# MATH1013

THE AUSTRALIAN NATIONAL UNIVERSITY

## Mathematics and Applications I
## ALGEBRA section
## Linear Algebra and Complex Numbers
## course notes by Linda Stals

lecturer: Neil Montgomery

John Dedman Building, Room 2135A
Ph: 61252689
email (Mon - Thurs): neil.montgomery@anu.edu.au

Week 1, Semester 1, 2018

# Contents

14 Subspaces (Lay 2.8)

- The Subspace Concept
- The Fundamental Subspaces of a Matrix
- Basis, Dimension and Rank

15 Determinants (Lay 3.1)

- Determinants for Small Matrices
- Determinants for Large Matrices
- Cofactor Expansion

16 Properties of Determinants (Lay 3.2)
Cramer's Rule, Volume and Linear Transformations (Lay 3.3)

- Use of Row Operations
- Cramer's Rule
- Determinants as Area and Volumes

# References

## Text Book

D.C. Lay, *Linear Algebra and its Applications*, Addison-Wesley

Section references relate to the current (5th) edition. These also work for the 3rd (and 3rd-Update) and 4th edition, but are sometimes slightly out for the 2nd Edition. Any edition is fine and can also be used in MATH1014

## Notes

These Algebra lecture notes are available on the course website in full and reduced form (4 slides per page). Minor modifications may occur during the semester.

## Supplementary Reference

The Lay text has a short appendix on complex numbers, but it does not go quite far enough for our purposes. A copy of a small extract from another text (Adams) will be made available on the course website.

# Course website I

The principal website, powered by ‘Wattle’, is at

`http://wattle.anu.edu.au`

It is important to check this site regularly.

Examples of what you will find on the Wattle website include:

- General course information, including assessment details.
- Tutorial worksheets.
- Sample exams.
- Lecture notes and diaries.
- Discussion boards.
- Further reference material.

# Course website II

**The second website, powered by 'WebAssign' is at**

`https://www.webassign.net/login.html`

**It is important to keep up-to-date with the quizzes on this site.**

Assessable quizzes will be released at regular intervals.

More information about quizzes and other assessment items can be found on the Wattle site.

# Other Resources

- The Library!
  - Take one of the introductory tours offered by the library.
- The Internet
  - Lay student resources
- Practice quizzes (with answers) on WebAssign.

# Feedback and Individual Help

The following lists some of the resources available to you.

- Tutorials.
  - one-and-a-half hour tutorials in weeks 3-12.
  - [Tutors are generally not available outside tutorial times.]
- My scheduled office hours (see Wattle for details).
- Discussion boards available on Wattle.
- Organisation of self-help tutorial sessions by groups of students is an excellent idea.

  You can use the Wattle discussion board to help organise groups.
- ANU Counselling.
- Academic Skills and Learning Centre.
- See 'Important Course Information' on our Wattle site.

# Textbook Exercises

**Textbook**

PLEASE DO RELEVANT EXERCISES FROM THE TEXTBOOK!.

If you get stuck on a question you can post a message on the discussion board, ask your tutor or come and see me.

- The answers to the odd numbered questions are given in the back of the textbook.
- A student study guide to Lay is also available that shows detailed workings for some of the exercises.
- With the exception of one or two applications sections, we will essentially be working sequentially through the first three chapters of the textbook (plus the complex numbers appendix).

# This Class

For the lectures to run efficiently it is important that people are mindful of creating a good learning environment.

- All mobile phones should be turned off!
- Classes start at 5 minutes past the hour.
- Please try not to leave the room before the lecture is finished.
- The Discussion Board on Wattle is to be used for postings related to mathematics only.
- Please be mindful of the distraction of talking in class.

# What is Linear Algebra?

- Linear algebra is the area of mathematics that has arisen from the systematic study of methods for solving systems of linear equations.
- These systems may be small in terms of variables and/or unknowns, or they may be very large.
- Many of the structures that are transparent for small systems also apply in the larger ones, though the latter create further difficulties from a numerical point of view.
- Linear equations appear everywhere. We shall be motivated in much of our discussion by problems arising in areas as diverse as economic models, tomography, computer graphics and nutrition.

# Topics in Linear Algebra

- We will be covering most of the material in Chapters 1, 2 and 3 of Lay.
  - Basic methods of solving systems of linear equations.
  - Vector and matrix algebra.
  - Determinants.
  - Applications.

# Complex Numbers

- The complex number system $\mathbb{C}$ is an extension of the real number system $\mathbb{R}$ crafted to include square roots of negative real numbers, and in particular square roots of $-1$.
- In fact $\mathbb{C}$ contains square roots of all its members. Indeed it contains $n$-th roots of all its members for every natural number $n$.
- We will be covering all the material in Appendix A of Adams, which includes:
  - $i = \sqrt{-1}$ and the complex number $z = x + iy$.
  - De Moivre's Theorem and $z = r \operatorname{cis} \theta$
  - Finding $n$-th roots of real and complex numbers
  - Factorizing polynomials
  - Exponential notation and $e^{i\pi} = -1$

# 2. Systems of Equations (Lay 1.1)

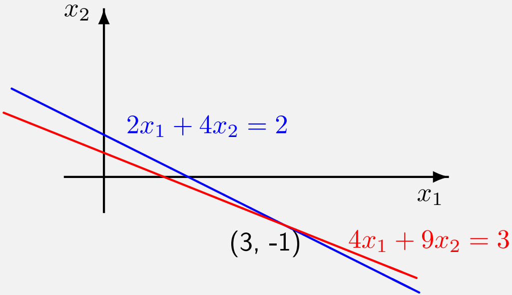
Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_updated_cw_3/cs_apps/lay03_01_cs.pdf
Lay's Review: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_1ov.pdf

# A Simple Problem

## Example

Solve the system of linear equations $\begin{aligned} 2x_1 + 4x_2 &= 2 \\ 4x_1 + 9x_2 &= 3 \end{aligned}$ .

Each equation represents a straight line in the $x_1$, $x_2$-plane.

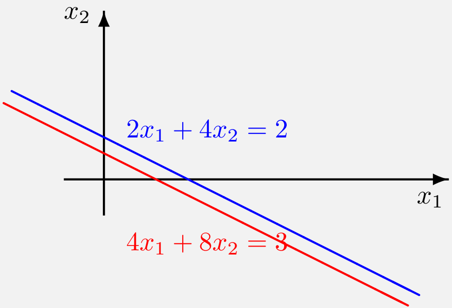



These lines intersect at the solution, namely $x_1 = 3$, $x_2 = -1$.
Check by substituting back into the equations.

# Slightly Different Problem

## Example

Solve the system of linear equations $\begin{array}{rcrcr} 2x_1 & + & 4x_2 & = & 2 \\ 4x_1 & + & 8x_2 & = & 3 \end{array}$.

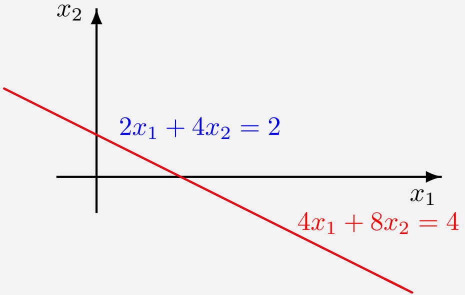



No solution!

# Another Slightly Different Problem

## Example

Solve the system of linear equations $\begin{array}{ccccc} 2x_1 & + & 4x_2 & = & 2 \\ 4x_1 & + & 8x_2 & = & 4 \end{array}$.




The two equations become the same: the two lines represented are the same, and any point on the line is a solution.

So there are infinitely many solutions.

# Solution Possibilities for Systems of Linear Equations

For larger systems of linear equations ('linear systems') the geometric aspects observed in the previous example problems remain, but their visualisation is more difficult.

**A System of linear equations has either:**

- No Solution, or
- Exactly one Solution, or
- Infinitely many solutions

**Definition**

We say a linear system is **consistent** if it has either one solution or infinitely many solutions.

**Definition**

A linear system is **inconsistent** if it has no solution.

# Interpolation

- Interpolation is fitting a function to a set of data points.
- Given: data $(t_i, y_i), \quad i = 1, \cdots, n$ and $t_i < t_j,$ for $i < j$.
- We seek: a function $P$ such that $P(t_i) = y_i, \quad i = 1, \cdots, n$.
- In particular, we often want to interpolate with polynomial functions.
- For $n$ points, a polynomial of degree $n - 1$ will do the trick.
- The next example demonstrates this.
- As a matter of fact, polynomials of high enough degree can 'approximately' interpolate whole functions, not just sets of points.
  The following theorem is for interest only.

**Theorem (Weierstrass Approximation Theorem)**

*If $f$ is defined and continuous on $[a, b]$ and $\epsilon > 0$ is given, then there exists a polynomial $P$, defined on $[a, b]$ such that*

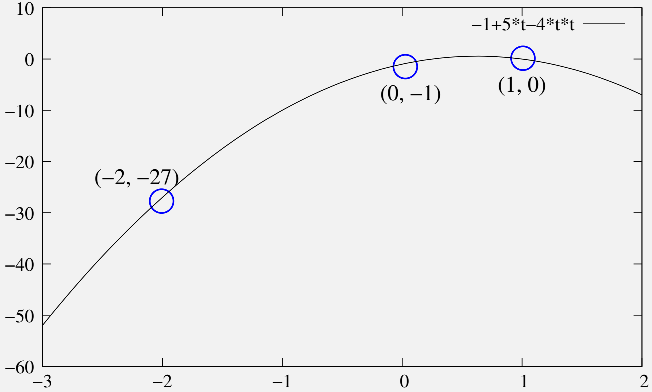
$$|f(t) - P(t)| < \epsilon \qquad \forall t \in [a, b].$$

## Example

Suppose we want to interpolate the data points (-2, -27), (0, -1), (1, 0). We try for an interpolating polynomial $P_2$ of degree 2:

$$P_2(t) = x_1 + x_2 t + x_3 t^2.$$

The problem is to find suitable values for the coefficients $x_1$, $x_2$ and $x_3$.



# Solution

To interpolate the point (-2,-27) we need $P_2(-2) = -27$ *i.e.*

$$x_1 + x_2(-2) + x_3(-2)^2 = -27$$

Together with the corresponding equations for the other two points, (0, -1) and (1, 0), we arrive at the system of linear equations

$$\begin{array}{ccccccc} x_1 & - & 2x_2 & + & 4x_3 & = & -27 \\ x_1 & + & 0x_2 & + & 0x_3 & = & -1 \\ x_1 & + & x_2 & + & x_3 & = & 0. \end{array}$$

Solving this system gives the coefficient values we need:

$$x_1 = -1,\ x_2 = 5,\ x_3 = -4.$$

Thus

$$P_2(t) = -1 + 5t - 4t^2.$$

# Coefficient Matrix

A system of $m$ linear equations in $n$ unknowns has the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where the $x_1, \ldots, x_n$ are unknowns, and the other terms are constants. The constants $a_{ij}$, for equations $i = 1 \ldots m$ and unknowns $x_j$, $j = 1 \ldots n$, are called the **coefficients** of the system.

## Definition

The array

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

is the **coefficient matrix** of the system.

# Augmented Matrix

Again for the linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

## Definition

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

*also written*

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

is the **augmented matrix** of the system.

The augmented matrix contains *all* the information of the original system of equations.

# Row Operations

There are three simple operations on a linear system that do not effect its solution:

- E1 Add a multiple of one equation to another equation.
- E2 Interchange two equations.
- E3 Multiply an equation by a non-zero constant.

The corresponding operations applied to the augmented coefficient matrix will similarly have no effect on the solution of the system.

**Definition (The elementary row operations)**

- R1 Replace one row by itself plus a multiple of another row.
- R2 Interchange two rows.
- R3 Multiply a row by a non-zero constant.

# Illustration of Elementary Operations I

## Example

Use row operations to simplify the system of linear equations

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$

$$\frac{3x}{2} + \frac{5y}{4} + 15z = 24. \tag{2}$$

First, for illustration purposes only, we will apply elementary operations to the equations directly.

After that we show how much simpler it looks when exactly the same steps are applied to the augmented matrix instead.

# Illustration of Elementary Operations II

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$

$$\frac{3x}{2} + \frac{5y}{4} + 15z = 24. \tag{2}$$

Our first step will be to eliminate $x$ from the second equation.
Multiply Equation (1) by -3, then add it to Equation (2):
(This is an example of an E1 operation.)
$\Rightarrow$

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$

$$\frac{1}{2}y + 12z = 12. \tag{3}$$

( The new equation (3) does not involve $x$. )

# Illustration of Elementary Operations III

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$
$$\frac{1}{2}y + 12z = 12. \tag{3}$$

Next we choose to simplify the coefficient of $y$ in the second equation.
Multiply Equation (3) by 2:
(This is an example of an E3 operation.)
$\Rightarrow$

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$
$$y + 24z = 24. \tag{4}$$

( The coefficient of $y$ in the new equation (4) is 1. )

# Illustration of Elementary Operations IV

$$\frac{x}{2} + \frac{y}{4} + z = 4, \tag{1}$$
$$y + 24z = 24. \tag{4}$$

Next we should eliminate $y$ from the first equation.
Multiply Equation (4) by -1/4, then add it to Equation (1):
(This is another example of an E1 operation.)

$$\frac{x}{2} - 5z = -2. \tag{5}$$
$$y + 24z = 24. \tag{4}$$

( The new equation (5) does not involve $y$. )

# Illustration of Elementary Operations V

$$\frac{x}{2} - 5z = -2. \tag{5}$$

$$y + 24z = 24. \tag{4}$$

Finally we will simplify the coefficient of $x$ in the first equation.
Multiply Equation (5) by 2:
(This is another example of an E3 operation.)
$\Rightarrow$

$$x - 10z = -4. \tag{6}$$

$$y + 24z = 24. \tag{4}$$

( The coefficient of $x$ in the new equation (6) is 1. )

No further simplification is possible.
The system has an infinite set of solutions.
We will discuss this phenomenon later.

