$ ,  $A$  is not invertible.

**If  $A$  is an invertible  $n \times n$  matrix, then for each  $\vec{b}$  in  $\mathbb{R}^n$ , the equation  $Ax = \vec{b}$  has the unique solution  $\vec{x} = A^{-1}\vec{b}$**

While the equation above can be used to solve an equation  $Ax = b$ , row reduction of  $[A \quad \mathbf{b}]$  is almost always faster.

Properties of invertible matrices:

1.  $(A^{-1})^{-1} = A$
2.  $(AB)^{-1} = B^{-1}A^{-1}$
3.  $(A^T)^{-1} = (A^{-1})^T$

*Elementary matrix:* a matrix that is obtained by performing a single elementary row operation on an identity matrix.

If an elementary row operation is performed on an  $m \times n$  matrix  $A$ , the resulting matrix can be written as  $EA$ , where the  $m \times m$  matrix  $E$  is created by performing the same row operation on  $I_m$ .

**Every elementary matrix is invertible and square** where the inverse of E is the elementary matrix that transforms E back into I.

*An  $n \times n$  matrix is invertible iff A is row equivalent to  $I_n$ , in which case any sequence of elementary row operations which maps  $A \mapsto I_n$  also transforms  $I_n \mapsto A^{-1}$*

A has an inverse iff for all  $\vec{b} \in \mathbb{R}^n$ ,  $Ax = b$  has a unique solution

2.2.2 To find  $A^{-1}$ :

Row reduce the augmented matrix  $[A \ I]$  If A is row equivalent to I, then  $[A \ I]$  is row equivalent to  $[I \ A^{-1}]$

Example: Find the inverse of  $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$

Solution:

$$[A \ I] = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}$$

Since  $A \sim I$ , A is invertible and

$$A^{-1} = \begin{bmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}$$

Row reduction of  $[A \ I]$  to  $[I \ A^{-1}]$  can be viewed as the simultaneous solution of the n systems

$$Ax = \mathbf{e}_1, \quad Ax = \mathbf{e}_2, \quad \dots, \quad Ax = \mathbf{e}_n$$

## 2.3 TOPIC 3: Characterisations of Invertible Matrices

*Invertible matrix theorem:* for a given square matrix  $A$ , the following are all true or all false:

1.  $A$  is an invertible matrix
2.  $A$  is row equivalent to  $I_n$
3.  $A$  has  $n$  pivot positions
4. The equation  $Ax = 0$  has only the trivial solution
5. The columns of  $A$  form a linearly independent set
6. The linear transformation  $x \mapsto Ax$  is one-to-one
7. The equation  $Ax = b$  has at least one solution for each  $b \in \mathbb{R}^n$
8. The columns of  $A$  span  $\mathbb{R}^n$
9. The linear transformation  $x \mapsto Ax$  maps  $\mathbb{R}^n$  onto  $\mathbb{R}^n$
10. There is an  $n \times n$  matrix  $C$  such that  $CA = I$
11. There is an  $n \times n$  matrix  $D$  such that  $AD = I$
12.  $A^T$  is an invertible matrix

If  $A$  and  $B$  are two square matrices and  $AB = I$ , then both  $A$  and  $B$  are invertible with  $B = A^{-1}$  and  $A = B^{-1}$

*Invertible linear transformation:* a transform is invertible if there exists a function  $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that

$$S(T(x)) = x \quad \forall x \in \mathbb{R}^n \quad (1)$$

$$T(S(x)) = x \quad \forall x \in \mathbb{R}^n \quad (2)$$

$$(3)$$

If such  $S$  exists, it is unique and must be a linear transform. This transform is invertible iff its standard matrix  $A$  is nonsingular. In that case,  $S(x) = A^{-1}x$  is the unique function satisfying (1) and (2).

*Ill-conditioned matrix:* an invertible matrix that can become singular if some of its entries are changed just slightly. This becomes dangerous if roundoff error makes a nonsingular matrix appear invertible.

## 2.4 TOPIC 4: Partitioned Matrices

*Partitioned matrix:* a matrix whose entries are meant to be considered differently than a list of column vectors based on the system it represents. For example, the matrix

$$A = \left[ \begin{array}{ccc|cc|c} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -2 & 1 \\ -8 & -6 & 3 & 1 & 7 & -4 \end{array} \right]$$

can also be the  $2 \times 3$  partitioned matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

whose entries are the blocks (submatrices)

$$\begin{aligned} A_{11} &= \begin{bmatrix} 3 & 0 & -1 \\ -5 & 2 & 4 \end{bmatrix}, & A_{12} &= \begin{bmatrix} 5 & 9 \\ 0 & -3 \end{bmatrix}, & A_{13} &= \begin{bmatrix} -2 \\ 1 \end{bmatrix} \\ A_{21} &= \begin{bmatrix} -8 & -6 & 3 \end{bmatrix}, & A_{22} &= \begin{bmatrix} 1 & 7 \end{bmatrix}, & A_{23} &= \begin{bmatrix} -4 \end{bmatrix} \end{aligned}$$

Assuming the blocks of two partitioned matrices are the same size, addition between them is defined as the block-by-block sum of corresponding partitions. Similarly, multiplication by a scalar is performed piecewise.

Partitioned matrices can be multiplied by the usual row-column rule as if the block entires were scalars, provided that for a product  $AB$ , the column partition of  $A$  matches the row partition of  $B$ .

*Conformable:* property of two matrices that are partitioned the same way

*Block diagonal matrix:* a partitioned matrix with zero blocks off the main diagonals - such a matrix is invertible iff each block on the diagonal is invertible

Partitioned matrices are especially useful in numerical solutions of especially large systems where partitioning can make more efficient use of system resources

## 2.5 TOPIC 5: MATRIX FACTORISATIONS

*Factorization:* an equation that expresses a matrix  $A$  as a product of two or more matrices.

Whereas multiplication was a synthesis of data, factorisation is an analysis.

### 2.5.1 LU Factorisation

This is a more efficient process for solving the system

$$Ax = b_1, \quad Ax = b_2, \quad \dots, \quad Ax = b_p$$

The method:

1. Assume  $A$  is an  $m \times n$  matrix that can be row reduced to echelon form without row interchanges
2.  $A$  can be written in the form  $A = LU$  where  $L$  is an  $m \times n$  lower triangular matrix with 1's on the diagonal and  $U$  is an  $m \times n$  echelon form of  $A$

$$A = LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix} \begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

3. When  $A = LU$ ,  $Ax = b$  can be written as  $L(Ux) = b$
4. We can then find  $x$  by solving the pair of equations

$$Ly = b$$

$$Ux = y$$

5. First solve for  $y$ , then for  $x$ . Because  $L$  and  $U$  are triangular, both equations are easy to solve.

Example: Given

$$A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & - & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 1 \\ -3 & 8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & -1 \end{bmatrix} = LU$$

use this LU factorisation to solve  $Ax = b$ , where  $b = \begin{bmatrix} -9 \\ 5 \\ 7 \\ 11 \end{bmatrix}$

Solution:

$$[L \ b] = \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 1 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ 0 & 1 & 0 & 0 & -4 \\ 0 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} = [I \ y]$$

Then,

$$[U \ y] = \begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & -1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

In this method, finding  $x$  requires 28 arithmetic operations. Meanwhile, row reduction of  $[A \ b]$  to  $[I \ x]$  takes 62 operations.

### 2.5.2 An LU factorisation algorithm

Suppose  $A$  can be reduced to an echelon form  $U$  using only row replacements that add a multiple of one row to another row *below it*. In this case, there exist unit lower triangular elementary matrices such that

$$E_p \dots E_1 A = U$$

Then

$$A = (E_p \dots E_1)^{-1} U = LU$$

where

$$L = (E_p \dots E_1)^{-1}$$

It can be shown that the row operations which reduce  $A$  to  $U$  also reduce  $L$  to  $I$ , because  $E_p \dots E_1 L = (E_p \dots E_1)(E_p \dots E_1)^{-1} = I$

In summary:

1. Reduce  $A$  to an echelon form  $U$  by a sequence of row replacement operations, if possible.
2. Place entries in  $L$  such that the same sequence of row operations reduces  $L$  to  $I$ .



Example: Find an LU factorization of

$$A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix}$$

Solution: Since A has four rows, L should be  $4 \times 4$ . The first column of L is the first column of A divided by the top pivot entry

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

Next row reduce A to its echelon form U:

$$\begin{aligned} A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix} &\sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & -9 & -3 & -4 & 10 \\ 0 & 12 & 4 & 12 & -5 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \\ &\sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 4 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ 0 & 3 & 1 & 2 & -3 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix} = U \end{aligned}$$

The columns eliminated in each step of the row reduction process divided by the pivot form L:

$$\begin{bmatrix} 2 & & & & \\ -4 & 3 & & & \\ 2 & -9 & 2 & & \\ -6 & 12 & 4 & 5 & \end{bmatrix} \implies L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ -3 & 4 & 2 & 1 \end{bmatrix}$$

## 2.6 TOPIC 6: The Leontief Input-Output Model

The interactions between sectors can be represented by the matrix equation

$$(I - C)x = d$$

Where:  $I$  is the identity matrix  $C$  is a coefficient matrix representing the demand of each sector for goods from each other sector  $\mathbf{x}$  is a unit consumption vector describing the number of units of output which are routed to each sector based on the consumption matrix  $\mathbf{d}$  is the final demand vector (the output of the system after accounting for intermediate demand,  $Cx$ )

If  $C$  and  $\mathbf{d}$  have nonnegative entries and if each column sum of  $C$  is less than 1,

$$\mathbf{x} = (I - C)^{-1}\mathbf{d}$$

and

$$(I - C)^{-1} \cong I + C + C^2 + \dots C^m$$

which, in most real-world circumstances, will have powers of the consumption matrix approach 0 quite quickly.