# Illustration of Elementary Operations VI

Now we show exactly the same sequence of operations applied to the augmented matrix of the system:

$$
\begin{aligned}
\left[\begin{array}{rrrr} 1/2 & 1/4 & 1 & 4 \\ 3/2 & 5/4 & 15 & 24 \end{array}\right]
&\overset{\text{R2-3R1}}{\longrightarrow} \left[\begin{array}{rrrr} 1/2 & 1/4 & 1 & 4 \\ 0 & 1/2 & 12 & 12 \end{array}\right] \\
&\overset{\text{2R2}}{\longrightarrow} \left[\begin{array}{rrrr} 1/2 & 1/4 & 1 & 4 \\ 0 & 1 & 24 & 24 \end{array}\right] \\
&\overset{\text{R1-1/4R2}}{\longrightarrow} \left[\begin{array}{rrrr} 1/2 & 0 & -5 & -2 \\ 0 & 1 & 24 & 24 \end{array}\right] \\
&\overset{\text{2R1}}{\longrightarrow} \left[\begin{array}{rrrr} 1 & 0 & -10 & -4 \\ 0 & 1 & 24 & 24 \end{array}\right].
\end{aligned}
$$

As expected this represents exactly the same pair of simplified equations:

$$
\begin{aligned}
x - 10z &= -4 \\
y + 24z &= 24.
\end{aligned}
$$

# Second Illustration of Row Operations I

## Example

Use row operations to simplify the system

$$\begin{array}{ccccccc} 2x & + & 2y & - & 4z & = & 0 \\ 2x & + & 2y & - & z & = & 1 \\ 3x & + & 2y & - & 3z & = & 3. \end{array}$$

The system has augmented matrix $\left[\begin{array}{cccc} 2 & 2 & -4 & 0 \\ 2 & 2 & -1 & 1 \\ 3 & 2 & -3 & 3 \end{array}\right]$.

On the next slide we show a sequence or row operations that leads to a unique solution of the system.

The sequence of operations was chosen in this case to exploit particular features of the matrices.

In the next section we will examine a systematic sequence, called Guass-Jordan Elimination, that can be applied to *any* matrix.

# Second Illustration of Row Operations II

$$\begin{bmatrix} 2 & 2 & -4 & 0 \\ 2 & 2 & -1 & 1 \\ 3 & 2 & -3 & 3 \end{bmatrix} \xrightarrow{\text{R1R3; R2R3}} \begin{bmatrix} 3 & 2 & -3 & 3 \\ 2 & 2 & -4 & 0 \\ 2 & 2 & -1 & 1 \end{bmatrix} \xrightarrow{\text{R3-R2}}$$

$$\begin{bmatrix} 3 & 2 & -3 & 3 \\ 2 & 2 & -4 & 0 \\ 0 & 0 & 3 & 1 \end{bmatrix} \xrightarrow{\text{R2-2/3R1; 1/3R3}} \begin{bmatrix} 3 & 2 & -3 & 3 \\ 0 & 2/3 & -2 & -2 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \xrightarrow{\text{R2+2R3}}$$

$$\begin{bmatrix} 3 & 2 & -3 & 3 \\ 0 & 2/3 & 0 & -4/3 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \xrightarrow{\text{3/2R2}} \begin{bmatrix} 3 & 2 & -3 & 3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \xrightarrow{\text{R1-2R2}}$$

$$\begin{bmatrix} 3 & 0 & -3 & 7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \xrightarrow{\text{R1+3R3}} \begin{bmatrix} 3 & 0 & 0 & 8 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \xrightarrow{\text{1/3R}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 8/3 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1/3 \end{bmatrix} \rightarrow \begin{array}{rcl} x & = & 8/3 \\ y & = & -2 \\ z & = & 1/3 \end{array}$$

# Last Illustration of Row Operations

## Example

Use row operations to simplify the system of linear equations

$$\begin{array}{ccccccc} x & & & + & z & = & 1 \\ & & y & - & z & = & -1 \\ 2x & + & y & + & z & = & 2. \end{array}$$

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & -1 \\ 2 & 1 & 1 & 2 \end{bmatrix} \overset{\text{R2R3}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 0 & 1 & -1 & -1 \end{bmatrix} \overset{\text{R2-2R1}}{\longrightarrow}$$

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix} \overset{\text{R3-R2}}{\longrightarrow} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

The last row here represents $\quad 0x + 0y + 0z = -1$

which has no solution. So the whole system has no solution.

# 3. Reduced Row Echelon Form (Lay 1.2)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_updated_cw_3/cs_apps/network.pdf
Lay's Review: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_2ov.pdf

# Leading Entries

We have seen that solving linear systems corresponds to using elementary row operations to convert the system's augmented matrix to a form like one of these:

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & * \\ 0 & \mathbf{1} & 0 & * \\ 0 & 0 & \mathbf{1} & * \end{bmatrix} \qquad \begin{bmatrix} \mathbf{1} & 0 & 1 & * \\ 0 & \mathbf{1} & -1 & * \\ 0 & 0 & 0 & * \end{bmatrix}$$

## Definition

The first (=leftmost) non-zero entry in a non-zero row of a matrix is the **leading entry** or **pivot** of that row. See the bold entries above.

# Reduced Row Echelon Form

## Definition

A matrix is in **Reduced Row Echelon** form if:

1. Any zero rows are at the bottom of the matrix.
2. In any non-zero row the leading entry is to the right of the leading entry of any higher row.
3. Each column that contains a leading entry has zeros everywhere below it. (This is actually a consequence of 2.)
4. In any non-zero row, the leading entry is a 1.
5. Each *c*olumn that contains a leading 1 has zeros everywhere else.

When an augmented matrix has been converted to reduced row echelon form, the variables corresponding to the leading 1's are called **leading variables**, the others **free variables**.

## Definition

A matrix is in **Row Echelon** form if properties 1-3 are satisfied.

# Reduced Row Echelon Example

## Illustration

This matrix is in reduced row echelon form:

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

It corresponds to the system of equations

$$\begin{array}{ccccc} x_1 & & & = & 7 \\ & x_2 & & = & -2 \\ & & x_4 & = & 5 \end{array},$$

The solution to this system is

$$x_1 = 7, \quad x_2 = -2, \quad x_3 = t, \quad x_4 = 5,$$

where $t$ is arbitrary. (We sometimes call $t$ a '*parameter*'.)

The leading variables $x_1, x_2, x_4$ are completely determined.
The free variable $x_3$ is completely unconstrained.

# Are the following in Reduced Row Echelon Form?

## Example

$$\begin{bmatrix} 1 & 0 & 0 & 2 & 3 \\ 0 & 2 & 0 & 1 & -2 \\ 0 & 0 & 1 & 1 & 6 \end{bmatrix},$$

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

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

# An Example with Two Free Variables

## Illustration

$$\begin{bmatrix} 1 & 2 & 0 & 2 & 0 & 5 \\ 0 & 0 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

is in reduced row echelon form.

The corresponding equations are

$$\begin{array}{ccccccccc} x_1 & + & 2x_2 & & & + & 2x_4 & & & = & 5 \\ & & & & x_3 & & & & & = & 3 \\ & & & & & & & & x_5 & = & 4 \end{array}.$$

The leading variables are $x_1, x_3, x_5$. The free variables are $x_2, x_4$.

Setting $x_2 = s, x_4 = t$, the solution in terms of the parameters $s$ and $t$ is

$$x_1 = 5 - 2s - 2t, \quad x_2 = s, \quad x_3 = 3, \quad x_4 = t, \quad x_5 = 4.$$

# Algorithm for rref (Reduced Row Echelon Form)

## Gauss-Jordan Elimination

1. Find the leftmost non-zero *column.*
2. Interchange the top row with another, if necessary, to bring a non-zero entry to the top of the column determined in Step 1.
3. Add suitable multiples of the top row to the rows below so that all entries below the pivot are zero.
4. Now cover the top row and repeat from Step 1 until matrix is in row-echelon form.
5. Multiply the last non-zero row by a suitable constant to make the pivot value 1.
6. Add suitable multiples of the last non-zero row to the rows above to produce zero entries above the pivot.
7. Now cover the last non-zero row and repeat from step 5 until the matrix is in reduced row echelon form.

# Example of Row Reduction

## Example

Solve the system

$$\begin{array}{ccccccc} & & y & + & z & = & 0 \\ x & & & + & z & = & 2 \\ 2x & + & 2y & + & 4z & = & 2 \end{array},$$

using the algorithm for reduction to reduced echelon form.

First, the augmented matrix for the system is

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

# Eliminate Variables from First Column of $\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 2 \\ 2 & 2 & 4 & 2 \end{bmatrix}$

1. Leftmost nonzero column is the first column.
2. Interchange first and last rows: (could have interchanged first and second)

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

3. Subtract half Row 1 from Row 2, so all entries below pivot are zero,

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

4. Cover Row 1 and repeat from Step 1.

# Eliminate Variables from Second Column of $\begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$

1. Leftmost nonzero column is the second column, which is nonzero in current top entry.
2. Add Row 2 to Row 3:
$$\begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We now have row echelon form.

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

1. $$\text{Subtract Row 3 from Row 2}: \quad \begin{bmatrix} 2 & 2 & 4 & 2 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

2. $$\text{Subtract twice Row 3 from Row 1}: \quad \begin{bmatrix} 2 & 2 & 4 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

3. $$\text{Add twice Row 2 to Row 1}: \quad \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

4. $$\text{Now scale}: \quad \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

# Interpreting $\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

The corresponding equations are

$$\begin{array}{ccccccc} x & & & + & z & = & 0 \\ & & y & + & z & = & 0 \\ & & & & 0 & = & 1 \end{array}$$

The last equation shows inconsistency.

## Definition

As indicated earlier, this method of obtaining the reduced row echelon form to solve a system of linear equations is called **Gauss-Jordan elimination**.

# Gaussian Elimination

## Definition

If we stop when a row echelon form has been reached (as opposed to a *reduced* row echelon form) it is called **Gaussian elimination**, and we then use **back substitution** to solve the system as follows:

1. Solve the equations for the leading variables.
2. Starting with the last equation, substitute each equation into the one above it.
3. Treat the free variables, if any, as unconstrained parameters.

# Gaussian Elimination Example

## Example

Solve the system

$$\begin{array}{rcrcrcrcr} & & 10y & - & 4z & + & w & = & 1 \\ x & + & 4y & - & z & + & w & = & 2 \\ 3x & + & 2y & + & z & + & 2w & = & 5 \\ -2x & - & 8y & + & 2z & - & 2w & = & -4 \\ x & - & 6y & + & 3z & & & = & 1, \end{array}$$

using Gaussian Elimination.

The system has augmented matrix

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

# Reduction to Row Echelon Form

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

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

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

(all-zero rows can be ignored)

# Back Substitution applied to $\left[\begin{array}{ccccc} 1 & 4 & -1 & 1 & 2 \\ 0 & 10 & -4 & 1 & 1 \end{array}\right]$

The row echelon form gives the corresponding equations

$$\begin{array}{ccccccccc} x & + & 4y & - & z & + & w & = & 2 \\ & & 10y & - & 4z & + & w & = & 1. \end{array}$$

Solving for the leading variables gives:

$$\begin{array}{rcl} x & = & -4y + z - w + 2 \\ y & = & \frac{2}{5}z - \frac{1}{10}w + \frac{1}{10}. \end{array}$$

Back substituting now gives:

$$\begin{array}{rcl} x & = & -\frac{3}{5}z - \frac{3}{5}w + \frac{8}{5} \\ y & = & \frac{2}{5}z - \frac{1}{10}w + \frac{1}{10} \end{array}$$

with $z, w$ free variables.

# Alternate Jordan Reduction of $\begin{bmatrix} 1 & 4 & -1 & 1 & 2 \\ 0 & 10 & -4 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Instead of back substitution we could have continued with Gauss-Jordan reduction to convert the above row-echelon-form matrix to *reduced* row echelon form. This only requires the steps: $R_2 \leftarrow (R_2/10)$,
$R_1 \leftarrow (R_1 - 4R_2)$.

The result is

$$\begin{bmatrix} 1 & 0 & 3/5 & 3/5 & 8/5 \\ 0 & 1 & -2/5 & 1/10 & 1/10 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This leads to the same equations as did back substitution.

# Consistent and Inconsistent Matrices

## Theorem (c.f. Theorem 2 in §1.2 of Lay)

*(1) If a row is obtained of the form*

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & \alpha \end{bmatrix}$$

*where $\alpha \neq 0$, (there is a pivot in the final column), then the system is* **inconsistent** *and so has no solution. Otherwise the system is* **consistent**.

*(2) Let us suppose that the system is consistent. Then*

- *if there is a pivot element in every variable column (equivalently there are no free variables), the system has a unique solution,*
- *if there are some free variables, the system has infinitely many solutions.*

## Example

The reduced row echelon form for the system

$$\begin{array}{ccccccccccc} 3a & + & 4b & + & 5c & - & d & + & e & = & p \\ 4a & + & 2b & - & c & + & 7d & - & e & = & q \\ 2a & - & 3b & + & 4c & + & d & + & 3e & = & r \\ 5a & + & 9b & & & + & 5d & - & 3e & = & s. \end{array}$$

is

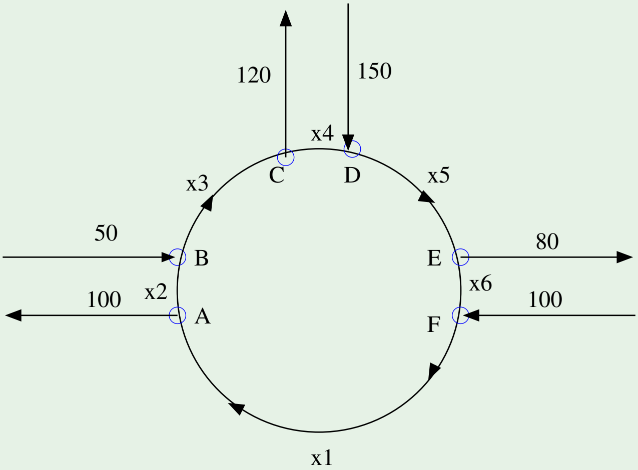
$$\begin{bmatrix} 1 & 0 & 0 & \frac{236}{137} & 6/137 & -36/137p + 31/137s + 45/137r \\ 0 & 1 & 0 & \frac{-55}{137} & -49/137 & -2/137s + 20/137p - 25/137r \\ 0 & 0 & 1 & \frac{-125}{137} & 63/137 & -7/137r + 33/137p - 17/137s \\ 0 & 0 & 0 & 0 & 0 & q + p - s - r \end{bmatrix}.$$

Under what conditions does the system have a solution?

# Traffic Flow Example

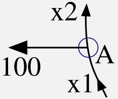
## Example

Traffic flow at a roundabout is as shown.




What is the smallest possible value of $x_6$?

# Solving the Traffic Flow Example I




- At each node, equate 'inputs' to 'outputs'.
- E.g. at node A input ($x_1$) must equal output ($x_2$+100).
- So node A provides the equation $\quad x_1 - x_2 = 100.$
- Create similar equations for the other five nodes.
- The augmented matrix for the resulting $6 \times 6$ linear system is shown below, together with its reduced row echelon form.