## 2.7 TOPIC 7: Computer Graphics

Often computer graphics will be represented by a list of points which can be connected by straight lines.

### 2.7.1 2D Graphics

A particular graphic, for example the letter "N" can be stored as coordinates in a data Matrix

$$D = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{bmatrix}$$

This representation makes the transformation or animation of pictures very easy: the resultant figure is described by the matrix product  $AD$  where  $A$  is any linear transformation.

Unfortunately, translation is not a linear transform so cannot be represented by matrix multiplication.

*Homogeneous coordinates:* corresponding sets of coordinates where  $\forall x, y \in \mathbb{R}^2, (x, y) \mapsto (x, y, 1) \in \mathbb{R}^3$  For example,  $(0, 0)$  has homogenous coordinates  $(0, 0, 1)$ . These coordinates are not multiplied or added by scalars but can be transformed via multiplication by  $3 \times 3$  matrices.

With homogeneous coordinates we are able to represent translations:

$$(x, y) \mapsto (x + h, y + k) \implies (x, y, 1) \mapsto (x + h, y + k, 1)$$

Thus,

$$\begin{bmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + h \\ y + k \\ 1 \end{bmatrix}$$

More broadly, any linear transformation in  $\mathbb{R}^2$  can be represented in homogeneous coordinates by a partitioned matrix of the form  $\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}$  where  $A$  is a  $2 \times 2$  standard matrix.

### 2.7.2 3D Graphics

Homogeneous coordinates for the point  $(x, y, z)$  can usually be expressed as  $(X, Y, Z, H)$  where

$$x = \frac{X}{H}, \quad y = \frac{Y}{H}, \quad z = \frac{Z}{H}$$

## 2.8 TOPIC 8: Subspaces of $\mathbb{R}^n$

*Subspace:* any set  $H$  in  $\mathbb{R}^n$  that has the properties:

- The zero vector is in  $H$
- $\forall \mathbf{u}, \mathbf{v} \in H \quad \mathbf{u} + \mathbf{v} \in H$
- $\forall \mathbf{u} \in H, \quad c\mathbf{u} \in H$

A subspace is closed under addition and scalar multiplication.

One of the most common ways of visualising a subspace is a plane through the origin.

Example: If  $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$  and  $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ , then  $H$  is a subspace of  $\mathbb{R}^n$

*Column space:* the subspace of a matrix  $A$  is the set  $\text{Col } A$  of all linear combinations of the columns of  $A$

$\text{Col } A$  equals  $\mathbb{R}^m$  only when the columns of  $A$  span  $\mathbb{R}^m$ . Otherwise,  $\text{Col } A$  is only part of  $\mathbb{R}^m$

A vector  $\mathbf{b}$  is in  $\text{Col } A$  iff the equation  $A\mathbf{x} = \mathbf{b}$  has a solution.

*Null space:* the null space of a matrix  $A$  is the set  $\text{Nul } A$  of all solutions of the homogeneous equation  $A\mathbf{x} = 0$

The null space of an  $m \times n$  matrix  $A$  is a subspace of  $\mathbb{R}^n$ . Equivalently, the set of all solutions of a system  $Ax = 0$  of  $m$  homogeneous linear equations in  $n$  unknowns is a subspace of  $\mathbb{R}^n$ .

*Basis:* for a subspace  $H$  of  $\mathbb{R}^n$ , the basis is a linearly independent set in  $H$  that spans  $H$

*Standard basis:* the set  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$

The pivot columns of a matrix  $A$  form a basis for the column space of  $A$ . Note: the columns of an echelon form  $B$  are often not in the column space of  $A$

## 2.9 TOPIC 9: Dimension and Rank

The main reason for selecting a basis for a subspace  $H$  rather than just a spanning set is that each vector in  $H$  can be written in only one way as a linear combination of the basis vectors.

*Coordinate vector:* for each  $x$  in a subspace  $H$ ,

$$[x]_{\mathfrak{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

for  $c_1, \dots, c_p$  such that

$$x = c_1 \mathbf{b}_1 + \dots + c_p \mathbf{b}_p$$

*Isomorphism:* a correspondence  $x \mapsto [x]_{\mathfrak{B}}$  which is a one-to-one mapping which preserves linear combinations.

*Dimension:* the dimension of a nonzero subspace  $H$ , denoted  $\dim H$ , is the number of vectors in any basis for  $H$

The space  $\mathbb{R}^n$  has dimension  $n$  so every basis for  $\mathbb{R}^n$  consists of  $n$  vectors. Similarly, a plane through  $\mathbf{0}$  is two-dimensional while a line through  $\mathbf{0}$  is one-dimensional.

*Rank:* the rank of a matrix  $A$ , denoted  $\text{rank } A$ , is the dimension of the column space of  $A$

Since the pivot columns of  $A$  form a basis for  $\text{Col } A$ , the rank of  $A$  is just the number of pivot columns in  $A$ .

*The Rank Theorem:* If a matrix  $A$  has  $n$  columns, then  $\text{rank } A + \dim \text{Nul } A = n$

*The Basis Theorem:* Any linearly independent set of exactly  $p$  elements in a  $p$ -dimensional subspace of  $\mathbb{R}^n$  is automatically a basis for  $H$ .

### 2.9.1 Extension of the Invertible Matrix Theorem

For an  $n \times n$  matrix  $A$ , the following statements are equivalent to the statement that  $A$  is nonsingular:

1. The columns of  $A$  form a basis of  $\mathbb{R}^n$
2.  $\text{Col } A = \mathbb{R}^n$
3.  $\dim \text{Col } A = n$
4.  $\text{rank } A = n$
5.  $\text{Nul } A = \{\mathbf{0}\}$
6.  $\dim \text{Nul } A = 0$

## 3 Module 3: Determinants, Eigenvalues, and Markov Chains

### 3.1 TOPIC 1: INTRODUCTION TO DETERMINANTS

A  $2 \times 2$  matrix is invertible iff its determinant is nonzero.

For a  $1 \times 1$ :  $\det A = a_{11}$

For a  $2 \times 2$ :  $\det A = a_{11}a_{22} - a_{12}a_{21}$

For a  $3 \times 3$ :

$$\Delta = a_{11} \cdot \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \cdot \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \cdot \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

This is equivalent to

$$\Delta = a_{11} \cdot \det A_{11} - a_{12} \cdot \det A_{12} + a_{13} \cdot \det A_{13}$$

Where  $A_{11}, A_{12}, A_{13}$  are obtained from  $A$  by deleting the first row and one of the three columns. (And where  $A_{ij}$  is the matrix formed by deleting the  $i$ -th row and  $j$ -th column of  $A$ ).

**In general, an  $n \times n$  determinant is defined by the determinants of  $(n - 1) \times (n - 1)$  submatrices.**

$$\det A = \sum_{j=1}^n (-1)^{1+j} a_{1j} \det A_{1j}$$

#### 3.1.1 Example:

Compute the determinant of

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

Solution:

$$\begin{aligned} \det A &= 1 \begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - 5 \begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 0 \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} \\ &= 1(0 - 2) - 5(0) + 0(-4 - 4) \\ &= -2 \end{aligned}$$

$(i, j)$ -cofactor of  $A$ : the number given by  $C_{ij} = (-1)^{i+j} \det A_{ij}$

Cofactor expansion across the  $i$ -th row of  $A$ : a method of computing the determinant with use of cofactors

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \dots + a_{in}C_{in}$$

The plus or minus in the  $(i, j)$ -cofactor depends on the position of  $a_{ij}$  in the matrix:

$$\begin{bmatrix} + & - & + & \dots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

### 3.1.2 Example:

Use a cofactor expansion across the third row to compute  $\det A$  for

$$A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

Solution:

$$\begin{aligned} \det A &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} \\ &= a_{31}(-1)^{3+1} \det A_{31} + a_{32}(-1)^{3+2} \det A_{32} + a_{33}(-1)^{3+3} \det A_{33} \\ &= 0 + (-2)(-1) \begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} + 0 \\ &= 2(-1) = -2 \end{aligned}$$

The cofactor approach is particularly useful in matrices with many zeros because those cofactors need not be calculated.

**If  $A$  is a triangular matrix, then  $\det A$  is the product of the entries on the main diagonal of  $A$**

Numerical note: in general, a cofactor expansion requires more than  $n!$  multiplications

## 3.2 TOPIC 2: Properties of Determinants

For a square matrix  $A$ :

1. Row replacement operations on  $A$  to produce a matrix  $B$  make  $\det B = \det A$
2. Row interchange operations on  $A$  to produce  $B$  make  $\det B = -\det A$
3. Row scaling operations make  $\det B = k \det A$

Using these operations, we can reduce a matrix to echelon form and use the fact that the determinant of a triangular matrix is the product of its diagonal entries to more efficiently calculate the determinant.

### 3.2.1 Example:

Compute  $\det A$  for

$$A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$$

Solution:

$$\det A = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix}$$

As replacement does not change the determinant and interchange only changes the sign,

$$\det A = -(1)(3)(-5) = 15$$

If a square matrix  $A$  has been reduced to echelon form  $U$  by row replacements and  $r$  interchanges alone,

$$\det A = (-1)^r \det U$$

Further,

$$\det A = \begin{cases} (-1)^r \cdot \prod_{i=1}^n U_{ii} & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases}$$

**A square matrix  $A$  is invertible iff  $\det A \neq 0$**

Numerical note: the row reduction method of calculating the determinant takes only about  $2n^3/3$  operations compared to the  $n!$  of cofactor expansion.

If  $A$  is an  $n \times n$  matrix, then  $\det A^T = \det A$



### 3.2.2 Determinants and Matrix Products

*Multiplicative property:* For square matrices A and B,

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

NOTE:  $\det(A + B)$  is usually *not* equal to  $\det A + \det B$

## 3.3 TOPIC 3: Cramer's rule, volume, and linear transformations

*Cramer's rule:* Let A be an invertible  $n \times n$  matrix and  $A_i(\vec{b})$  be the matrix obtained from A by replacing column  $i$  by the vector  $\vec{b}$ . For any  $\vec{b}$  in  $\mathbb{R}^n$ , the unique solution of  $A\vec{x} = \vec{b}$  has entries given by

$$x_i = \frac{\det A_i(\vec{b})}{\det A}, \quad i = 1, 2, \dots, n$$

Example: Use Cramer's rule to solve the system

$$3x_1 - 2x_2 = 6 \tag{4}$$

$$-5x_1 + 4x_2 = 8 \tag{5}$$

Solution:

$$A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}, \quad A_1(\vec{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, \quad A_2(\vec{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

Since  $\det A = 2$  the system has a unique solution. By Cramer's rule:

$$x_1 = \frac{\det A_1(\vec{b})}{\det A} = \frac{24 + 16}{2} = 20$$
$$x_2 = \frac{\det A_2(\vec{b})}{\det A} = \frac{24 + 30}{2} = 27$$

*Laplace transforms:* an approach, useful in many engineering problems, which involves converting an appropriate system of linear differential equations into a system of linear algebraic equations whose coefficients involve a parameter.

*Adjugate (classical adjoint) of A:* the transpose of the matrix of cofactors of A

$$\text{adj}A = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

For an invertible square matrix  $A$ ,

$$A^{-1} = \frac{1}{\det A} \text{adj} A$$

Cramer's rule is a theoretical tool for studying the tolerance of the solution to  $Ax = b$  to changes in entry  $b$  or  $A$ . It is sometimes useful for hand calculation of  $3 \times 3$  matrices with complex entries. Usually, *it is hopelessly inefficient*.

### 3.3.1 Determinants as Area or Volume

If  $A$  is  $2 \times 2$ , the area of the parallelogram determined by the columns of  $A$  is  $|\det A|$ .

If  $A$  is  $3 \times 3$ , the volume of the parallelepiped determined by the columns of  $A$  is also  $|\det A|$ .

Example: Calculate the area of the parallelogram determined by the points  $(-2, -2)$ ,  $(0, 3)$ ,  $(4, -1)$ , and  $(6, 4)$ .

Solution: First, translate the parallelogram to the origin by subtracting  $(-2, -2)$  from each of the other vertices. The new vertices have coordinates  $(0, 0)$ ,  $(2, 5)$ ,  $(6, 1)$ , and  $(8, 6)$ . This is in the Span of

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}$$

And  $|\det A| = -28$ , so the area of the parallelogram is 28.

### 3.3.2 Linear transformations

Let  $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$ . If  $S$  is a parallelogram in  $\mathbb{R}^2$  and  $T(S)$  is the set of images of points of  $S$ ,

$$\{\text{area of } T(S)\} = |\det A| \{\text{area of } S\}$$

This statement is analogous for volume in  $\mathbb{R}^3$ .

## 3.4 TOPIC 4: Markov Chains

*Probability vector*: a vector with nonnegative entries that add up to 1

*Stochastic matrix*: a square matrix whose columns are probability vectors

*Markov Chain:* a model used to describe an experiment or measurement which is performed many times in the same way, where the outcome of each trial depends only on the immediately preceding trial. It can also be represented as a sequence of probability vectors  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$  together with a stochastic matrix  $P$  such that

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P\mathbf{x}_1, \quad \mathbf{x}_3 = P\mathbf{x}_2, \dots$$

Thus the Markov chain is described by the first-order difference equation

$$\mathbf{x}_{k+1} = P\mathbf{x}_k \quad \text{for } k = 0, 1, 2, \dots$$

In general,  $\mathbf{x}_k = P^k \mathbf{x}_0$

When a Markov chain of vectors is in  $\mathbb{R}^n$ , the entries in  $\mathbf{x}_k$  (called the *state vector*) list the probabilities that the system is in each of  $n$  possible states or that the outcome of the experiment is one of  $n$  possible outcomes.

*Steady-state vector (equilibrium vector):* If  $P$  is a stochastic matrix, then a steady-state vector for  $P$  is a probability vector  $\mathbf{q}$  such that

$$P\mathbf{q} = \mathbf{q}$$

Every stochastic matrix has a steady-state vector

To find the steady-state vector solve the equation  $P\mathbf{x} = \mathbf{x}$ :

$$\begin{aligned} P\mathbf{x} - \mathbf{x} &= \mathbf{0} \\ P\mathbf{x} - I\mathbf{x} &= \mathbf{0} \\ (P - I)\mathbf{x} &= \mathbf{0} \end{aligned}$$

Which can be solved by row reduction to find the general solution.

Next, choose a simple basis for the solution space and divide that by the sum of its entries to find a probability vector  $\mathbf{q}$  in the set of solutions of  $P\mathbf{x} = \mathbf{x}$ .

*Regular:* a quality of a stochastic matrix where some matrix power of  $P^k$  contains only strictly positive entries.

**If  $P$  is an  $n \times n$  regular stochastic matrix, then  $P$  has a unique steady-state vector  $\mathbf{q}$ . Further, if  $x_0$  is any initial state and  $x_{k+1} = Px_k$  for  $k = 0, 1, 2, \dots$  then the Markov chain  $\{x_k\}$  converges to  $\mathbf{q}$  as  $k \rightarrow \infty$**

### 3.5 TOPIC 5: Eigenvalues and Eigenvectors

*Eigenvector:* a nonzero vector  $x$  such that  $Ax = \lambda x$  for some  $n \times n$  matrix  $A$  and scalar  $\lambda$

*Eigenvalue:* a scalar  $\lambda$  for which there is a nontrivial solution of  $Ax = \lambda x$  (where  $x$  is an eigenvector corresponding to  $\lambda$ )

Example: Show that 7 is an eigenvalue of  $\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$

7 is an eigenvalue of  $A$  iff

$$Ax = 7x$$

Thus,

$$\begin{aligned} Ax - 7x &= 0 \\ (A - 7I)x &= 0 \\ A - 7I &= \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix} \end{aligned}$$

The columns of  $A$  are linearly dependent so 7 is an eigenvalue of  $A$ . The eigenvectors are:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So, every vector of the form  $x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  ( $x_2 \neq 0$ ) is an eigenvector corresponding to  $\lambda = 7$

■

Warning: row reduction can be used to find eigenvectors but not eigenvalues. An echelon form of matrix  $A$  usually does not display the eigenvalues of  $A$ .

$\lambda$  is an eigenvalue of an  $n \times n$  matrix  $A$  iff the equation  $(A - \lambda I)x = 0$  has a nontrivial solution.

The set of all solutions of that equation is just the nullspace of  $A - \lambda I$  so this set is a subspace of  $\mathbb{R}^n$

*Eigenspace:* the subspace of  $\mathbb{R}^n$  consisting of the zero vector and all the eigenvectors corresponding to  $\lambda$

A scalar  $\lambda$  will be an eigenvalue of  $A$  if  $(A - \lambda I)x = 0$  has free variables

### 3.5.1 Difference equations:

Eigenvectors can be used to construct solutions of the first-order difference equation

$$x_{k+1} = Ax_k \quad (k = 0, 1, 2, \dots)$$

A solution of this equation is an explicit description of the sequence  $\{x_k\}$  whose formula for each  $x_k$  does not depend directly on  $A$  or on the preceding terms in the sequence. The simplest way to build this solution is to take an eigenvector  $x_0$  and a corresponding eigenvalue such that

$$x_k = \lambda^k x_0$$

Which is a solution because

$$Ax_k = A(\lambda^k x_0) = \lambda^k (Ax_0) = \lambda^k (\lambda x_0) = \lambda^{k+1} x_0 = x_{k+1}$$

Eigenvalue Theorems:

1. The eigenvalues of a triangular matrix are the entries on its main diagonal
2.  $0$  is an eigenvalue of  $A \iff A$  is not invertible
3. Stochastic matrices have an eigenvalue equal to  $1$
4. If  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$  are eigenvectors that correspond to distinct eigenvalues, then those eigenvectors are linearly independent

From 3Blue1Brown: An easier way to compute simple eigenvalues is by recognising certain geometric characteristics.