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 100 \\ 0 & 1 & -1 & 0 & 0 & 0 & -50 \\ 0 & 0 & 1 & -1 & 0 & 0 & 120 \\ 0 & 0 & 0 & 1 & -1 & 0 & -150 \\ 0 & 0 & 0 & 0 & 1 & -1 & 80 \\ -1 & 0 & 0 & 0 & 0 & 1 & -100 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & 100 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 50 \\ 0 & 0 & 0 & 1 & 0 & -1 & -70 \\ 0 & 0 & 0 & 0 & 1 & -1 & 80 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

# Solving the Traffic Flow Example II

The solution is:

$$\begin{aligned} x_1 &= x_6 + 100, \\ x_2 &= x_6, \\ x_3 &= x_6 + 50, \\ x_4 &= x_6 - 70, \\ x_5 &= x_6 + 80. \end{aligned}$$

Question: So what is the smallest possible value of $x_6$?

# 4. Vector Equations (Lay 1.3)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_updated_cw_3/cs_apps/diet.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_3ov.pdf

# Vectors

So far we have shown how to solve a system such as

$$\left.\begin{array}{rcl} 2x_1 + 4x_2 & = & 2 \\ 4x_1 + 9x_2 & = & 3 \end{array}\right\} \ldots (*)$$

by applying row operations to the augmented coefficient matrix $\mathbf{M}$.

The individual rows of $\mathbf{M}$, $[2\ 4\ 2]$ and $[4\ 9\ 3]$, are called *row matrices*.

The individual cols, $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$, $\begin{bmatrix} 4 \\ 9 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ are called *column matrices*.

Both row and column matrices are often called 'vectors'. But for us:

## Definition

A **vector** is a *column* matrix.

We shall see later that $(*)$ can be written as the single *vector equation*

$$x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

# Dimension

In this course we will mostly be concerned with vectors like

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

with only two or three entries (or 'components').

However a vector can have any number, $n$, of entries:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

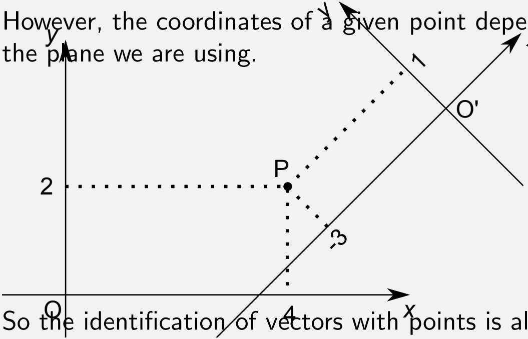
We call $n$ the **dimension** of the vector.

The notation $\mathbb{R}^2$, $\mathbb{R}^3$, $\mathbb{R}^n$ will be used for the sets of all 2-, 3- and $n$-dimensional vectors respectively. ($\mathbb{R}$ is the set of all real numbers; the set to which the entries in the vectors belong.)

# Graphical Interpretation of Vectors (i): **Points**

The set $\mathbb{R}^2$ is called **2-space** because we can identify $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ with the point with coordinates $(u_1, u_2)$ in the plane.

However, the coordinates of a given point depend on what coordinate system in the plane we are using.




Coordinates of P are

- (4,2) in one coordinate system
- (-3,1) in the other.

So the identification of vectors with points is always *relative* to some given or imagined fixed coordinate system.

# Graphical Interpretation of Vectors (ii): **Translations**

In practice we usually have in mind fixed *directions* for the coordinate axes. For example:

- 'east' or 'right' for the (positive) $x$-axis
- 'north' or 'up' for the $y$-axis

However we may not have in mind any particular position for the origin.

In this case we can think of a vector $\begin{bmatrix} a \\ b \end{bmatrix}$ as representing a **translation** (*i.e.* a 'shift') of

- $a$ units in the $x$ direction, and
- $b$ units in the $y$ direction, but
- independent of the position of the origin.

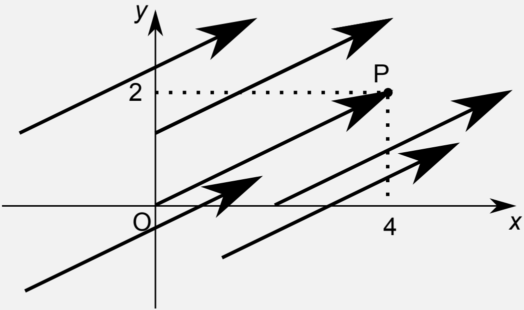
# Representation of Vectors as Arrows

Viewed as translations, vectors are represented by *arrows*.

An arrow has

| | |
|---|---|
| length; | showing the size of the translation, and |
| direction; | showing the direction of the translation, but |
| no position; | it can be drawn anywhere. |

For example, in the diagram **each** of the arrows represents the 'translation vector' $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$.




All the arrows

- have exactly the same length
- point in exactly the same direction

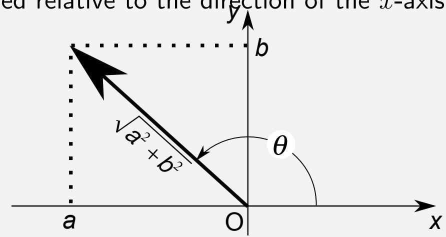
# Physical Vectors

In the physical sciences, a 'vector' is defined to be an entity that has magnitude and direction (but no location). The interpretation of a two-entry column matrix as a translation vector in the plane is in accord with this definition:

The magnitude of $\begin{bmatrix} a \\ b \end{bmatrix}$ is the length of any representing arrow, *i.e.* $\sqrt{a^2 + b^2}$ by Pythagoras.

The direction of $\begin{bmatrix} a \\ b \end{bmatrix}$ is specified relative to the direction of the $x$-axis.

The direction is measured by the angle $\theta$ that the representing arrow starting from the origin makes to the $x$-axis, measured counter-clockwise.



# 3-Space and $n$-Space

Vectors of dimension 3 can be interpreted as translations in 3D.

For example if

$$\left\{\begin{array}{l}\text{the } x\text{-axis points East}\\ \text{the } y\text{-axis points North}\\ \text{the } z\text{-axis points Up}\end{array}\right\} \text{ then } \begin{bmatrix}-2\\10\\-5\end{bmatrix} \text{ represents } \left\{\begin{array}{l}\text{2 units West}\\ \text{10 units North}\\ \text{5 units Down}\end{array}\right\}.$$

As in the 2D case, the position of the origin is irrelevant.

Remarks:

- interpretation in accord with usage in the physical sciences
- difficult to show 3D arrows convincingly on 2D paper/screen
- specifying direction by angles gets awkward
- for $n > 3$ $n$D vectors live in 'hyperspace' — difficult to visualize

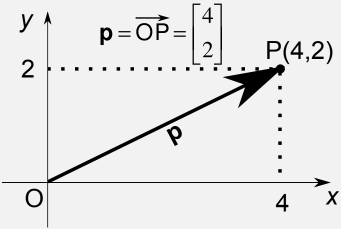
# Position Vectors

## Definition

The **position vector** of a point P relative to the origin O of a given coordinate system is the translation vector that represents, in that coordinate system, the shift from O to P.

The position vector of P relative to O is often denoted by $\mathbf{p} = \overrightarrow{OP}$.

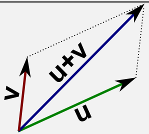
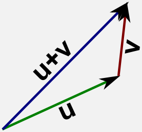
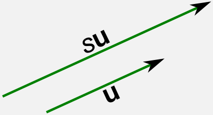
When the coordinate system is understood from the context, the phrase "relative to O" is often omitted. However it must always be borne in mind, because position vectors don't make sense without this.




- P$(a, b)$ has position vector $\mathbf{p} = \begin{bmatrix} a \\ b \end{bmatrix}$ (relative to O)
- $\mathbf{p}$ is *not the same* as $(a, b)$
- $\mathbf{p}$ *itself* is not 'tied' to O

# Addition and Scalar Multiplication of Vectors

The following is for $\mathbb{R}^2$, with the obvious extension to $\mathbb{R}^3$ and beyond.

| | Algebraic Definitions | Graphical Interpretation |
|---|---|---|
| Sum: | $\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}$ (component-wise addition) |   (parallelogram rule)   (triangle rule) |
| Scalar Product: | $s \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} sa_1 \\ sa_2 \end{bmatrix}$ <br> $s \in \mathbb{R}$ (*i.e.* $s$ is a 'scalar') |   (shrink/stretch rule) |

# A Linear System Revisited

Earlier we claimed that the linear system

$$\left.\begin{array}{rcl} 2x_1 + 4x_2 & = & 2 \\ 4x_1 + 9x_2 & = & 3 \end{array}\right\} \ldots (*)$$

could be written as the vector equation

$$x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \ldots (**).$$

Having now defined addition and scalar multiplication for vectors we can verify this claim, because $(**)$ is equivalent to

$$\begin{bmatrix} 2x_1 + 4x_2 \\ 4x_1 + 9x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

and this is true if and only if $(*)$ is true.

# Unit Vectors

## Definition

Any vector with magnitude 1 is called a **unit vector**.
Unit vectors in the directions of coordinate axes are called **standard** unit vectors and are denoted by $\mathbf{i}$, $\mathbf{j}$ etc. ($\mathbf{e}_1$, $\mathbf{e}_2$ etc. in Lay.)

In $\mathbb{R}^2$: $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ In $\mathbb{R}^3$: $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ $\mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Using the operations of addition and scalar multiplication we can now write, for any $a, b, c \in \mathbb{R}$,

$$\begin{aligned}
\begin{bmatrix} a \\ b \end{bmatrix} &= a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix} = a\mathbf{i} + b\mathbf{j} \\
\begin{bmatrix} a \\ b \\ c \end{bmatrix} &= a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}
\end{aligned}$$

This way of writing vectors is popular in the physical sciences.

# Linear Combinations

## Definition

Given vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ in $\mathbb{R}^n$ , and scalars $c_1, c_2..., c_p$, the vector

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \sum_{i=1}^{p} c_i\mathbf{v}_i$$

is a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ with **weights** $c_1, c_2..., c_p$.

## Example

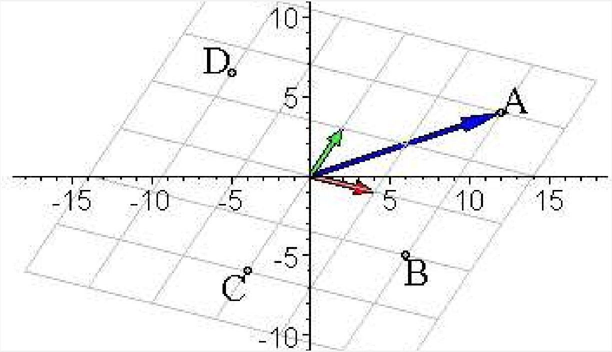
$$\mathbf{w} = \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 4\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

is a linear combination of $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$ with weights 2 and 4.

# Graphical Interpretation of Linear Combinations I

## Example

Points A, B, C, D have position vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$.
What weights are required to express these position vectors as linear combinations of the vectors $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ (green) and $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ (red)?
Hint: Use stretch and parallelogram rules.



# Graphical Interpretation of Linear Combinations II

## Answers

| Point | Weights $c_1$ | $c_2$ | Linear Combination |
|---|---|---|---|
| A | 2 | 2 | $\mathbf{a} = 2\mathbf{v}_1 + 2\mathbf{v}_2$ |
| B | -1 | 2 | $\mathbf{b} = -\mathbf{v}_1 + 2\mathbf{v}_2$ |
| C | -2 | 0 | $\mathbf{c} = -2\mathbf{v}_1$ |
| D | 1.5 | -2 | $\mathbf{d} = 1.5\mathbf{v}_1 - 2\mathbf{v}_2$ |

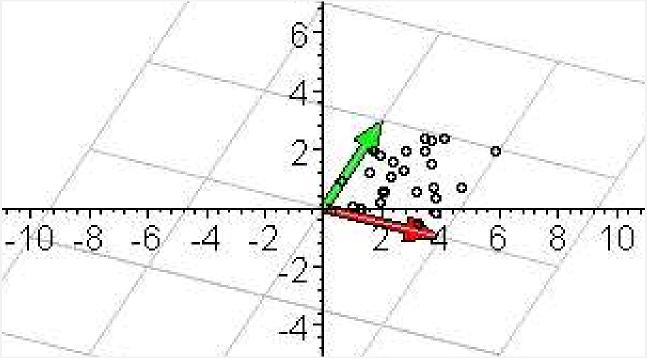
Comments:

- It appears that the position vector of **any** point in the plane can be expressed as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$.
- The next few slides aim to show this is true.
- The set of *all* linear combinations of $\mathbf{v}_1, \mathbf{v}_2$ is called the **span** of $\mathbf{v}_1, \mathbf{v}_2$. (Formal definition later)

# Graphical Interpretation of Linear Combinations III

## Illustration

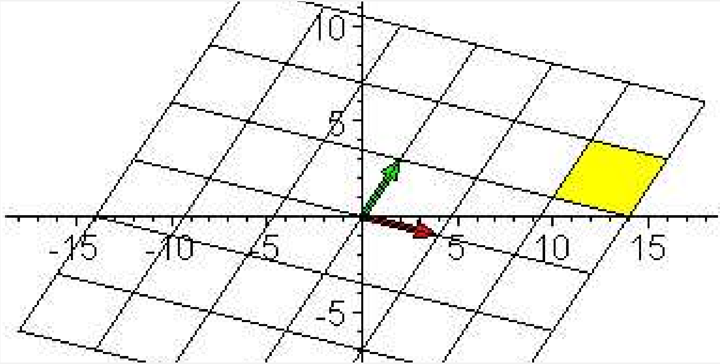
All the dot points have position vectors in the span of $\mathbf{v}_1, \mathbf{v}_2$.
For each point the weights are between 0 and 1 inclusive.



# Graphical Interpretation of Linear Combinations IV

## Illustration

What would be the ranges of weights $c_1$ and $c_2$ for the shaded region in the span of $\mathbf{v}_1$ (green), $\mathbf{v}_2$ (red)?



# Graphical Interpretation of Linear Combinations V

## Example

For the vectors $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$ of this running example **prove** that the span of $\mathbf{v}_1, \mathbf{v}_2$ is $\mathbb{R}^2$. (We say "$\mathbf{v}_1, \mathbf{v}_2$ spans $\mathbb{R}^2$".)

We need to establish that for any vector $\begin{bmatrix} a \\ b \end{bmatrix}$ in $\mathbb{R}^2$ we can find weights $c_1, c_2$ so that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_1 \begin{bmatrix} 2 \\ 3 \end{bmatrix} + c_2 \begin{bmatrix} 4 \\ -1 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Thus we need to show that the system with augmented matrix

$$\mathbf{M} = \begin{bmatrix} 2 & 4 & a \\ 3 & -1 & b \end{bmatrix}$$

has a solution (is consistent) for all values of $a, b \in \mathbb{R}$.