1.  $\frac{1}{2} \text{tr} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{a+d}{2} = m$
2.  $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc = \lambda_1 \lambda_2 = p$
3.  $\lambda_1, \lambda_2 = m \pm \sqrt{m^2 - p}$

This is equivalent to computing  $(x - \lambda_1)(x - \lambda_2) = x^2 - (\lambda_1 + \lambda_2)x + \lambda_1 \lambda_2$  with the quadratic formula

## 3.6 TOPIC 6: The Characteristic Equation

A scalar  $\lambda$  is an eigenvalue of an  $n \times n$  matrix  $A$  iff  $\lambda$  satisfies the characteristic equation

$$\det(A - \lambda I) = 0$$

(Because a matrix is invertible iff 0 is not an eigenvalue of a square matrix  $A$  and the determinant is also not 0 by the Invertible Matrix Theorem)

*Characteristic polynomial of  $A$ :* the expanded form of  $\det(A - \lambda I)$  for an  $n \times n$  matrix which will be a polynomial of degree  $n$

*Trace:* the sum of the diagonal elements of its matrices

When a  $(2 \times 2)$  matrix is singular, one eigenvalue is 0 and the other is the trace.

Warnings:

1. Matrices can have the same eigenvalues but not be similar
2. Similarity is not the same as row equivalence
3. Row operations on a matrix usually change its eigenvalues

*Algebraic multiplicity:* the multiplicity of an eigenvalue as a root of the characteristic polynomial

*Geometric multiplicity:* the dimension of  $\text{Null}(A - \lambda I)$

- Geometric multiplicity is always at least 1

For an  $n \times n$  matrix  $A$  with  $a_i$  as the algebraic multiplicity of  $\lambda_i$  and  $g_i$  the geometric multiplicity:

1.  $1 \leq a_i \leq n$
2.  $1 \leq g_i \leq a_i$

### 3.6.1 Similarity

If  $A$  and  $B$  are square matrices, then  $A$  is similar to  $B$  if there is a matrix  $P$  such that  $P^{-1}AP = B$ . If so,  $A$  and  $B$  have the same characteristic polynomial and hence the same eigenvalues with the same multiplicities.

This can lead to a more efficient algorithm for calculating large powers of  $A^k$

### 3.7 TOPIC 7: Diagonalisation

In many cases, the eigenvalue-eigenvector information of a matrix  $A$  can be displayed in the factorisation  $A = PDP^{-1}$  where  $D$  is a diagonal matrix.

This can be especially useful for quickly calculating large values of  $A^k$  because, in general,

$$D^k = \begin{bmatrix} a^k & 0 \\ 0 & b^k \end{bmatrix}$$

*Diagonalisable:* characteristic of a square matrix if it is similar to a diagonal matrix. In other words,  $A = PDP^{-1}$  for invertible  $P$  and diagonal  $D$ .

*The Diagonalisation Theorem:* an  $n \times n$  matrix  $A$  is diagonalisable iff  $A$  has  $n$  linearly independent eigenvectors. In other words,  $A$  is diagonalisable iff there are enough eigenvectors to form a basis of  $\mathbb{R}^n$  (the *eigenvector basis of  $\mathbb{R}^n$* )

#### 3.7.1 Diagonalising matrices

1. Find the eigenvalues of  $A$  (usually by solving  $\det(A - \lambda I) = 0$  with the use of a computer)
2. Find  $n$  linearly independent eigenvectors of  $A$  (if this step fails,  $A$  is not diagonalisable)
3. Construct  $P$  from the vectors of step 2  $P = [v_1 \ \dots \ v_n]$
4. Construct  $D$  from the corresponding eigenvalues in the order which matches the order of the columns of  $P$

*An  $n \times n$  matrix with  $n$  distinct eigenvalues is diagonalisable*

Let  $A$  be an  $n \times n$  matrix whose distinct eigenvalues are  $\lambda_1, \dots, \lambda_p$

- For  $1 \leq p$ , the dimension of the eigenspace for  $\lambda_k$  is less than or equal to the multiplicity of  $\lambda_k$
- $A$  is diagonalisable iff the sum of the dimensions of the eigenspace equals  $n$   
 $\iff$  the characteristic polynomial factors completely into linear factors or the dimension of the eigenspace for each  $\lambda_k$  equals the multiplicity of  $\lambda_k$
- If  $A$  is diagonal and  $\mathfrak{B}_k$  is a basis for the eigenspace corresponding to  $\lambda_k$  for each  $k$ , the total collection of vectors in  $\mathfrak{B}_1, \dots, \mathfrak{B}_p$  forms an eigenvector basis for  $\mathbb{R}^n$

### 3.8 TOPIC 8: Complex Eigenvalues

A complex scalar  $\lambda$  satisfies  $\det(A - \lambda I) = 0$  iff there is a nonzero vector  $\vec{x} \in \mathbb{C}^n$  such that  $A\vec{x} = \lambda\vec{x}$

Example: Find a basis for the eigenspace of  $A = \begin{bmatrix} 0.5 & -0.6 \\ 0.75 & 1.1 \end{bmatrix}$

Solution:

$$\begin{aligned} 0 &= \begin{vmatrix} 0.5 - \lambda & -0.6 \\ 0.75 & 1.1 - \lambda \end{vmatrix} = (0.5 - \lambda)(1.1 - \lambda) - (-0.6)(0.75) \\ &= \lambda^2 - 1.6\lambda + 1 \end{aligned}$$

From the quadratic formula,

$$\lambda = \frac{1}{2}[1.6 \pm \sqrt{(-1.6)^2 - 4}] = 0.8 \pm 0.6i$$

For  $\lambda = 0.8 - 0.6i$ ,

$$A - (0.8 - 0.6i)I = \begin{bmatrix} -0.3 + 0.6i & -0.6 \\ 0.75 & 0.3 + 0.6i \end{bmatrix}$$

Row reduction is difficult but since  $0.8 - 0.6i$  is an eigenvalue, the system

$$\begin{aligned} (-0.3 + 0.6i)x_1 - 0.6x_2 &= 0 \\ 0.75x_1 + (0.3 + 0.6i)x_2 &= 0 \end{aligned}$$

has a nontrivial solution and either equation can be used to represent the other variable.

$$\begin{aligned} 0.75x_1 &= (-0.3 - 0.6i)x_2 \\ x_1 &= (-0.4 - 0.8i)x_2 \end{aligned}$$

So,  $x_1 = -2 - 4i$  when  $x_2$  is chosen to be 5 to eliminate the decimals. The basis for the eigenspace corresponding to  $\lambda = 0.8 - 0.6i$  is thus

$$v_1 = \begin{bmatrix} -2 - 4i \\ 5 \end{bmatrix}$$

Analogous calculations for  $\lambda = 0.8 + 0.6i$  yields

$$v_2 = \begin{bmatrix} -2 + 4i \\ 5 \end{bmatrix}$$



### 3.8.1 Real and Imaginary parts of vectors

For some complex vector,  $x$

$$x = \Re x + i\Im x$$

and its complex conjugate is

$$\bar{x} = \Re x - i\Im x$$

Properties of complex matrix algebra:

- $r\bar{x} = \bar{r}x$
- $\bar{B}x = B\bar{x}$
- $\bar{B}C = \bar{B}\bar{C}$
- $r\bar{B} = \bar{r}B$

*When an  $n \times n$  matrix  $A$  is real, its complex eigenvalues occur in conjugate pairs*

For  $A$  an  $n \times n$  matrix with real entries where  $\lambda$  is an eigenvalue of  $A$  and  $x$  is a corresponding eigenvector in  $\mathbb{C}^n$ , then  $\bar{\lambda}$  is an eigenvalue with  $\bar{x}$  a corresponding eigenvector.

If  $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$  where  $a$  and  $b$  are real and not both zero, then the eigenvalues of  $C$  are  $\lambda = a \pm bi$ . If  $r = |\lambda| = \sqrt{a^2 + b^2}$ , then

$$C = r \begin{bmatrix} a/r & -b/r \\ b/r & a/r \end{bmatrix} = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}$$

*Argument of  $\lambda = a + bi$ :*  $\phi$ , the angle between the positive x-axis and the ray from  $(0, 0)$  through  $(a, b)$ .

Thus, the transformation  $\vec{x} \mapsto C\vec{x}$  may be viewed as the composition of a rotation through the angle  $\phi$  and a scaling by  $|\lambda|$

Let  $A$  be a real  $2 \times 2$  matrix with a complex eigenvalue  $\lambda = a - bi$  ( $b \neq 0$ ) and an associated eigenvector  $v \in \mathbb{C}^2$ .

Then  $A = PCP^{-1}$ , where  $P = [\Re v_1 \quad \Im v_1]$  and  $C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$

*Invariant:* property of a space on which  $A$  acts as a rotation (with or without scaling) where every vector in that space is rotated into another point *on that same space*

### 3.9 TOPIC 9: Google Page Rank

If  $P$  is a regular  $m \times m$  matrix with  $m \geq 2$ , the following statements are all true:

1. There is a stochastic matrix  $\Pi$  such that  $\lim_{n \rightarrow \infty} P^n = \Pi$
2. Each column of  $\Pi$  is the same probability vector  $q$
3. For any initial probability vector  $x_0$ ,  $\lim_{n \rightarrow \infty} P^n x_0 = q$
4. The vector  $q$  is the unique probability vector which is an eigenvector of  $P$  associated with the eigenvalue 1
5. All eigenvalues of  $P$  other than 1 have  $|\lambda| < 1$

Because The World Wide Web can be modeled as a directed graph, with vertices representing webpages and arrows representing links between pages. Let  $P$  be the huge transition matrix for this Markov chain. If  $P$  were regular, there is a vector  $q$  with entries representing occupation times. Because some "important" pages have links from other "important" pages and spend more time in that small group of important pages, occupation time correlates with the importance of a page, allowing Google to rank pages by the relative size of the corresponding entry in  $q$  for an appropriately chosen Markov chain.