# Graphical Interpretation of Linear Combinations V(cont.)

Using Gauss-Jordan elimination we find that $\mathbf{M}$ has the following reduced row echelon form

$$\left[\begin{array}{ccc} 1 & 0 & (a+4b)/14 \\ 0 & 1 & (3a-2b)/14 \end{array}\right].$$

Thus given any $a, b \in \mathbb{R}$ whatsoever, we simply set

$$\begin{array}{rcl} c_1 & = & (a+4b)/14 \\ c_2 & = & (3a-2b)/14 \end{array}$$

and we are assured that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \left[\begin{array}{c} a \\ b \end{array}\right].$$

So every vector $\left[\begin{array}{c} a \\ b \end{array}\right]$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, which is what we wanted to prove.

# Definition and Interpretation of Span

## Definition

Given a set of $p$ vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ in $\mathbb{R}^n$, the set of all possible linear combinations of these vectors is called the set **spanned** by $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$, and is denoted by

$$\text{Span}\left\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\right\}.$$

## Graphical Interpretation

If every vector in $S = \text{Span}\left\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p\right\}$ is interpreted as the position vector of a point in $n$-space, then $S$ represents all the points that make up a 'linear object' containing the origin.

Examples of 'linear objects':

In $\mathbb{R}^2$: points, lines, the whole of the 2-space plane.

In $\mathbb{R}^3$: points, lines, planes, the whole of 3-space itself.

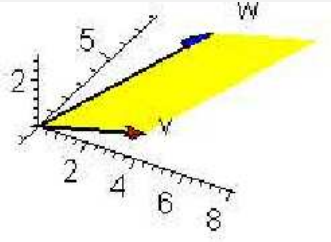
In $\mathbb{R}^n,\ n > 3$: hyperpoints, hyperlines, hyperplanes etc.

# An Example in $\mathbb{R}^3$

## Example

For $\mathbf{v} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 3 \\ 7 \\ 1 \end{bmatrix}$, what is the graphical interpretation of Span$\{\mathbf{v}, \mathbf{w}\}$?

The Span of $\mathbf{v}$ and $\mathbf{w}$ is a plane in $\mathbb{R}^3$. Below is a plot of all the points generated using weights between 0 and 1 inclusive.



# Another Example in $\mathbb{R}^3$.

## Example

Let
$$\mathbf{u} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -5 \\ -8 \\ 2 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 3 \\ -5 \\ h \end{bmatrix}.$$
For which values of $h$ is $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$?

To answer this, we have to check the consistency of the system
$$x_1 \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} -5 \\ -8 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -5 \\ h \end{bmatrix}.$$

Looking at each row in turn, we have the linear system
$$\begin{array}{rcrcr} x_1 & - & 5x_2 & = & 3 \\ 3x_1 & - & 8x_2 & = & -5 \\ -x_1 & + & 2x_2 & = & h \end{array}.$$

# Another Example in $\mathbb{R}^3$ (cont.)

This system in turn has augmented matrix and reduced row echelon matrix given by

$$\begin{bmatrix} 1 & -5 & 3 \\ 3 & -8 & -5 \\ -1 & 2 & h \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -5 & 3 \\ 0 & 7 & -14 \\ 0 & -3 & h+3 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & -5 & 3 \\ 0 & 1 & -2 \\ 0 & -3 & h+3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & -2 \\ 0 & 0 & h-3 \end{bmatrix}.$$

From this we see that the equation is consistent if and only if $h = 3$.
If $h \neq 3$ then $\mathbf{w} \notin \text{Span}\,\{\mathbf{u}, \mathbf{v}\}$.

# Extra Examples

Previous examples suggest a systematic way to answer the following:

## Example

What is the span of the following vectors?

(1)

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(2)

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

(3)

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}.$$

Brief solutions follow.

# Extra Example 1

## Example (1)

Find Span$\{\mathbf{v}_1, \mathbf{v}_2\}$; $\quad \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

We need to check the consistency of the system

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \begin{bmatrix} a \\ b \end{bmatrix}.$$

The augmented matrix and its reduced row echelon form are

$$\begin{bmatrix} 2 & 1 & a \\ 1 & 1 & b \end{bmatrix} \to \begin{bmatrix} 1 & 1/2 & a/2 \\ 0 & 1 & 2b-a \end{bmatrix} \to \begin{bmatrix} 1 & 0 & a-b \\ 0 & 1 & 2b-a \end{bmatrix}.$$

So the system is consistent and hence Span $\{\mathbf{v}_1, \mathbf{v}_2\} = \mathbb{R}^2$.

# Extra Example 2

## Example (2)

Find Span$\{\mathbf{v}_1, \mathbf{v}_2\}$; $\quad \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.

We need to check the consistency of the system

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \begin{bmatrix} a \\ b \end{bmatrix}.$$

The augmented matrix and its reduced row echelon form are

$$\begin{bmatrix} 2 & 4 & a \\ 1 & 2 & b \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 4 & a \\ 0 & 0 & b - a/2 \end{bmatrix}$$

The system is consistent if and only if $b = a/2$. Hence Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ consists of all vectors of the form $\begin{bmatrix} 2b \\ b \end{bmatrix}$, $b \in \mathbb{R}$.

# Extra Example 3

## Example (3)

Find Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$; $\quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}.$

We need to check the consistency of the system

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \begin{bmatrix} a \\ b \\ d \end{bmatrix}.$$

The augmented matrix and its row echelon form are

$$\begin{bmatrix} 1 & 1 & 2 & a \\ 1 & 2 & 3 & b \\ 1 & 1 & 2 & d \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 2 & a \\ 0 & 1 & 1 & b-a \\ 0 & 0 & 0 & d-a \end{bmatrix}.$$

The system is consistent if and only if $d = a$. Hence Span $\{\mathbf{v}_1, \mathbf{v}_2.\mathbf{v}_3\}$ consists of all vectors of the form $\begin{bmatrix} a \\ b \\ a \end{bmatrix}$, $b \in \mathbb{R}$.

# 5. Matrix Equation $\mathbf{Ax} = \mathbf{b}$ (Lay 1.4)

Case Study:

Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_4ov.pdf

# Matrix Vector Multiplication $\mathbf{A}\mathbf{x}$

## Definition

Let $\mathbf{A}$ be an $m \times n$ matrix with column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$.
Let $\mathbf{x} \in R^n$.

$$\mathbf{A}\mathbf{x} = \overset{\overset{\mathbf{A}}{\downarrow}}{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}} \overset{\overset{\mathbf{x}}{\downarrow}}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \sum_{j=1}^{n} x_j\mathbf{a}_j.$$

So $\mathbf{A}\mathbf{x}$ is the linear combination of the columns of $\mathbf{A}$ with weights given by the entries of $\mathbf{x}$.

## Illustration

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 6 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 4x_1 + 5x_2 + 6x_3. \end{bmatrix}$$

# Calculation of individual components of $\mathbf{Ax}$

Each component (single entry row) of $\mathbf{Ax}$ can be obtained directly — *i.e.* without first forming the linear combination of the columns of $\mathbf{A}$.

With $\mathbf{A}$, $\mathbf{x}$ as before, the $i$-th component of $\mathbf{Ax}$, written $(\mathbf{Ax})_i$, is given by

$$(\mathbf{Ax})_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j.$$

This is called the **inner product** of the *row* $(\mathbf{A})_i = [a_{i1}, a_{i2}, \ldots, a_{in}]$ and the *column* $\mathbf{x}$. It is the **sum of products** of corresponding entries.

## Illustration

The 4th component of $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \\ 5 & 6 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is $4x_1 + 5x_2 + 6x_3.$

# Dimensions

Using inner products we see that

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{n \times 1} = \underbrace{\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}}_{m \times 1}.$$

So:

- An $m \times n$ matrix multiplied by an $n \times 1$ matrix gives an $m \times 1$ matrix.
- The two 'inner' (indigo) dimensions are the same (they are both $n$).
- The product is not defined unless this is the case.
- The two 'outer' (orange) dimensions, $m$ and 1, give the dimensions of the product.

# Linearity

## Theorem

*The following 'linearity' rules hold for matrix vector multiplication:*

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

*where* $\mathbf{A}$ *is an* $m \times n$ *matrix,* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*.*

## Example

Verify that $\mathbf{A}(\mathbf{u}+\mathbf{v}) = \mathbf{A}\mathbf{u}+\mathbf{A}\mathbf{v}$ when $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}$.

$$\mathbf{A}(\mathbf{u}+\mathbf{v}) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \left( \begin{bmatrix} 5 \\ 6 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \end{bmatrix} \right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 12 \\ 14 \end{bmatrix} = \begin{bmatrix} 12+28 \\ 36+56 \end{bmatrix} = \begin{bmatrix} 40 \\ 92 \end{bmatrix}$$
$$\mathbf{A}\mathbf{u}+\mathbf{A}\mathbf{v} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 17 \\ 39 \end{bmatrix} + \begin{bmatrix} 23 \\ 53 \end{bmatrix} = \begin{bmatrix} 40 \\ 92 \end{bmatrix}. \checkmark$$

# The Equation $\mathbf{Ax} = \mathbf{b}$

In Section 4 we saw that the linear system

$$\begin{aligned} 2x_1 + 4x_2 &= 2 \\ 4x_1 + 9x_2 &= 3 \end{aligned}$$

can be written as the vector equation

$$x_1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

We now see that this in turn can be rewritten as the matrix equation

$$\begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

which has the 'generic' form $\boxed{\mathbf{Ax} = \mathbf{b}}$, where in this case

$$\mathbf{A} = \begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

# Four Ways to write a Linear System

## Theorem

*Let* $\mathbf{A} = (a_{ij})$ *be an* $m \times n$ *matrix, with columns* $\mathbf{a}_1, \mathbf{a}_2, \cdots \mathbf{a}_n$*, and let* $\mathbf{b} \in \mathbb{R}^m$*. The following four statements are equivalent:*

1. $$\left\{ \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \right\}.$$
2. $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$.
3. $$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$
4. $\mathbf{A}\mathbf{x} = \mathbf{b}$.
5. *Equations (1)-(4) all have augmented matrix* $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$.

# Spanning $\mathbb{R}^m$

## Conventions and Observations

In view of the previous theorem and earlier definitions:

- We will henceforth refer to a general linear system as just $\mathbf{A}\mathbf{x} = \mathbf{b}$.
- Since $\mathbf{A}\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j$, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution (is consistent) if and only if $\mathbf{b}$ is a linear combination of the columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ of $\mathbf{A}$.
- Using the 'span' terminology this says that $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b} \in \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$.
- It follows that, for a given coefficient matrix $\mathbf{A}$, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent for **every** right-hand side $\mathbf{b} \in \mathbb{R}^m$ if and only if $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} = \mathbb{R}^m$.
- In that case we say that $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ **span** $\mathbb{R}^m$.

# Consistency Conditions

Summarizing the previous slide:

**Spanning Condition for Consistency**

$\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b}$ if and only if the columns of $\mathbf{A}$ span $\mathbb{R}^m$.

Using Gaussian elimination there is a systematic way to check this condition:

**Main Theorem (part of "THEOREM 4" in Lay)**

The columns of an $m \times n$ matrix $\mathbf{A}$ span $\mathbb{R}^m$ if and only if a row echelon form of $\mathbf{A}$ has a pivot in every row.

We finish this section with a sketch of the proof of this theorem,
but first an example:

# Span Example

## Example

Do the columns of $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 4 & -1 & 2 \\ 1 & 2 & 2 & 1 \end{bmatrix}$ span $\mathbb{R}^3$?

First convert the matrix to row echelon form:

$$\begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 4 & -1 & 2 \\ 1 & 2 & 2 & 1 \end{bmatrix} \overset{\text{R1,R1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 0 & -7 & -4 \\ 0 & 0 & -1 & -2 \end{bmatrix} \overset{\text{R2,R3}}{\longrightarrow}$$

$$\begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & -7 & -4 \end{bmatrix} \overset{\text{R1}}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

Question: So do the columns of $\mathbf{A}$ span $\mathbb{R}^3$?

Answer: Yes, by the main theorem, since every row in the echelon form contains a pivot. (The pivots are 1, 1 and 10.)

# Proof Outline for Main Theorem I

**if part:** Suppose that a row echelon form of $\mathbf{A}$ has a pivot in every row. Then for any $\mathbf{b}$, the reduced row echelon form of the *augmented* matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ has a form like

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & * & * & 0 & * & 0 & * & \cdots & c_1 \\ 0 & \cdots & 0 & 1 & * & \vdots & * & \cdots & c_2 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 & * & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 1 & * & \cdots & c_m \end{bmatrix}$$

for some $c_1, \ldots, c_m$ (determined by $\mathbf{b}$).

There is no row with a last-column pivot.

So the system is consistent and hence $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution (a unique one if there are no free variables, infinitely many otherwise.)

So any $\mathbf{b}$ can be expressed as a linear combination of the columns of $\mathbf{A}$.

In other words, the columns of $\mathbf{A}$ span $\mathbb{R}^m$.

# Proof Outline for Main Theorem II

<u>**only if part:**</u> Conversely, suppose that a row echelon form of $\mathbf{A}$ fails to have a pivot entry in every row.

In this case the row reduced form of the *augmented* matrix $\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix}$ looks like:

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & * & * & 0 & * & 0 & * & \cdots & c_1 \\ 0 & \cdots & 0 & 1 & * & \vdots & * & \cdots & c_2 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 & * & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & c_m \end{bmatrix}.$$

Some choice of $\mathbf{b}$ will ensure that $c_m \neq 0$. (Let $c_m = 1$ and all the other $c$'s = 0. Reverse the row operations to get back to $\mathbf{A}$ and associated $\mathbf{b}$).

So with this choice of $\mathbf{b}$ the system $\mathbf{Ax} = \mathbf{b}$ is inconsistent, which means that $\mathbf{b}$ *cannot* be expressed as a linear combination of the columns of $\mathbf{A}$.

Hence the columns of $\mathbf{A}$ do *not* span $\mathbb{R}^m$.

# 6. Solution Sets of Linear Systems (Lay 1.5)

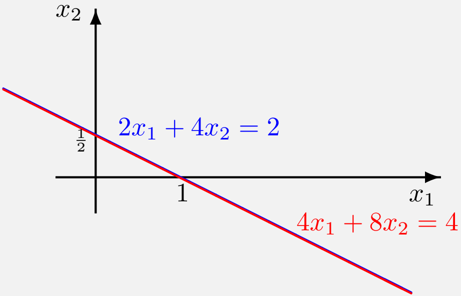
Case Study:
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_5ov.pdf

# An Example Revisited

In Section 2 we had the example

$$\begin{array}{ccccc} 2x_1 & + & 4x_2 & = & 2 \\ 4x_1 & + & 8x_2 & = & 4 \end{array}.$$

We saw that the system has infinitely many solutions because the two equations represent the same line.