In general, a number of adjustments have to be made to regularise a simple random walk on the directed graph model:

1. If  $j$  is a dangling node (absorbing state), replace column  $j$  of  $P$  with  $\begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix}$  to form the matrix  $P_*$
2. To correct for cycles, let  $p$  be the probability a surfer will pick from all possible links from page  $j$  with equal probability,  $1 - p$  is the probability the surfer will pick *any* web page so the new transition matrix will be

$$G = pP_* + (1 - p)K$$

where  $K$  is an  $n \times n$  matrix whose columns are  $\begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix}$

Google is said to use a value  $p = 0.85$ . Furthermore, the true PageRank algorithm uses a  $K$  whose columns are a probability vector  $v$  which can be different between individuals or groups of searchers.

Though Google uses a variation of the Power Method of computing eigenvalues to compute the eigenvector associated with the eigenvalue 1 and it only takes between 50-100 iterations for Google-level accuracy, the computation of a new  $q$  takes Google days and is repeated every month.

## 4 Module 4: Orthogonality

### 4.1 TOPIC 1: Inner product, Length, Orthogonality

*The Inner Product:* the number  $\vec{u}^T \vec{v}$  for two  $n \times 1$  vectors, often written  $\vec{u} \cdot \vec{v}$  and also called the dot product.

In other words,

$$u \cdot v = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n$$

Inner product theorems:

1.  $u \cdot v = v \cdot u$
2.  $(u + v) \cdot w = u \cdot w + v \cdot w$
3.  $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
4.  $u \cdot u \geq 0, \quad u \cdot u = 0 \iff u = \vec{0}$

*Norm:* the length of  $v$  is the nonnegative scalar  $\|v\|$  defined by

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

For any scalar  $c$ ,

$$\|c\vec{v}\| = |c|\|v\|$$

*Unit vector:* a vector whose length is 1

*Normalising:* the process of creating a unit vector  $u$  in the same direction as some vector  $v$  by multiplying  $v$  by  $1/\|v\|$

*Distance in  $\mathbb{R}^n$ :* for  $u, v \in \mathbb{R}^n$ ,

$$\text{dist}(u, v) = \|u - v\|$$

#### 4.1.1 Orthogonal vectors

Two lines through the origin in  $\mathbb{R}^2$  or  $\mathbb{R}^3$  determined by vectors  $u, v$  are geometrically perpendicular iff the distance from  $u$  to  $v$  is the same as distance from  $u$  to  $-v$ .

$$\begin{aligned} [\text{dist}(u, -v)]^2 &= \|u - (-v)\|^2 = \|u + v\|^2 \\ &= (u + v) \cdot (u + v) \\ &= u \cdot (u + v) + v \cdot (u + v) \\ &= u \cdot u + u \cdot v + v \cdot u + v \cdot v \\ &= \|u\|^2 + \|v\|^2 + 2u \cdot v \end{aligned}$$

Two squared distances are equal iff  $2u \cdot v = -2u \cdot v$  which happens iff  $u \cdot v = 0$

**Two vectors  $u, v \in \mathbb{R}^n$  are orthogonal iff**

$$u \cdot v = 0 \iff \|u + v\|^2 = \|u\|^2 + \|v\|^2$$

#### 4.1.2 Orthogonal complements

If a vector  $z$  is orthogonal to every vector in a subspace  $W$  of  $\mathbb{R}^n$ ,  $z$  is orthogonal to  $W$

*Orthogonal complement of  $W$ :*  $W^\perp$ , the set of all vectors that are orthogonal to  $W$

Orthogonal complement theorems:

1. A vector  $x$  is in  $W^\perp$  iff  $x$  is orthogonal to every vector in a set that spans  $W$
2.  $W^\perp$  is a subspace of  $\mathbb{R}^n$
3.  $(\text{Row } A)^\perp = \text{Nul } A, \quad (\text{Col } A)^\perp = \text{Nul } A^T$

#### 4.1.3 Angles in $\mathbb{R}^2$ and $\mathbb{R}^3$

$$u \cdot v = \|u\| \|v\| \cos \theta$$

## 4.2 TOPIC 2: Orthogonal Sets

*Orthogonal set:* a set of vectors  $\{u_1, \dots, u_p\} \in \mathbb{R}^n$  if each pair of distinct vectors from the set is orthogonal ( $u_i \cdot u_j = 0, i \neq j$ )

If  $S = \{u_1, \dots, u_p\}$  is an orthogonal set of nonzero vectors in  $\mathbb{R}^n$ , then  $S$  is linearly independent and hence is a basis for the subspace spanned by  $S$

*Orthogonal basis:* a basis for a subspace  $W$  in  $\mathbb{R}^n$  that is also an orthogonal set

For  $\{u_1, \dots, u_p\}$  an orthogonal basis for  $W$ ,  $\forall y \in W$ ,

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$$

#### 4.2.1 Orthogonal Projection

Given a nonzero vector  $u$  in  $\mathbb{R}^n$ , we wish to decompose a vector  $y$  into

$$y = \hat{y} + z$$

where  $\hat{y} = \alpha u$  for some scalar  $\alpha$  and  $z$  is some vector orthogonal to  $u$ . Thus,

$$\begin{aligned}\alpha &= \frac{y \cdot u}{u \cdot u} \\ \hat{y} &= \frac{y \cdot u}{u \cdot u} u\end{aligned}$$

The vector  $\hat{y}$  is called the orthogonal projection of  $y$  onto  $u$  and the vector  $z$  is called the component of  $y$  orthogonal to  $u$ .

This projection is actually determined by the subspace  $L$  spanned by  $u$ .

Therefore:

$$\hat{y} = \text{proj}_L y = \frac{y \cdot u}{u \cdot u} u$$

#### 4.2.2 Orthonormal Sets

*Orthonormal set:* a set  $\{u_1, \dots, u_p\}$  if it is an orthogonal set of unit vectors. If  $W$  is the subspace spanned by such a set, then  $\{u_1, \dots, u_p\}$  is an orthonormal basis for  $W$

A set is an orthonormal basis if:

1. It is an orthogonal set  $\iff$  every pair of vectors is orthogonal  $\iff$  all the dot products are zero
2. It is a set of unit vectors  $\iff v_i \cdot v_i = 1$
3. The set forms a basis for  $\mathbb{R}^n \iff$  the vectors are linearly independent

An  $m \times n$  matrix  $U$  has orthonormal columns iff  $U^T U = I$

Let  $U$  be an  $m \times n$  matrix with orthonormal columns, for  $x, y \in \mathbb{R}^n$ :

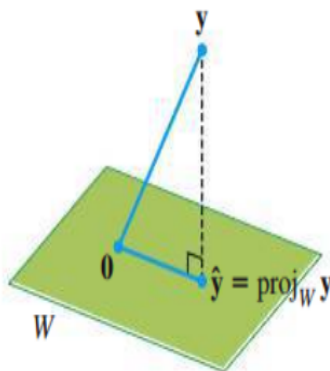
1.  $\|Ux\| = \|x\|$
2.  $(Ux) \cdot (Uy) = x \cdot y$
3.  $(Ux) \cdot (Uy) = 0 \iff x \cdot y = 0$

Thus, the mapping  $x \mapsto Ux$  preserves lengths and orthogonality.

*Orthogonal matrix:* a square invertible matrix such that  $U^{-1} = U^T$ . Any square matrix with orthonormal columns is an orthogonal matrix and also has orthonormal rows

### 4.3 TOPIC 3: Orthogonal Projections

Given a vector  $y$  and a subspace  $W$  in  $\mathbb{R}^n$ , there is a vector  $\hat{y}$  in  $W$  such that  $\hat{y}$  is the unique vector in  $W$  for which  $y - \hat{y}$  is orthogonal to  $W$ , and  $\hat{y}$  is the unique vector in  $W$  closest to  $y$ .



Observe that whenever a vector  $y$  is written as a linear combination of vectors  $u_1, \dots, u_n \in \mathbb{R}^n$ , the terms in the sum for  $y$  can be grouped into two parts so that  $y$  can be written as

$$y = z_1 + z_2$$

where  $z_1$  is a linear combination of some of the  $u_i$  and  $z_2$  is a linear combination of the rest of the  $u_i$ . This idea is particularly useful when  $\{u_1, \dots, u_n\}$  is an orthogonal basis.

The Orthogonal Decomposition Theorem: Let  $W$  be a subspace of  $\mathbb{R}^n$ . Then each  $y \in \mathbb{R}^n$  can be written uniquely in the form

$$y = \hat{y} + z$$

where  $\hat{y} \in W, z \in W^\perp$ . In fact,  $\{u_1, \dots, u_p\}$  is any orthogonal basis of  $W$ , then

$$\begin{aligned}\hat{y} &= \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \dots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \\ z &= y - \hat{y}\end{aligned}$$

This is a more general form of projection onto a line – projection onto the subspace  $W$ .

#### 4.3.1 Properties of Orthogonal Projections

If  $\{u_1, \dots, u_p\}$  is an orthogonal basis for  $W$  and if  $y$  happens to be in  $W = \text{Span}\{u_1, \dots, u_p\}$ , then  $\text{proj}_W y = y$

The Best Approximation Theorem: Let  $W$  be a subspace of  $\mathbb{R}^n$ , let  $y$  be any vector in  $\mathbb{R}^n$ , and let  $\hat{y}$  be the orthogonal projection of  $y$  onto  $W$ . Then,

$$\|y - \hat{y}\| < \|y - v\|$$

for all  $v$  in  $W$  distinct from  $\hat{y}$  (" $\hat{y}$  is the closest point in  $W$  to  $y$ ")

Finally, If  $\{u_1, \dots, u_p\}$  is an orthonormal basis for a subspace  $W$  of  $\mathbb{R}^n$ , then

$$\text{proj}_W y = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \dots + (y \cdot u_p)u_p$$

If  $U = \begin{bmatrix} u_1 & u_2 & \dots & u_p \end{bmatrix}$ , then

$$\text{proj}_W y = UU^T y \quad \forall y \in \mathbb{R}^n$$

## 4.4 TOPIC 4: The Gram-Schmidt Process

*The Gram-Schmidt Process:* a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of  $\mathbb{R}^n$

Example: Let  $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ . Then,  $\{x_1, x_2, x_3\}$  is a basis for subspace  $W$  of  $\mathbb{R}^4$ . Construct an orthogonal basis for  $W$ .

Solution:



1. Step 1: Let  $v_1 = x_1$  and  $W_1 = \text{Span}\{x_1\} = \text{Span}\{v_1\}$
2. Let  $v_2$  be the vector

$$\begin{aligned}
 v_2 &= x_2 - \text{proj}_{W_1} x_2 \\
 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\
 &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}
 \end{aligned}$$

Thus,  $\{v_1, v_2\}$  is an orthogonal basis for the subspace spanned by  $x_1$  and  $x_2$

3. (Optional) Scale  $v_2$  to simplify later computations, making

$$\{v_1, v_2\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

4. Let  $v_3$  be the vector  $x_3 - \text{proj}_{W_2} x_3$  where

$$\text{proj}_{W_2} x_3 = \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x_3 \cdot v'_2}{v'_2 \cdot v'_2} v'_2 = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Then

$$v_3 = x_3 - \text{proj}_{W_2} x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

To summarize: Given a basis  $\{x_1, \dots, x_p\}$  for a nonzero subspace  $W$  of  $\mathbb{R}^n$ , define

$$\begin{aligned} v_1 &= x_1 \\ v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\ v_3 &= x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} \\ &\vdots \\ v_p &= x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} - \frac{x_p \cdot v_2}{v_2 \cdot v_2} - \dots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} \end{aligned}$$

Then,  $\{v_1, \dots, v_p\}$  is an orthogonal basis for  $W$  and

$$\text{Span}\{v_1, \dots, v_k\} = \text{Span}\{x_1, \dots, x_k\} \quad \text{for } 1 \leq k \leq p$$

#### 4.4.1 Orthonormal Bases

*Orthonormal basis:* an orthogonal basis which has been normalized

Example:  $v_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$

Solution:

$$\begin{aligned} u_1 &= \frac{1}{\|v_1\|} v_1 = \frac{1}{\sqrt{45}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix} \\ u_2 &= \frac{1}{\|v_2\|} v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned}$$

#### 4.4.2 QR Factorization of Matrices

If an  $m \times n$  matrix has linearly independent columns  $x_1, \dots, x_n$ , then applying the Gram-Schmidt process with normalizations is equivalent to factoring  $A$ .

*The QR Factorization:* If  $A$  is an  $m \times n$  matrix with linearly independent columns, then  $A$  can be factored as  $A = QR$  where  $Q$  is an  $m \times n$  matrix whose columns form an orthonormal basis for  $\text{Col } A$  and  $R$  is an  $n \times n$  upper triangular invertible matrix with positive entries on its diagonal. Further,  $R = Q^T A$ .

## 4.5 TOPIC 5: Least Squares Problem

*The general least-squares problem:* question involving finding a close approximate  $x$  for an inconsistent equation  $Ax = b$

If  $A$  is  $m \times n$  and  $b$  is in  $\mathbb{R}^m$ , a least-squares solution of  $Ax = b$  is  $\hat{x}$  in  $\mathbb{R}^n$  such that

$$\|b - A\hat{x}\| \leq \|b - Ax\|$$

for all  $x$  in  $\mathbb{R}^n$

No matter what  $x$  we select,  $Ax$  will necessarily be in the column space. So we seek an  $x$  that makes  $Ax$  the closest point in  $\text{Col } A$  to  $b$ .

### 4.5.1 Solution of the general problem

Apply the Best Approximation Theorem

$$\hat{b} = \text{proj}_{\text{Col } A} b$$

Then, because  $\hat{b}$  is in the column space of  $A$ ,  $Ax = \hat{b}$  is consistent and there is a least-squares solution to

$$A\hat{x} = \hat{b}$$

The set of least-squares solutions of  $Ax = b$  coincides with the nonempty set of solutions of the normal equations

$$A^T Ax = A^T b$$

For an  $m \times n$  matrix  $A$ , the following statements are logically equivalent:

1. The equation  $Ax = b$  has a unique least-squares solution for each  $b$  in  $\mathbb{R}^m$
2. The columns of  $A$  are linearly independent
3. The matrix  $A^T A$  is invertible

When these statements are true,

$$\hat{x} = (A^T A)^{-1} A^T b$$

*Least-squares error:* the distance from  $b$  to  $A\hat{x}$

In some cases, the normal equations for a least-squares problem can be ill-conditioned. If the columns of  $A$  are linearly independent, the least-squares solution can often be computed more reliably through a QR factorisation:

$$\hat{x} = R^{-1}Q^Tb$$

Numerical note: Since  $R$  is upper triangular,  $\hat{x}$  should be calculated as the exact solution of

$$Rx = Q^Tb$$

## 4.6 TOPIC 6: Application to Linear Models

The simplest relation between two variables  $x$  and  $y$  is the linear equation  $y = \beta_0 + \beta_1x$ . Experimental data often produce points  $(x_1, y_1), \dots, (x_n, y_n)$  that lie close to a line.

*Least-squares line:* the line  $y = \beta_0 + \beta_1x$  that minimizes the sum of the squares of the residuals (errors). This line is also called a line of regression of  $y$  on  $x$ .

We can write this system as

$$X\beta = y$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then, computing the least-squares solution of  $X\beta = y$  is equivalent to finding the  $\beta$  that determines the least-squares line.

Example: Find the equation of the least-squares line that best fits the data points (2,1), (5,2), (7,3), and (8,3)

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

Then, computing  $X^T X \beta = X^T y$  yields:

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence,

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

And the least-squares equation has the form

$$y = \frac{2}{7} + \frac{5}{14}x$$

A common practice before computing a least-squares line is to form a new variable  $x^* = x - \bar{x}$  from the average of the original x-values. The new x-data are in mean-deviation form and the two columns of the design matrix will be orthogonal, simplifying solution of the normal equations.

#### 4.6.1 Least-squares fitting of other curves

*Residual vector:*  $\epsilon = y - X\beta$ , whose length should be minimized to find a least-squares solution.

For a quadratic curve:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$y = X\beta + \epsilon$$

Data can be fit to curves which have the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \dots + \beta_k f_k(x)$$

where  $f_0, \dots, f_k$  are known functions and  $\beta_0, \dots, \beta_k$  are parameters.

#### 4.6.2 Multiple regression

When an experiment involves two independent variables, the least-squares fit is called a trend surface and will be described by a linear model in the form

$$y = \beta_0 f_0(u, v) + \beta_1 f_1(u, v) + \dots + \beta_k f_k(u, v)$$

## 5 Module 5: Symmetric Matrices and Quadratic Forms

### 5.1 TOPIC 1: Diagonalisation of Symmetric Matrices

*Symmetric matrix:* a matrix  $A$  such that  $A^T = A$  which is necessarily square. Its diagonal entries are arbitrary but all others occur in pairs on the main diagonal.

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}$$

for example.

Diagonalisation review:

$$A = \begin{bmatrix} 6 & -2 & -1 \\ -2 & 6 & -1 \\ -1 & -1 & 5 \end{bmatrix}$$

The characteristic equation is

$$0 = -\lambda^3 + 17\lambda^2 - 90\lambda + 144 = -(\lambda - 8)(\lambda - 6)(\lambda - 3)$$

We can get a basis for each eigenspace:

$$\lambda = 8 : \vec{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}; \quad \lambda = 6 : \vec{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}; \quad \lambda = 3 : \vec{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

These vectors form an orthogonal basis for  $\mathbb{R}^3$ . The normalised basis eigenvectors are:

$$\vec{u}_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Thus,

$$A = PDP^{-1}$$
$$P = \begin{bmatrix} -1 & -1/\sqrt{6} & 1/\sqrt{3} \\ 1 & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}, \quad D = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Since  $P$  is square and an orthonormal matrix and  $P^{-1} = P^T$

**If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.**

*Orthogonally diagonalisable:* An  $n \times n$  matrix if there are an orthogonal matrix P and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

This diagonalisation requires  $n$  linearly independent and orthonormal eigenvectors.