Let's look again at this example, using techniques discussed in the interim. Using the augmented matrix we get

$$\begin{bmatrix} 2 & 4 & 2 \\ 4 & 8 & 4 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

So $x_2$ is a free variable. We can write the solution in terms of a parameter:

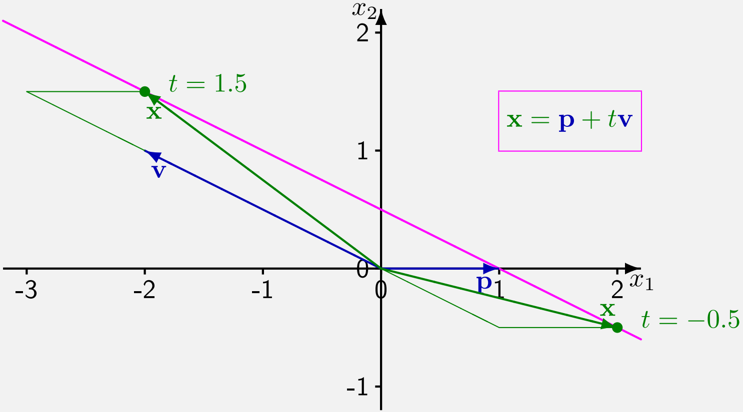
$$\begin{array}{ccc} x_1 & = & 1 - 2t \\ x_2 & = & t \end{array} \qquad \text{or, using vectors,} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

# An Example Revisited (continued)

Thus a **parametric equation** for our line is

$$\mathbf{x} = \mathbf{p} + t\mathbf{v} \qquad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Every point on the line corresponds to a value of $t$, and vice-versa.



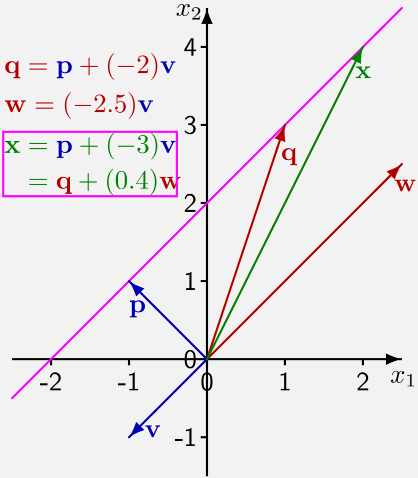
# Lines in $\mathbb{R}^n$

## Definition

For $\mathbf{p}, \mathbf{v} \in \mathbb{R}^n$, with $\mathbf{v}$ non-zero, the vector set $L = \{\mathbf{x} : \mathbf{x} = \mathbf{p} + t\mathbf{v}, t \in \mathbb{R}\}$ is called a **line** in $\mathbb{R}^n$. We refer to $L$ as the line with **parametric equation** $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ or, more informally, as just 'the line $\mathbf{x} = \mathbf{p} + t\mathbf{v}$'.

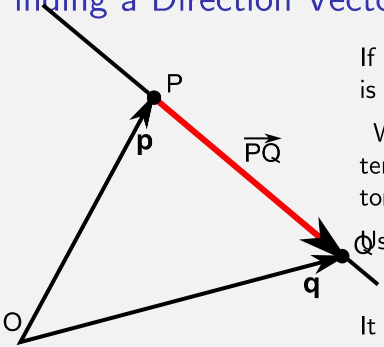
- Our previous example was for $\mathbb{R}^2$ — a line in 2D.
- Shortly we will look at examples for $\mathbb{R}^3$ — lines in 3D.
- In 2D and 3D we visualize $L$ as the geometrical line comprising all points with position vectors in $L$ (with respect to some chosen coordinate system).
- For $n > 3$ we cannot fully visualize $L$ since we live in an only three dimensional world. Such lines are sometimes called "**hyperlines**".

# Properties of the line $\mathbf{x} = \mathbf{p} + t\mathbf{v}$




- The vector $\mathbf{p}$ represents 'a point on the line'.
- $\mathbf{p}$ is not unique, it can be replaced by any other 'point on the line'; *i.e*
  $\mathbf{p}$ can be replaced by
  $\mathbf{q} = \mathbf{p} + r\mathbf{v}$ for any $r \in \mathbb{R}$.
- The vector $\mathbf{v}$ represents 'the direction of the line'.
- $\mathbf{v}$ is not unique, it can be replaced by any other vector in the same direction; *i.e*
  $\mathbf{v}$ can be replaced by $\mathbf{w} = s\mathbf{v}$ for any non-zero $s \in \mathbb{R}$.

# Finding a Direction Vector




If P and Q are points on a line then $\vec{PQ}$ is a vector in the direction of the line.

With respect to a fixed coordinate system, suppose P and Q have position vectors $\mathbf{p}$ and $\mathbf{q}$ respectively.

Using the the triangle rule we see that

$$\mathbf{p} + \vec{PQ} = \mathbf{q}.$$

It follows that

$$\boxed{\vec{PQ} = \mathbf{q} - \mathbf{p} = \text{ 'head minus tail' } \quad \text{is a direction vector for the line}}$$

---

Another way to see this is: If $\mathbf{v}$ is *any* direction vector for the line, then the line has equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$. Since Q is on the line this means that $\mathbf{q} = \mathbf{p} + t\mathbf{v}$ for some suitable value of $t$. Hence $\mathbf{q} - \mathbf{p} = t\mathbf{v}$. It follows that $\mathbf{q} - \mathbf{p}$, as a multiple of $\mathbf{v}$, is an alternative direction vector for the line.

# Equation of Line through Two Points

## Example

What is the equation of the line through the points $(8, -5, 4)$ and $(6, 3, 2)$?

For a direction vector $\mathbf{v}$ we can take the difference of the position vectors of the points:

$$\mathbf{v} = \begin{bmatrix} 8 \\ -5 \\ 4 \end{bmatrix} - \begin{bmatrix} 6 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -8 \\ 2 \end{bmatrix}.$$

For a position vector $\mathbf{p}$ we can take the position vector of either point.

Using the first point: $$\mathbf{p} = \begin{bmatrix} 8 \\ -5 \\ 4 \end{bmatrix}.$$

So a parametric equation is

$$\mathbf{x} = \overset{\overset{\displaystyle\mathbf{p}}{\downarrow}}{\begin{bmatrix} 8 \\ -5 \\ 4 \end{bmatrix}} + t \overset{\overset{\displaystyle\mathbf{v}}{\downarrow}}{\begin{bmatrix} 2 \\ -8 \\ 2 \end{bmatrix}}.$$

# Another Example in 3D

## Example

Solve the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 2 \\ 3 & 3.5 & 2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix},$$

and interpret the answer geometrically.

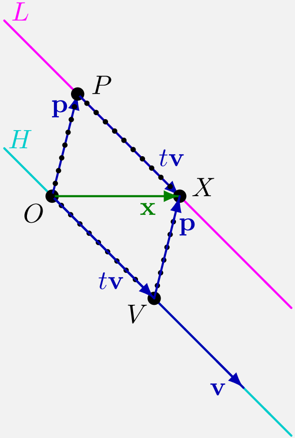
$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 & 6 \\ 2 & 5 & 2 & 8 \\ 3 & 3.5 & 2 & 7 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 3/8 & 7/8 \\ 0 & 1 & 1/4 & 5/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\longrightarrow \quad \begin{aligned} x_1 &= 7/8 - 3/8x_3 \\ x_2 &= 5/4 - 1/4x_3 \\ x_3 &= x_3 \text{ (free)} \end{aligned} \quad \longrightarrow \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7/8 \\ 5/4 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3/8 \\ -1/4 \\ 1 \end{bmatrix}$$

This has the form $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ and so represents a line through $(7/8, 5/4, 0)$ parallel to $(-3/8)\mathbf{i} + (-1/4)\mathbf{j} + \mathbf{k}$.

# A Different Interpretation of $\mathbf{x} = \mathbf{p} + t\mathbf{v}$

In the diagram: $L$ is the line with equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$;
$H$ is parallel to $L$, with equation $\mathbf{x} = t\mathbf{v}$.




So far we have interpreted $\mathbf{p} + t\mathbf{v}$ thus:

- $\mathbf{p}$ is the position vector of $P$ on $L$
- $t\mathbf{v}$ translates $P$ to $X$ along $L$.

By 'going round the parallelogram the other way' we get a different interpretation:

- $t\mathbf{v}$ is the position vector of $V$ on $H$
- $\mathbf{p}$ translates $V$ to $X$ across from $H$ to $L$.

Every point on $L$ can be reached by applying translation $\mathbf{p}$ to a corresponding point on $H$.

We say that $L$ is a **translate** of $H$.

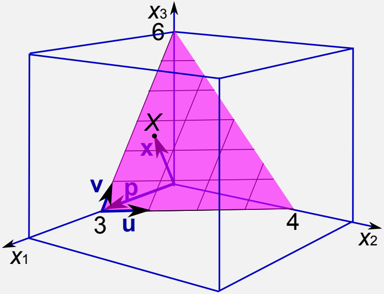
.

# Parametric equation for a plane

You are probably aware that an equation such as $4x_1 + 3x_2 + 2x_3 = 12$ can be visualized as a plane. We can find a parametric equation for this:

$$4x_1 + 3x_2 + 2x_3 = 12 \quad \to \quad \begin{bmatrix} 4 & 3 & 2 & 12 \end{bmatrix} \quad \to \quad \begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{2} & 3 \end{bmatrix}$$

$$\begin{matrix} x_2, x_3 \\ \text{free} \end{matrix} \quad \to \quad \begin{aligned} x_1 &= 3 - \tfrac{3}{4}s - \tfrac{1}{2}t \\ x_2 &= s \\ x_3 &= t \end{aligned} \quad \to \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -\frac{3}{4} \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix}$$

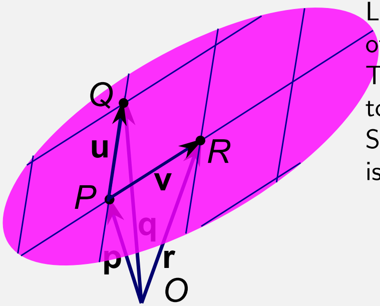



- This is a parametric equation $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$.
- Part of the plane is shown in pink.
- The parameters $s$ and $t$ serve as 'local coordinates' on the plane.
- The point $X$ has $(s, t) = (0.5, 2.5)$.
- Position vector of $X$ is $\mathbf{x} = \mathbf{p} + (0.5)\mathbf{u} + (2.5)\mathbf{v}$.

# Parametric Equation of a Plane through Three Points

**Example**

Find a parametric equation for the plane through the points $P(1, -1, 2)$, $Q(-1, -1, 6)$ and $R(-1, 1, 4)$.




Let $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r}$ be the position vectors of $P$, $Q$ and $R$ respectively.

Then $\mathbf{u} = \mathbf{q} - \mathbf{p}$ and $\mathbf{v} = \mathbf{r} - \mathbf{p}$ are vectors with directions lying in the plane.

So a parametric equation for the plane is $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$, *i.e.*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + s \begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix} + t \begin{bmatrix} -2 \\ 2 \\ 2 \end{bmatrix}.$$

Note that, as for the parametric equation of a line, there is nothing unique about $\mathbf{p}$, $\mathbf{u}$ or $\mathbf{v}$. For example another equation for the plane is $\mathbf{x} = \mathbf{q} + s(\mathbf{p} - \mathbf{q}) + t(\mathbf{r} - \mathbf{q})$.

# Implicit Equation of a Plane

Given a parametric equation for a plane, how to we convert it to a linear equation, called an **implicit equation**, for that plane?

## Example

Find an implicit equation for the plane with parametric equation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + t \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}$$

We seek implicit equation $c_1x_1 + c_2x_2 + c_3x_3 = b$. Take $b = 1$ (by scaling).

Substitute into this equation the coordinate of three points on the plane. Taking $(s, t) = (0, 0)$, $(1, 0)$ and $(0, 1)$ gives

$$\begin{aligned} c_1 + \ \ c_2 + 2c_3 &= 1 \\ 2c_1 + 2c_2 + 0c_3 &= 1 \\ 3c_1 + 0c_2 + 0c_3 &= 1 \end{aligned} \quad \rightarrow \quad \begin{bmatrix} 1 & 1 & 2 & 1 \\ 2 & 2 & 0 & 1 \\ 3 & 0 & 0 & 1 \end{bmatrix} \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{6} \\ 0 & 0 & 1 & \frac{1}{4} \end{bmatrix}$$

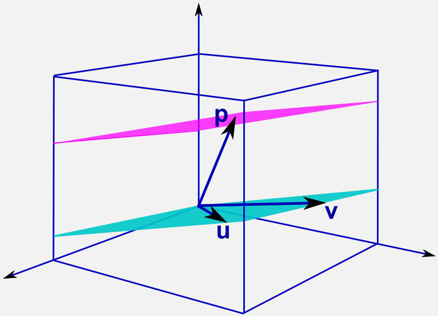
So our equation is $\frac{1}{3}x_1 + \frac{1}{6}x_2 + \frac{1}{4}x_3 = 1$, or $4x_1 + 2x_2 + 3x_3 = 12$.

# Planes in $\mathbb{R}^n$

## Definition

For $\mathbf{p}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, with $\mathbf{u}$ and $\mathbf{v}$ non-zero and not multiples of each other, the vector set $\{\mathbf{x} : \mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v},\ s, t \in \mathbb{R}\}$ is called a **plane** in $\mathbb{R}^n$. We call it the plane with **parametric equation** $\mathbf{x} = \mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$ or, more informally, as just 'the plane $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$'.

- The restrictions on $\mathbf{u}$ and $\mathbf{v}$ ensure that the plane is not a line.
- The definition generalizes the concept of a 'real' plane in 3D.
- For $n = 2$ the plane is just $\mathbb{R}^2$.
- For $n > 3$ the plane is sometimes called a hyperplane of dimension 2 or just a 'hyperplane'.
- Any plane can be viewed as a translate of a plane through the origin. (See diagram for $n = 3$.)



# General Solution of $\mathbf{Ax} = \mathbf{b}$

We now wish to use the basic ideas of parametric description of lines and planes to describe the solution set of a general linear system $\mathbf{Ax} = \mathbf{b}$.

# Homogeneous Systems

## Definition

A system of equations is **homogeneous** if it is of the form $\mathbf{Ax} = \mathbf{0}$.