An  $n \times n$  matrix A is orthogonally diagonalisable iff A is a symmetric matrix.

### 5.1.1 The Spectral Theorem

*Spectrum:* the set of eigenvalues of a matrix A

An  $n \times n$  symmetric matrix A has the following properties

1. A has  $n$  real eigenvalues, counting multiplicities
2. The dimension of the eigenspace for each eigenvalue  $\lambda$  equals the multiplicity of  $\lambda$  as a root of the characteristic equation
3. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal
4. A is orthogonally diagonalisable.

### 5.1.2 Spectral Decomposition:

For  $A = PDP^{-1}$  where the columns of P are orthonormal eigenvectors  $u_1, \dots, u_n$  of A and the corresponding eigenvalues  $\lambda_1, \dots, \lambda_n$  are in the diagonal matrix D. Thus:

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \dots + \lambda_n u_n u_n^T$$

Every term in this spectral decomposition is an  $n \times n$  matrix of rank 1 (e.g. every column of  $\lambda_1 u_1 u_1^T$  is a multiple of  $u_1$ ). Every matrix  $u_j u_j^T$  is a projection matrix such that  $(u_j u_j^T)x = \text{proj}_{\text{Span}\{u_j\}}x$

## 5.2 TOPIC 2: Quadratic Forms

*Quadratic form:* a function Q defined on  $\mathbb{R}^n$  whose value at a vector  $x \in \mathbb{R}^n$  can be computed by an expression of the form  $Q(x) = x^T A x$  where A is an  $n \times n$  symmetric matrix

Example: For  $x \in \mathbb{R}^3$ ,

$$\begin{aligned} Q(x) &= 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3 \\ &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned}$$

### 5.2.1 Change of Variable in a Quadratic Form

Using a change of variable  $y = P^{-1}x$  (where  $y$  is the coordinate vector of  $x$  relative to the basis of  $\mathbb{R}^n$  determined by the columns of  $P$ ),

$$x^T A x = (Py)^T A (Py) = y^T P^T A P y = y^T (P^T A P) y$$

and so the matrix of the quadratic form is  $P^T A P$ . Since  $A$  is symmetric, there is an orthogonal matrix  $P$  such that  $P^T A P = D$  and the quadratic form is just

$$y^T D y$$

### 5.2.2 Principal Axes

*The Principal Axes Theorem:* If  $A$  is an  $n \times n$  symmetric matrix, then there is an orthogonal change of variable  $x = Py$  that transforms the quadratic form  $x^T A x$  into a quadratic form  $y^T D y$  with no cross-product term.

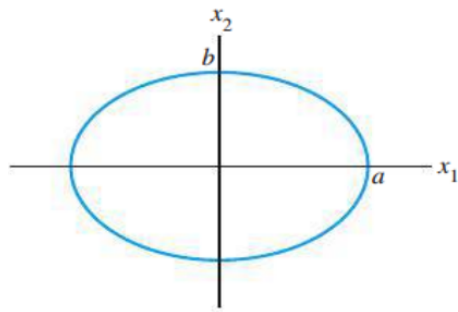
*Principal axes:* the columns of  $P$  in the theorem which form the orthonormal basis of  $\mathbb{R}^n$  for which  $y$  is the relative coordinate vector of  $x$ .

The set of all  $\vec{x} \in \mathbb{R}^2$  that satisfy  $x^T A x = c$  for some constant  $c$  corresponds to

1. an ellipse (or circle)
2. a hyperbola
3. two intersecting lines
4. a single point
5. or contains no points at all

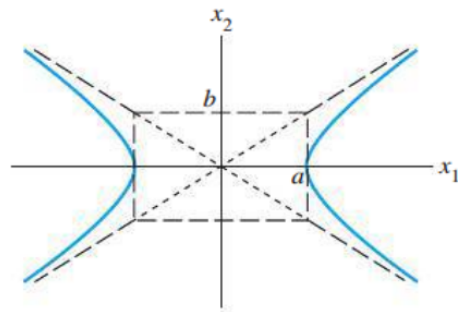


If A is a diagonal matrix, the graph is in standard position:



$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, \quad a > b > 0$$

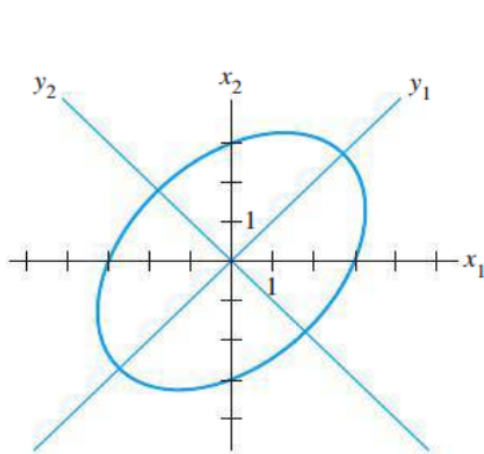
ellipse



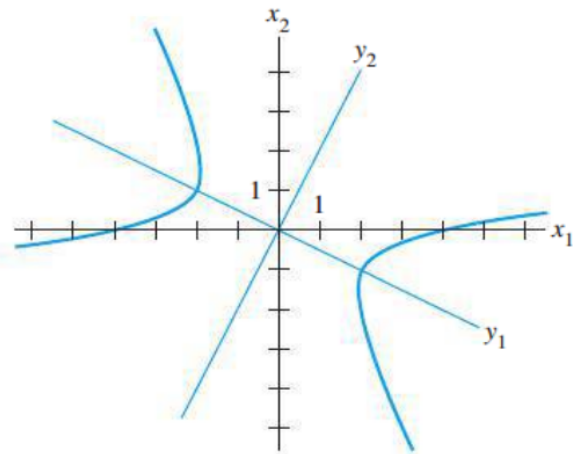
$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1, \quad a > b > 0$$

hyperbola

Else, the graph is rotated out of standard position and finding the principal axes (determined by the eigenvectors of A) amounts to finding a new coordinate system with respect to which the graph is in standard position.



(a)  $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$



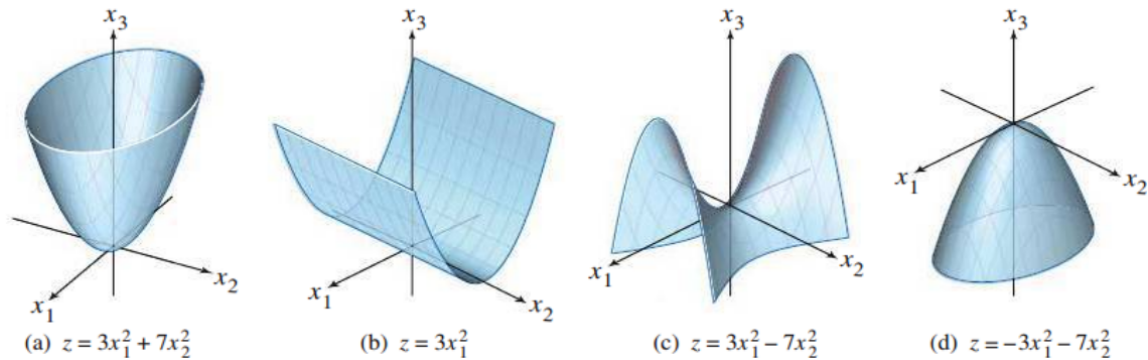
(b)  $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

### 5.2.3 Classifying Quadratic Forms

A quadratic form Q is:

1. positive definite if  $Q(x) > 0, \quad \forall x \neq 0$  (as in Fig A)

2. Negative definite  $Q(x) < 0, \quad \forall x \neq 0$  (as in Fig D)
3. Positive semidefinite  $Q(x) \geq 0, \quad \forall x \neq 0$  (as in Fig A and B)
4. Negative semidefinite  $Q(x) \leq 0, \quad \forall x \neq 0$  (as in Fig A and B)
5. indefinite if  $Q(X)$  assumes both positive and negative values (as in Fig C)



Let  $A$  be an  $n \times n$  symmetric matrix. Then a quadratic form  $x^T A x$  is:

1. positive definite iff the eigenvalues of  $A$  are all positive
2. negative definite iff the eigenvalues of  $A$  are all negative
3. indefinite iff  $A$  has both positive and negative eigenvalues

*Cholesky Factorisation:* this attempt to factor  $A$  into the form  $A = R^T R$  where  $R$  is upper triangular with positive diagonal entries can be a fast way to determine whether a symmetric matrix  $A$  is positive definite but is possible iff  $A$  is positive definite.

### 5.3 TOPIC 3: Constrained Optimization

Example: The surface of a unit sphere in  $\mathbb{R}^3$  is given by

$$1 = x_1^2 + x_2^2 + x_3^2 = \|\vec{x}\|^2$$

$Q$  is a quantity (for example, temperature) that we wish to optimize

$$Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

Identify the largest and smallest values of  $Q$  on the surface.