In this case there is always the **trivial solution**, namely $\mathbf{x} = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

## Theorem

*The homogeneous equation $\mathbf{Ax} = \mathbf{0}$ has a non-trivial solution if and only if the equation has at least one free variable.*

In particular; if $\mathbf{A}$ is an $m \times n$ matrix with $m < n$, then the equation $\mathbf{Ax} = \mathbf{0}$ has a non-trivial solution.

# Example of a Homogeneous System

## Example

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

How many leading variables?
How many free variables?
Does the corresponding homogeneous matrix equation have non-trivial solutions?

# A Closer Look at the Last Example

Using the rref of the augmented matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{array}{rcl} x_1 & = & -x_2 - x_5, \\ x_2 & = & x_2 \text{ (free)}, \\ x_3 & = & -x_5, \\ x_4 & = & 0, \\ x_5 & = & x_5 \text{ (free )}. \end{array} \rightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

Now the solution is written in the parametric form

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v}; \quad \text{where} \quad \mathbf{u} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus the solution set is $\text{span}\{\mathbf{u}, \mathbf{v}\}$, which is a (hyper)plane passing through the origin.

# More Examples.

## Example

For each matrix:

- How many free variables?
- Are there any non-trivial solutions to $\mathbf{A}\mathbf{x} = \mathbf{0}$?
- Describe the solution set.

1. $\mathbf{A} = \begin{bmatrix} 2 & -3 & 1 \end{bmatrix}$ (So $\mathbf{A}\mathbf{x} = \mathbf{0}$ is just $2x_1 - 3x_2 + x_3 = 0$.)
2. $\mathbf{A} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 2 \\ 3 & 3.5 & 2 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 3/8 \\ 0 & 1 & 1/4 \\ 0 & 0 & 0 \end{bmatrix}$
3. $\mathbf{A} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 2 \\ 3 & 1.5 & 5 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

# Comparing Homogeneous and Non-Homogeneous Systems

## Example

Let $\mathbf{A}, \mathbf{b} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 5 & 2 \\ 3 & 3.5 & 2 \end{bmatrix}, \begin{bmatrix} 6 \\ 8 \\ 7 \end{bmatrix}$. Compare solutions to $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Ax} = \mathbf{0}$.

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 & 6 & 0 \\ 2 & 5 & 2 & 8 & 0 \\ 3 & 3.5 & 2 & 7 & 0 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 3/8 & 7/8 & 0 \\ 0 & 1 & 1/4 & 5/4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

| Non-homogeneous case $\mathbf{Ax} = \mathbf{b}$ | Homogeneous case $\mathbf{Ax} = \mathbf{0}$ |
|---|---|
| $\begin{aligned} x_1 &= \tfrac{7}{8} - \tfrac{3}{8}x_3, \\ x_2 &= \tfrac{5}{4} - \tfrac{1}{4}x_3, \\ x_3 &= x_3\text{(free)}. \end{aligned}$ *i.e.* $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{7}{8} \\ \frac{5}{4} \\ 0 \end{bmatrix} + t \begin{bmatrix} -\frac{3}{8} \\ -\frac{1}{4} \\ 1 \end{bmatrix}$ | $\begin{aligned} x_1 &= -\tfrac{3}{8}x_3, \\ x_2 &= -\tfrac{1}{4}x_3, \\ x_3 &= x_3\text{(free)}. \end{aligned}$ *i.e.* $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -\frac{3}{8} \\ -\frac{1}{4} \\ 1 \end{bmatrix}$ |

Both solutions represent lines. The line representing the solution of the homogeneous equation passes through the origin.
The other line is a translate of this.

# General Solution to a Non-homogeneous System

## Theorem (Lay Theorem 6 §1.5)

*If $\mathbf{p}$ is any solution to the equation $\mathbf{Ax} = \mathbf{b}$, then the solutions of $\mathbf{Ax} = \mathbf{b}$ are all vectors of the form $\mathbf{p} + \mathbf{h}$ where $\mathbf{h}$ is any solution of the homogeneous equation $\mathbf{Ax} = \mathbf{0}$.*

Proof: Let $\mathbf{p}$ be any fixed ('particular') solution to th equation $\mathbf{Ax} = \mathbf{b}$.
Then by definition $\mathbf{Ap} = \mathbf{b}$ (is true).
We have to prove, for any $\mathbf{x}$:

$$\mathbf{Ax} = \mathbf{b} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{p} + \mathbf{h} \text{ where } \mathbf{Ah} = \mathbf{0}.$$

Check that each implication in the chain below really is two-way:

$$\begin{aligned} \mathbf{Ax} = \mathbf{b} \; &\Leftrightarrow \; \mathbf{Ax} = \mathbf{Ap} \quad \text{def. of } \mathbf{p} \\ &\Leftrightarrow \; \mathbf{Ax} - \mathbf{Ap} = \mathbf{0} \\ &\Leftrightarrow \; \mathbf{A}(\mathbf{x} - \mathbf{p}) = \mathbf{0} \quad \text{(linearity)} \\ &\Leftrightarrow \; \mathbf{x} - \mathbf{p} = \mathbf{h} \text{ where } \mathbf{Ah} = \mathbf{0}. \\ &\Leftrightarrow \; \mathbf{x} = \mathbf{p} + \mathbf{h} \text{ where } \mathbf{Ah} = \mathbf{0} \end{aligned}$$

# Graphical Interpretations

We have defined lines and planes in $\mathbb{R}^n$.

- A **line** is a one parameter family given by $\mathbf{x} = \mathbf{p} + t\mathbf{v}$.
- A **plane** is a two parameter family given by $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$.

For $n > 3$ these are *generalizations* of the the geometrical objects with these names, so we sometimes use the terms 'hyperline' and 'hyperplane' instead. However, the term 'hyperplane' is used even more generally:

- A **hyperplane** is a $k$ parameter family (any $k \geq 1$) given by
$$\mathbf{x} = \mathbf{p} + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + +t_k\mathbf{v}_k.$$
- Lines and planes are special cases of hyperplanes.

## Graphical Interpretations of Solution Sets of Linear Systems

The solutions of a homogeneous equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ form a hyperplane through the origin. The solutions, *if there are any*, of a corresponding inhomogeneous equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ (with the same coefficient matrix $\mathbf{A}$), form a translate of this hyperplane.

# Existence and Uniqueness Conditions

Suppose $\mathbf{A}$ is an $m \times n$ matrix. The table below gathers together some properties of $\mathbf{A}$ that we have observed to have an influence on the solution(s) to the equation $\mathbf{Ax} = \mathbf{b}$ for general $\mathbf{b}$.

| Existence | Uniqueness |
|---|---|
| Equivalent conditions for $\mathbf{Ax} = \mathbf{b}$ to have at **least** one solution for every $\mathbf{b} \in \mathbb{R}^m$:<br>• The columns of $\mathbf{A}$ span $\mathbb{R}^m$.<br>• $\mathbf{A}$ has a pivot position in every *row*. This condition implies (but is NOT equivalent to) $n \geq m$. | Equivalent conditions for $\mathbf{Ax} = \mathbf{b}$ to have at **most** one solution for every $\mathbf{b} \in \mathbb{R}^m$:<br>• Equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.<br>• $\mathbf{A}$ has a pivot position in every *column*. This condition implies (but is NOT equivalent to) $n \leq m$. |

# Square Matrices

For obvious reasons, an $m \times n$ matrix with $m = n$ is called **square**.

Using the conditions listed on the previous slide we obtain:

### Theorem

- *If a square matrix* $\mathbf{A}$ *has a pivot position in every row (equivalently in every column) then the equation* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for every right hand side* $\mathbf{b}$.
- *Conversely, if a matrix* $\mathbf{A}$ *has the property that the equation* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for every right hand side* $\mathbf{b}$*, then* $\mathbf{A}$ *must be square and have a pivot position in every row (equivalently in every column).*

Note that all the conditions in this theorem relate directly to $\mathbf{A}$, not to some augmentation of $\mathbf{A}$.

# Independent Vectors

The condition that the homogeneous equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$ is very important in its own right.

Explicitly, the condition says that if $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n\end{bmatrix}$ then the only solution to
$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$
is $x_1 = \cdots = x_n = 0$.

When this condition is met we say that the columns of $\mathbf{A}$ are **independent**.

We will look at his case again later when we study more general sets of independent vectors.

# 7. Linear Independence (Lay 1.7)

Case Study:
Lay's Summary: `http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_7ov.pdf`

# Linear Independence

## Definition

To say that a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of vectors in $\mathbb{R}^m$ is **linearly independent** means that the only way the equation

$$x_1\mathbf{v}_1 + \cdots + x_k\mathbf{v}_k = \mathbf{0}$$

can hold is when all the weights $x_1, \ldots, x_k$ are zero.

A set of vectors which is not linearly independent is linearly **de**pendent:

## Definition

The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is **linearly dependent** if there are weights $x_1, \ldots, x_k$, *not all zero* such that

$$x_1\mathbf{v}_1 + \cdots + x_k\mathbf{v}_k = \mathbf{0}.$$

# Linear Independence of Matrix Columns

Recall that if $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$, then saying that the equation $\mathbf{Ax} = \mathbf{0}$ has no non-trivial solutions is equivalent to saying that the equation

$$x_1\mathbf{a}_1 + \cdots + x_k\mathbf{a}_k = \mathbf{0}$$

has only the trivial solution $x_1 = x_2 = \cdots = x_n = 0$.

This can be expressed by saying that $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent.

We also had a theorem that said that for the equation $\mathbf{Ax} = \mathbf{0}$ to have no non-trivial solutions there must be no free variables. This is equivalent to saying that every variable is leading, or in other words, there is a pivot position in every column.

## Theorem

*The following statements are equivalent:*

- *The columns of $\mathbf{A}$ are linear independent.*
- *The equation $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.*
- *Every column of $\mathbf{A}$ contains a pivot position.*

# Linear Independence Example

## Example

Are the columns of the following matrix linearly independent?

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

$$\begin{bmatrix} 3 & 4 & 3 \\ -1 & -7 & 7 \\ 1 & 3 & -2 \\ 0 & 2 & -6 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

So are the columns of $\mathbf{A}$ linearly independent?

# Some Simple Cases

## Illustration 1

Two vectors $\mathbf{u}, \mathbf{v}$ are linearly dependent if and only if one is a (scalar) multiple of the other.

Example $\Rightarrow$: If $2\mathbf{u} + 3\mathbf{v} = \mathbf{0}$ then $\mathbf{u} = \left(\frac{-3}{2}\right)\mathbf{v}$.
Example $\Leftarrow$: If $\mathbf{u} = 6\mathbf{v}$ then $(1)\mathbf{u} + (-6)\mathbf{v} = \mathbf{0}$.

## Illustration 2

Any collection of $n$ vectors in $\mathbb{R}^m$, with $n > m$, is linearly dependent.

Proof: Think of the vectors as columns of an $m \times n$ matrix $\mathbf{A}$.
More columns than rows $\Rightarrow$ some column(s) has no pivot position.

## Illustration 3

Three vectors in 2-space must be linearly dependent.

Proof: Special case of Illustration 2; $m = 2$, $n = 3$.

# A Characterisation of Linear Dependence

**Theorem (c.f. Lay, Theorem 7, §1.6 $2^{\text{nd}}$ ed. and §1.7 $3^{\text{rd}}$ ed.)**

*A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, $k > 1$, of non-zero vectors, is linearly dependent if and only if some $\mathbf{v}_j$, $j > 1$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.*

Illustration:

The vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 3 \\ -2 \\ 5 \end{bmatrix}$, $\mathbf{v}_4 = \begin{bmatrix} 3 \\ -3 \\ 0 \\ 1 \end{bmatrix}$

are linearly dependent because $2\mathbf{v}_1 - \mathbf{v}_2 - \mathbf{v}_3 = \mathbf{0}$, or, more formally,

$$(2)\mathbf{v}_1 + (-1)\mathbf{v}_2 + (-1)\mathbf{v}_3 + (0)\mathbf{v}_4 = \mathbf{0}.$$

In this case we can take $j = 3$ in the theorem, since

$$\mathbf{v}_3 = 2\mathbf{v}_1 - \mathbf{v}_2.$$

# Proof of Characterisation Theorem

$\Rightarrow$: Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent, say

$$x_1\mathbf{v}_1 + \cdots + x_k\mathbf{v}_k = \mathbf{0}$$

where $x_i \neq 0$ for at least one $i$. Choose $j$ to be the greatest of these $i$'s. Then we have

$$\frac{x_1}{x_j}\mathbf{v}_1 + \cdots + \frac{x_j}{x_j}\mathbf{v}_j = \mathbf{0},$$

the other terms being zero. The last coefficient here is 1, and so

$$\mathbf{v}_j = -\tfrac{x_1}{x_j}\mathbf{v}_1 + \cdots + \tfrac{x_{j-1}}{x_j}\mathbf{v}_{j-1}$$

as required.

$\Rightarrow$: Suppose $\mathbf{v}_j$, $j > 1$, is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$, say

$$\mathbf{v}_j = x_1\mathbf{v}_1 + \cdots + x_{j-1}\mathbf{v}_{j-1}.$$

Then

$$x_1\mathbf{v}_1 + \cdots + x_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j + (0)\mathbf{v}_{j+1} + \cdots + (0)\mathbf{v}_k = \mathbf{0}$$

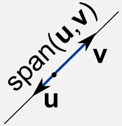
that is, $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is linearly dependent.

# Graphical Interpretation of Linear Dependence I

## Theorem (Two linearly dependent vectors)

*Non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent if and and only if they have the same (or opposite) direction.*
*In this case span$\{\mathbf{u}, \mathbf{v}\}$ is a (hyper)line through the origin.*

Proof idea:
By the characterisation theorem $\mathbf{v}$ is a multiple of $\mathbf{u}$.



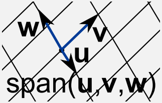
# Graphical Interpretation of Linear Dependence II

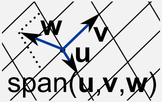
## Theorem (Three linearly dependent vectors)

*Non-zero vectors* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{w}$ *are linearly dependent if and and only if they are coplanar (i.e. their directions are all parallel to a common plane).*
*In this case span*$\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ *is either a (hyper)plane, or a (hyper)line, through the origin.*

Proof idea:
By the characterisation theorem

- either (i) $\mathbf{v}$ and $\mathbf{w}$ are multiples of $\mathbf{u}$
- or (ii) $\mathbf{v}$ but not $\mathbf{w}$ is multiple of $\mathbf{u}$
- or (iii) $\mathbf{w}$ but not $\mathbf{v}$ is multiple of $\mathbf{u}$
- else (iv) $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$




Case (i) gives a line, the others a plane.
Cases (iii) and (iv) are illustrated.

# Example: Linear Independence

## Example

Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. Are these vectors coplanar?

If $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ lie in a plane then $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ must be linearly dependent.
To test this, let $\mathbf{A} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$.
We need to know whether the equation $\mathbf{Ax} = \mathbf{0}$ has a nontrivial solution.

By a previous theorem this will be the case if and only if $\mathbf{A}$ fails to have a pivot position in every column.

The reduced row echelon form is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

So there is a pivot in every column.
So there are no non-trivial solutions.
So the vectors are linearly independent.
So the vectors are **not** coplanar.

# Example: Linear Dependence

## Example

Let $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 7 \\ -6 \end{bmatrix}$.
Are these vectors coplanar? If so, give an equation for a suitable plane.

This time $\mathbf{A} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3] \quad \xrightarrow{\text{rref}} \quad \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

# Implicit Definition of a Plane

Finding an implicit definition of a plane was demonstrated in section 6.
Here is another example; this time the plane passes through the origin.
The example follows on from the previous slide, and uses the same data.

## Example

Find an implicit equation for the plane containing the points with position vectors

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 7 \\ -6 \end{bmatrix}.$$

From the previous example we know that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are coplanar.
This implies that our plane passes through the origin. (*Explain*)
So we seek a *homogeneous* equation

$$ax_1 + bx_2 + cx_3 = 0$$

that is satisfied by the coordinates of each of the three points.

# Implicit Definition of a Plane (cont.)

Substituting each coordinate triple $(x_1, x_2, x_3)$ into $ax_1 + bx_2 + cx_3 = 0$ gives

$$\begin{aligned} 2a - \ b + 4c &= 0 \\ 4a + 2b + 3c &= 0 \\ 2a + 7b + 6c &= 0 \end{aligned} \qquad \textit{i.e.} \qquad \begin{bmatrix} 2 & -1 & 4 \\ 4 & 2 & 3 \\ 2 & 7 & -6 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

*Note that the matrix here is* **not** *what we had before.*
*It has been transposed – the rows and columns have been interchanged.*

$$\begin{bmatrix} 2 & -1 & 4 & 0 \\ 4 & 2 & 3 & 0 \\ 2 & 7 & -6 & 0 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 11/8 & 0 \\ 0 & 1 & -5/4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{aligned} a &= (-11/8)c \\ b &= (5/4)c \\ c &= c \end{aligned} .$$

To eliminate fractions take say $c = -8$, giving $a = 11$ and $b = -10$.
So an implicit equation for our plane is

$$11x_1 - 10x_2 - 8x_3 = 0.$$

## 8. Linear Transformations (Parts of Lay 1.8 and 1.9)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/lay03_02_cs.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_8ov.pdf and
http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_9ov.pdf

# Matrices as a Shorthand Notation

- Matrices have so far been viewed mainly as a shorthand notation for representing systems of linear equations.
- That is, we encapsulate a system of $m$ linear equations in $n$ unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

  as simply $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{Ax} = \mathbf{b} \text{ means } \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

- 'Solving the system of equations' becomes 'solving $\mathbf{Ax} = \mathbf{b}$'.
- This means finding $\mathbf{x}$ given $\mathbf{A}$ and $\mathbf{b}$.
  There may be a unique such $\mathbf{x}$, or none, or an infinite many.

# Matrices as Transformations

We now look at the matrix $\mathbf{A}$ in a more active role:

$\mathbf{A}$ acts on $\mathbf{x}$ to produce $\mathbf{Ax}$.

We capture this idea formally by defining a function (a mapping, a transformation) associated with the matrix $\mathbf{A}$.

## Definition (Matrix Transformation)

Let $\mathbf{A}$ be and $m \times n$ matrix. The **Matrix Transformation** $T$ associated with $\mathbf{A}$ is the function $T : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$T(\mathbf{x}) = \mathbf{Ax} \text{ for every } \mathbf{x} \in \mathbb{R}^n.$$

- The matrix transformation may also be written as $T : \mathbf{x} \mapsto \mathbf{Ax}$. This is read as "ex maps to A ex'.
- $T$ has domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$.
- 'Solving $T(\mathbf{x}) = \mathbf{b}$' means finding which $\mathbf{x}$, if any, are mapped to $\mathbf{b}$ (by $T$). This is the same as 'solving $\mathbf{Ax} = \mathbf{b}$'.

# Matrix Transformation Example

## Example

Let $\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -4 \\ 3 & 2 & 1 \\ -2 & -1 & -2 \end{bmatrix}$ and let $\mathbf{b} = \begin{bmatrix} 1 \\ -5 \\ -7 \\ 3 \end{bmatrix}$.

Let $T$ be the matrix transformation associated with $\mathbf{A}$, given by

$$T(\mathbf{x}) = \mathbf{A} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 3x_3 \\ x_2 - 4x_3 \\ 3x_1 + 2x_2 + x_3 \\ -2x_1 - x_2 - 2x_3 \end{bmatrix}.$$

1. Is there *some* $\mathbf{x}$ which maps to $\mathbf{b}$?
2. Which vectors $\mathbf{x}$ map to $\mathbf{0}$?
3. Does $T$ map *onto* $\mathbb{R}^4$?

# Answering Question 1

*Is there some $\mathbf{x}$ which maps to $\mathbf{b}$?*

We must solve $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -4 \\ 3 & 2 & 1 \\ -2 & -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ -7 \\ 3 \end{bmatrix}$.

The augmented matrix of this system reduces to $\begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & -4 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

This yields the solution in parametric form

$$\mathbf{x} = \begin{bmatrix} 1 \\ -5 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$$

for any $t \in \mathbb{R}$. So the short answer to Question 1 is YES.

# Answering Question 2

*Which* $\mathbf{x}$ *maps to* $\mathbf{0}$*?*

We need to do the same as for Question 1 but with $\mathbf{b} = \mathbf{0}$.

So the answer can be read off the previous parametric solution, giving

$$\mathbf{x} = t \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$$

for any $t \in R$.

# Answering Question 3

*Does $T$ map onto $\mathbb{R}^4$?*

In this case a general $\mathbf{b}$ must be considered:

$$\begin{bmatrix} 1 & 0 & 3 & a \\ 0 & 1 & -4 & b \\ 3 & 2 & 1 & c \\ -2 & -1 & -2 & d \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 3 & a \\ 0 & 1 & -4 & b \\ 0 & 0 & 0 & c-3a-2b \\ 0 & 0 & 0 & d+2a+b \end{bmatrix}.$$

So the **two** conditions

$$\begin{aligned} c &= 3a+2b \\ d &= -2a-b \end{aligned}$$

must be satisfied.

This condition is enough to show that $T$ does **NOT** map onto $\mathbb{R}^4$.

# More on Answering Question 3

The conditions mean that the vector

$$\mathbf{b} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} a \\ b \\ 3a+2b \\ -2a-b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 3 \\ -2 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 2 \\ -1 \end{bmatrix}$$

is the image of

$$\mathbf{x} = \begin{bmatrix} a - 3x_3 \\ b + 4x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$$

for any $t \in \mathbb{R}$.

Only such $\mathbf{b}$ are in the range of $T$.

# Comments on Answering Questions 1 - 3

It is apparent that answering these questions about $T$ required nothing new.

In particular

- the range of $T$ is exactly the span of the columns of $A$.
- $T(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x}$ solves the homogeneous equation $\mathbf{A}\mathbf{x} = \mathbf{0}$.

# Linear Transformations

All these transformations $T$ have an important property which comes from similar properties of matrices noted previously – they are linear.

## Definition

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is **linear** if

1. $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}),$ (for all $\mathbf{u}$,$\mathbf{v} \in \mathbb{R}^n$).
2. $T(c\mathbf{u}) = cT(\mathbf{u}),$ (for all $\mathbf{u} \in \mathbb{R}^n$ and $c \in \mathbb{R}$).

Equivalently,

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}),$$

for all $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$, and $c, d \in \mathbb{R}$.

In other words, a linear transformation preserves the operations of addition of vectors and of multiplication by scalars.

# Comments on Linear Transformations

- Linear transformations are ubiquitous in modern mathematics and engineering.
- Many physical, chemical and electrical processes behave in a 'linear' fashion. (Though of course some important ones do not.)
- It is ideas of linearity which underlie calculus where we are for ever approximating non-linear functions locally by linear functions (the tangent or the tangent plane in higher dimensions).

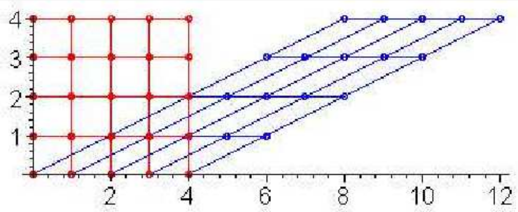
# Example: Shear Transformations.

Lay gives the example of a *shear* transformation in 2D, which is one associated with a matrix of the form

$$(1) \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad (2) \begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}.$$

## Example

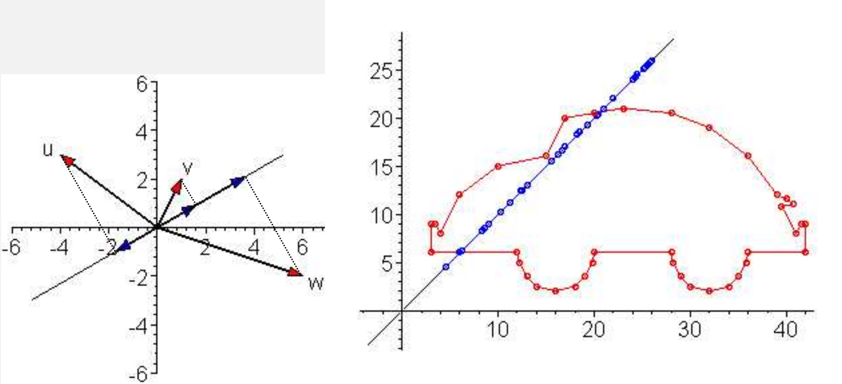
In the diagram below the red grid is sheared to the blue one.
What matrix $\mathbf{A}$ is associated with this transformation $T$?



# Projections onto a line in 2D

## Example

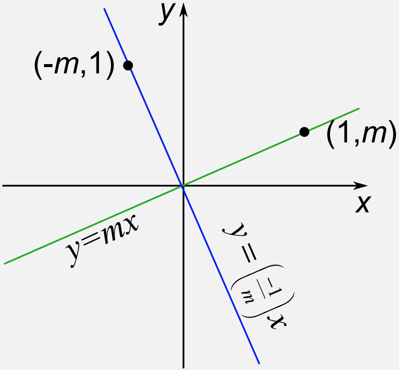
Find the matrix $\mathbf{A}$ such that the transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ is a projection $P$ that projects all points in the $x, y$-plane onto the line $y = mx$.
See the diagrams below, where again red is transformed to blue.



# Projections onto a line in 2D (continued)

It's not immediately obvious that *any* matrix transformation will work. However, assuming there is one, what does it need to do?

- The transformation $P$ should map any point on the line $y = mx$ to itself.
  *Such a point is* $(1, m)$.
- The transformation $P$ should map any point on the line $y = (\frac{-1}{m})x$ to the origin, since this line is perpendicular to $y = mx$ and passes through O.
  *Such a point is* $(-m, 1)$.




So the associated matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ should satisfy

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ m \end{bmatrix} = \begin{bmatrix} 1 \\ m \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} -m \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

# Projections onto a line in 2D (concluded)

This gives us 4 linear equations for *a*,*b*,*c*,*d*, namely

$$\begin{array}{rcccc} a & + & mb & = & 1, \\ -ma & + & b & = & 0, \\ c & + & md & = & m, \\ -mc & + & d & = & 0. \end{array}$$

Solving for $a$, $b$, $c$ and $d$ gives

$$\mathbf{A} \;=\; \frac{1}{1+m^2}\left[\begin{array}{cc} 1 & m \\ m & m^2 \end{array}\right].$$

The corresponding transformation is given by

$$\begin{aligned} P(\mathbf{x}) &= \frac{1}{1+m^2}\begin{bmatrix} 1 & m \\ m & m^2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\ &= \frac{x+my}{1+m^2}\begin{bmatrix} 1 \\ m \end{bmatrix}. \end{aligned}$$

# The Standard Matrix of a Linear Transformation

## Theorem ("THEOREM 10" in Lay)

*Let* $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then there exists a unique* $m$ *by* $n$ *matrix* $\mathbf{A}$ *such that*

$$T(\mathbf{x}) = \mathbf{A}\mathbf{x} \textit{ for all } \mathbf{x} \in \mathbb{R}^n.$$

*In fact* $\mathbf{A}$ *is the matrix whose* $j$*th column is the vector* $T(\mathbf{e}_j)$, *where* $\mathbf{e}_j$ *is the unit vector in* $\mathbb{R}^n$ *whose* $j$*-th entry is 1 (with all other entries 0).*

$$\mathbf{A} = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix}.$$

# Standard Matrix Example

## Example

Determine the matrix for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ for which

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The standard matrix of $T$ is given by

$$\mathbf{A} = [T(\mathbf{e}_1) \;\; T(\mathbf{e}_2)] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

# 9. Applications (parts of Lay 1.6,9,10 and 2.6,7)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/currents.pdf
Lay's Summaries: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_6ov.pdf, http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec1_10ov.pdf and http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c01/pdf_files/sec2_7ov.pdf

# Applications

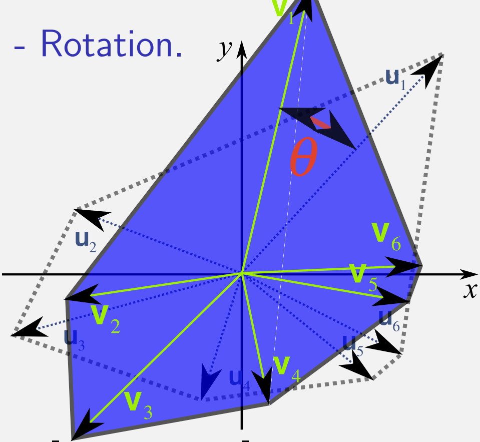
In this section we will briefly survey some uses of linear algebra in the following fields:

- Computer graphics
- Population dynamics
- An elementary economic model
- Simple circuits

The aim is to try to give an impression of the diversity of applications of linear algebra.

There will be no attempt to provide any kind of systematic introduction to any of these applications.

# Computer Graphics I - Rotation.




Suppose the vertices of a polygon in $\mathbb{R}^2$ are given by position vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$.

If this polygon is rotated (anticlockwise) around the origin by an angle $\theta$, it can be shown that the position vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, of the resulting polygon are given by

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \mathbf{u_1} & \cdots & \mathbf{u_k} \end{bmatrix}. \tag{1}$$

Remark: The rotation matrix in formula (1) is easily obtained using the standard matrix theorem.