Solution:

$$Q = \vec{x}^T \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix} \vec{x} = 9x_1^2 + 4x_2^2 + 3x_3^2$$

$$\leq 9x_1^2 + 9x_2^2 + 9x_3^2$$

$$= 9(x_1^2 + x_2^2 + x_3^2) = 9\|\vec{x}\|^2 = 9$$

Because  $\|\vec{x}\|^2 = 1$ , the max occurs at  $\vec{x} = \begin{bmatrix} \pm 1 \\ 0 \\ 0 \end{bmatrix}$ . Similarly,  $\min Q(\vec{x}) : \|\vec{x}\| = 1\} = 3$

which occurs at  $\vec{x} = \begin{bmatrix} 0 \\ 0 \\ \pm 1 \end{bmatrix}$ . The minimum and maximum values of  $Q$  were the eigenvalues of  $A$  and the corresponding eigenvectors gave their location.

### 5.3.1 The General Form

If  $Q = x^T A x$ ,  $A$  is a real  $n \times n$  symmetric matrix, with eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$$

and with associated normalised eigenvectors

$$u_1, u_2, \dots, u_n$$

Then, subject to the constraint  $\|x\| = 1$ ,

1. the maximum of  $Q(x) = \lambda_1$ , attained at  $x = \pm u_1$
2. the minimum of  $Q(x) = \lambda_n$ , attained at  $x = \pm u_n$

### 5.3.2 Constrained optimisation with a repeated eigenvalue

Example: Calculate the max and min of  $Q = x_1^2 + 2x_2x_3$  subject to  $\|x\| = 1$

Solution:

$$Q = x^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} x$$

By inspection,

$$\lambda = 1 \implies v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Because A is symmetric, the eigenvector for  $\lambda = -1$  must be orthogonal to  $v_1$  and  $v_2$ . By inspection,

$$v_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Therefore, the minimum value of Q is -1, obtained at  $\pm v_3$  and the max is 1, obtained at a vector in the span of  $v_1$  and  $v_2$

### 5.3.3 Orthogonality Constraints

If  $Q = x^T A x$ , A is a real  $n \times n$  symmetric matrix, with eigenvalues

$$\lambda_1 \geq \lambda_2 \dots \geq \lambda_n$$

and with associated normalised eigenvectors

$$u_1, u_2, \dots, u_n$$

Then, subject to the constraints  $\|x\| = 1$  and  $\vec{x} \cdot \vec{u}_1 = 0$ ,

1. the maximum of  $Q(x) = \lambda_2$ , attained at  $x = u_2$
2. the minimum of  $Q(x) = \lambda_n$ , attained at  $x = u_n$

## 5.4 TOPIC 4: The Singular Value Decomposition

*Singular value:* the square roots of the eigenvalues of  $A^T A$  for some real  $m \times n$  A, usually arranged in decreasing order  $\sigma_1 \dots \sigma_n$  In general,

$$\sigma_i = \|A \vec{v}_i\|$$

where A is a linear transform acting on the unit vector v.

The eigenvalues of  $A^T A$  will always be real and non-negative

*Singular vectors:* the eigenvectors of  $A^T A$

Suppose that  $v_i$  are the ordered  $n$  orthogonal eigenvectors of  $A^T A$  and that  $A$  has  $r$  non-zero singular values ( $r \leq n$ ). Then the set of vectors

$$\{v_{r+1}, v_{r+2}, \dots, v_n\}$$

is an orthogonal basis for  $\text{Nul } A$  and the set

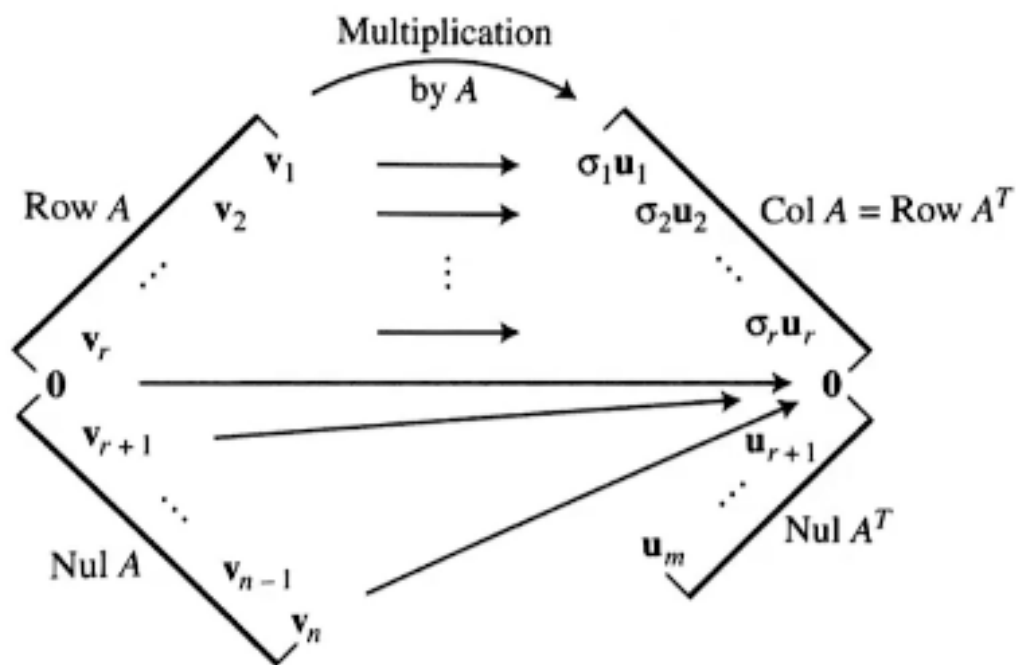
$$\{v_1, v_2, \dots, v_r\}$$

is an orthogonal basis for  $\text{Row } A$  and  $\text{rank } A = r$

If those  $v_i$  are orthonormal, then

$$\{Av_1, Av_2, \dots, Av_r\}$$

are an orthogonal basis for  $\text{Col } A$ .



Summary:

Examples:  $A \in \mathbb{R}^{12 \times 4}$  with 3 non-zero singular values. Indicate whether the following statements are true or false

1. A basis for  $\text{Col } A$  is given by  $Av_1, Av_2, Av_3$
2. A basis for  $\text{Nul } A$  is given by  $v_1, v_2, v_3$

Solutions:

1. True: A has rank 3, the vectors are independent and in Col A, so they form a basis for Col A
2. False: A has rank 3, so the vectors form a basis for Row A.  $v_4$  forms a basis for Nul A

#### 5.4.1 The SVD

Suppose A is an  $m \times n$  matrix with singular values  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$  and  $m \geq n$ . Then A has the decomposition

$$A = U\Sigma V^T$$

$$\Sigma = \begin{bmatrix} D & \\ \vec{0}_{m-n,n} & \end{bmatrix}, \quad D = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix}$$

U is a  $m \times m$  orthogonal matrix and V is an  $n \times n$  orthogonal matrix.

If  $m < n$ , then  $\Sigma = \begin{bmatrix} D & \vec{0}_{m,n-m} \end{bmatrix}$  with everything else the same.

A Procedure for Constructing the SVD: Suppose A is  $m \times n$  and has rank r.

1. Compute the squared singular values of  $A^T A$  ( $\sigma_i^2$ ) and construct  $\Sigma$
2. Compute the unit singular vectors of  $A^T A$  ( $v_i$ ) and use them to form V
3. Compute an orthonormal basis for Col A using

$$u_i = \frac{1}{\sigma_i} A v_i, \quad i = 1, 2, \dots, r$$

If necessary, extend  $\{u_i\}$  to form an orthonormal basis for  $\mathbb{R}^m$  and use the basis to form U.

Example: Construct the SVD for  $A = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

Solution:

$$A^T A = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} \implies \lambda_1 = 9, \lambda_2 = 4$$

Thus,

$$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Next,

$$\begin{aligned} A^T A - \lambda_1 I &= \begin{bmatrix} -5 & 0 \\ 0 & 0 \end{bmatrix} \implies v_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ A^T A - \lambda_2 I &= \begin{bmatrix} -5 & 0 \\ 0 & 0 \end{bmatrix} \implies v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \implies V &= [\vec{v}_1 \quad \vec{v}_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{aligned}$$

Next we construct left-singular unit vectors in  $\mathbb{R}^4$

$$\begin{aligned} u_1 &= \frac{1}{\sigma_1} A v_1 = \frac{1}{3} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \\ u_2 &= \frac{1}{\sigma_2} A v_2 = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

$u_3$  and  $u_4$  can be any two vectors orthogonal to  $u_1$  and  $u_2$ . For example

$$u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad u_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Finally,

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

Example: Construct the SVD of  $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$

Solution:

$$A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$$

So  $\sigma_1 = \sqrt{18}$  and  $\sigma_2 = 0$  and

$$\Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Constructing the right-singular vectors to form V:

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So

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

Next we construct the left-singular vectors.

$$u_1 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

By inspection, two other vectors orthogonal to  $u_1$  are

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

Because  $u_1 \in \text{Col } A$ , these vectors are in  $(\text{Col } A)^\perp$ . U is an orthogonal matrix so we use the Gram-Schmidt method to create an orthogonal basis for  $(\text{Col } A)^\perp$ :

$$u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad u_3 = \frac{1}{\sqrt{45}} \begin{bmatrix} -2 \\ 4 \\ -5 \end{bmatrix}$$



Finally,

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4\sqrt{45} \\ 2/3 & 0 & -5\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$