# Computer Graphics II - Rotation Example

## Example

The line segment from (-1,1) to (0,1) is rotated clockwise around the origin through $60°$. Calculate the locations of the end points of the line segment after the rotation.

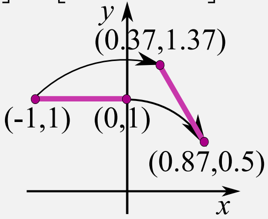
A clockwise rotation is negative, so the rotation matrix is

$$\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} = \begin{bmatrix}\cos(-60°) & -\sin(-60°) \\ \sin(-60°) & \cos(-60°)\end{bmatrix} = \begin{bmatrix}0.5 & 0.87 \\ -0.87 & 0.5\end{bmatrix}.$$

So (-1,1) and (0,1) are rotated to the new positions given by

$$\begin{bmatrix}0.5 & 0.87 \\ -0.87 & 0.5\end{bmatrix}\begin{bmatrix}-1 & 0 \\ 1 & 1\end{bmatrix} = \begin{bmatrix}0.37 & 0.87 \\ 1.37 & 0.5\end{bmatrix}.$$

That is, the new coordinates are (0.37,1.37) and (0.87,0.5).



# Computer Graphics III - Homogeneous Coordinates

Translation by a fixed non-zero vector is **not** a linear transformation.

To force a matrix to effect a translation we use the following device:

### Definition

Make *all* the vectors $\mathbf{u}_i, \mathbf{v}_i, (i = 1 \dots k)$ into 3-vectors by setting the third co-ordinate to 1. These are called **homogeneous coordinates**.

For example $\begin{bmatrix} 3 \\ -5 \end{bmatrix}$ becomes $\begin{bmatrix} 3 \\ -5 \\ 1 \end{bmatrix}$.

With homogeneous coordinates the rotation matrix is almost unchanged:

$$\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{bmatrix}.$$

# Computer Graphics IV - Translation

But now if $\mathbf{w} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \mathbb{R}^2$ is a fixed vector, then

$$\begin{bmatrix} x+\alpha \\ y+\beta \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

and we have a representation of the translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{w}$ via matrix multiplication.

Using homogeneous coordinates, some linear transformations will result in a vector whose last component is not 1; say $d \neq 1$ instead. Unless $d = 0$ this vector is deemed to be equivalent to the vector obtained from it by dividing all components by $d$.

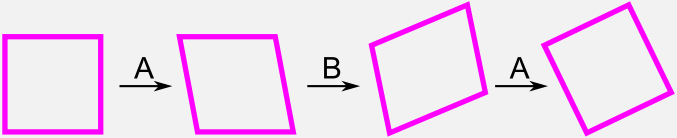
So $\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \equiv \begin{bmatrix} 3/5 \\ 4/5 \\ 1 \end{bmatrix}$; both 'represent' the point with $x, y$-coords (0.6,0.8).

A vector with 3rd component zero is deemed to represent a 'point at infinity'.

# Computer Graphics V - Rotation as Sequence of Shears

Rotation can be realized as a product of three shears:

$$\overset{\textstyle\mathbf{A}\atop\downarrow}{\begin{bmatrix} 1 & -\tan\frac{\theta}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \overset{\textstyle\mathbf{B}\atop\downarrow}{\begin{bmatrix} 1 & 0 & 0 \\ \sin\theta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \overset{\textstyle\mathbf{A}\atop\downarrow}{\begin{bmatrix} 1 & -\tan\frac{\theta}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$




This 'factorization' of a rotation matrix effectively reduces the number of scalar multiplications needed to perform a rotation of one vector from 4 to 3, though the number of additions has increased.

At the lowest level, it can take up to 5 times as long to do multiplication as it does to do an addition.

# Population Dynamics I

## Example

Consider a biological population in which there are three stages of development of the females:

- dependent (age up to 1 year),
- juvenile (age between 1 and 2 years),
- adult (age over 2 years).

Suppose that, in each year, 60% of the dependents, 75% of the juveniles, and 90% of the adults survive. Suppose also that 42% of adult females produce (a single) female offspring each year.
Investigate how the population develops over time.

Let $\mathbf{x}^{(k)} \in \mathbb{R}^3$ give the numbers of dependents, juveniles and adults at the end of the $k$-th year. Then using the given transition proportions we get:

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} 0 & 0 & 0.42 \\ 0.6 & 0 & 0 \\ 0 & 0.75 & 0.9 \end{bmatrix} \mathbf{x}^{(k)} = \mathbf{A}\mathbf{x}^{(k)}.$$

# Population Dynamics II

So starting with 100 adult females (only), $\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 100 \end{bmatrix}$,

$$\begin{aligned}
\mathbf{x}^{(1)} &= \mathbf{A}\mathbf{x}^{(0)} = \begin{bmatrix} 0 & 0 & 0.42 \\ 0.6 & 0 & 0 \\ 0 & 0.75 & 0.9 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 100 \end{bmatrix} = \begin{bmatrix} 42 \\ 0 \\ 90 \end{bmatrix}, \\
\mathbf{x}^{(2)} &= \mathbf{A}\mathbf{x}^{(1)} = \begin{bmatrix} 37.8 \\ 25.2 \\ 81 \end{bmatrix}, \mathbf{x}^{(3)} = \mathbf{A}\mathbf{x}^{(2)} = \begin{bmatrix} 34.0 \\ 22.7 \\ 91.8 \end{bmatrix}, \ldots, \\
\mathbf{x}^{(20)} &= \begin{bmatrix} 108.3 \\ 60.9 \\ 274.9 \end{bmatrix}, \ldots, \mathbf{x}^{(50)} = \begin{bmatrix} 741.9 \\ 417.5 \\ 1883.5 \end{bmatrix}, \ldots
\end{aligned}$$

It is clear that the population is increasing.

# Population Dynamics III

There are a number of more quantitative questions we may like to ask.

1. Exactly how fast is the population increasing?
   - in fact $\approx (1.0662)^k$
2. What is the population distribution across the stages after many years?
   - in fact $\approx 39 : 22 : 100$

You will be able to answer these questions yourselves after you have learned about eigenvalues and eigenvectors, next semester.

# An Elementary Economic Model I

## Example

Three neighbours grow peas, beans and carrots respectively and they agree to share the vegetables amongst themselves in the following manner (amounts in kilos of vegetables).

| | Peas | Beans | Carrots |
|---|---|---|---|
| Pea grower uses | 1 | 1 | 2 |
| Bean grower uses | 2 | 3 | 1 |
| Carrot grower uses | 2 | 1 | 2 |

What is a reasonable price for each to charge per kilo so the net cost is zero to all?

# An Elementary Economic Model II

| | Peas | Beans | Carrots |
|---|---|---|---|
| Pea grower uses | 1 | 1 | 2 |
| Bean grower uses | 2 | 3 | 1 |
| Carrot grower uses | 2 | 1 | 2 |

Set $u, v$ and $w$ to be the price (\$) per kilo of peas, beans and carrots respectively. Equating the outlays (left side) and income (right side) of each grower we find:

$$\begin{array}{ccccccc} u & + & v & + & 2w & = & 5u \\ 2u & + & 3v & + & w & = & 5v \\ 2u & + & v & + & 2w & = & 5w \end{array}.$$

This leads to the homogeneous system

$$\begin{array}{ccccccc} -4u & + & v & + & 2w & = & 0 \\ 2u & - & 2v & + & w & = & 0 \\ 2u & + & v & - & 3w & = & 0 \end{array}.$$

# An Elementary Economic Model III

The coefficient matrix reduces to $\begin{bmatrix} 1 & 0 & -5/6 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}$

and so the solution is

$$u = (5/6)w, \quad v = (4/3)w \quad \text{for general } w.$$

Taking $w = 6$ for example yields $u = 5, \ \ v = 8$; *i.e*

| | |
|---|---|
| Peas: | \$5/kg |
| Beans: | \$8/kg |
| Carrots: | \$6/kg |

This is an example of a closed Leontief economy – the production of each grower is completely distributed amongst the growers.

For problems such as this, is there always a non-zero solution?
Indeed, is there a realistic solution, that is, with $u, v, ...$ all positive ?

Analysis shows that the answer, for this closed situation, is YES.

# An Economic Model with Demand I

## Example

A company has two interacting branches, X and Y.

- Half of the total output, by value, of branch X is actually consumed internally by the branch itself.
- Branch X also uses 20 cents worth of branch Y's output for every \$1's worth of its own production.
- For every \$1's worth of its own output, branch Y consumes 60 cents worth of branch X's output.
- Branch Y consumes 40%, by value, of its own output.

What total output, by value, is required from each branch in order to meet an (external) demand of

— \$50 000 worth of X-product and

— \$40 000 worth of Y-product?

# An Economic Model with Demand II

Let $x$ and $y$ be the dollar values of the required total outputs from branches X and Y respectively.

Each of $x$ and $y$ are the sum of the three amounts arising from:

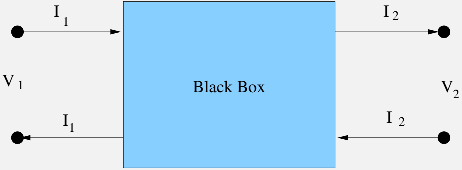
- internal consumption;
- consumption by the other branch;
- external demand.

From the given data:
$$\begin{aligned} x &= (0.5)x + (0.6)y + 50\,000 \\ y &= (0.2)x + (0.4)y + 40\,000 \end{aligned}.$$

This is an example of an *open* Leontief model, because there is surplus production to meet external demand. It leads to the *non*-homogeneous system
$$\begin{bmatrix} 0.5 & -0.6 \\ -0.2 & 0.6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 50\,000 \\ 40\,000 \end{bmatrix}$$

which has unique solution
$$\begin{aligned} x &= \$300\,000.00 \\ y &= \$166\,666.67 \end{aligned}.$$

# Simple Circuits



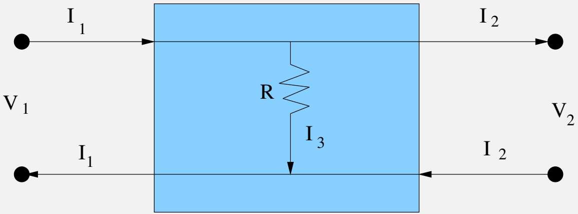

A component in an electrical circuit can be considered as a black box with known transfer function, that is, the transformation between the input and output voltages and currents,

$$T : \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} \mapsto \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}.$$

Often $T$ is a linear transformation, so there is a transfer matrix $\mathbf{A}$ such that

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \mathbf{A} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}.$$

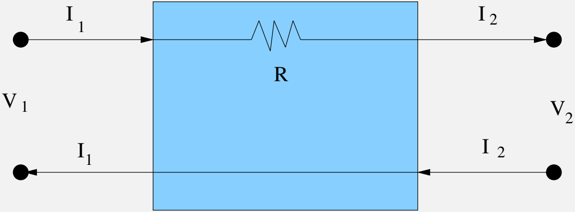
# Simple Shunt




Kirchhoff's and Ohm's laws show that $I_2 + I_3 = I_1$ and $V_1 = V_2 = I_3 R$, whence

$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{R} & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}.$$

# Series Resistor

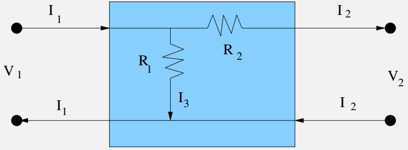



Ohm's laws gives

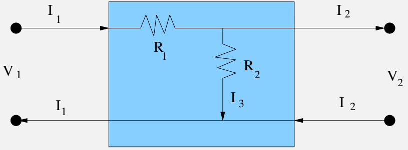
$$\begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 & -R \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}.$$

# Combinations

If we cascade these circuits the resulting transfer matrix will be the product, in the appropriate order, of the individual transfer matrices.
Here are the two possibilities:




$$\begin{bmatrix} 1 & -R_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\frac{1}{R_1} & 1 \end{bmatrix} = \begin{bmatrix} 1 + \frac{R_2}{R_1} & -R_2 \\ -\frac{1}{R_1} & 1 \end{bmatrix}$$

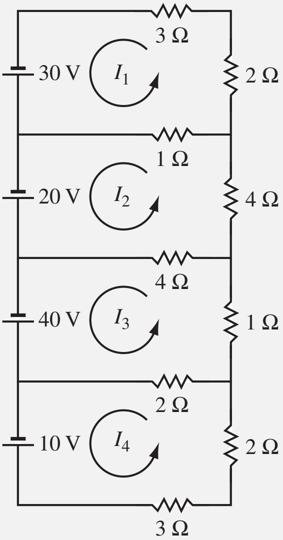



$$\begin{bmatrix} 1 & 0 \\ -\frac{1}{R_2} & 1 \end{bmatrix} \begin{bmatrix} 1 & -R_1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -R_1 \\ -\frac{1}{R_2} & 1 + \frac{R_1}{R_2} \end{bmatrix}$$

So the effect is decidedly different depending on the sequencing.

With more active circuitry like capacitors, inductors etcetera, the details are more complicated (Laplace transforms are needed), but the basic agenda is the same.

Linear algebra can also be used to calculate current flow in various components of a circuit, for example:



# 10. Matrix Operations - (Lay 2.1)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/adjacency.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_1ov.pdf

# Matrix Addition

## Definition

Given two $m \times n$ matrices $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$ and $\mathbf{B} = [\mathbf{b}_1 \ \ldots \ \mathbf{b}_n]$ the **sum** $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is defined by adding corresponding column vectors.

$$[\mathbf{a}_1 \ \ldots \ \mathbf{a}_n] + [\mathbf{b}_1 \ \ldots \ \mathbf{b}_n] = [\mathbf{a}_1 + \mathbf{b}_1 \ \ldots \ \mathbf{a}_n + \mathbf{b}_n]$$

This is just element-wise addition: $c_{ij} = a_{ij} + b_{ij}, \quad 1 \le i \le m, 1 \le j \le n.$

Warning! If $\mathbf{A}, \mathbf{B}$ are not the same size then their sum is not defined.

## Example

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 2 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} 1+0 & 2+1 \\ 3+(-1) & 4+2 \\ 5+(-2) & 6+(-3) \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 3 & 3 \end{bmatrix}$$

# Matrix Scalar Multiplication

## Definition

If $\alpha$ is a scalar (*i.e.* any real number) and $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_n]$ is an $m \times n$ matrix then the **scalar product** $\alpha\mathbf{A}$ is the matrix $\mathbf{C}$ obtained by replacing each column $\mathbf{a}_i$ by the scalar product $\alpha\mathbf{a}_i$.

$$\alpha[\mathbf{a}_1 \ \ldots \ \mathbf{a}_n] = [\alpha\mathbf{a}_1 \ \ldots \ \alpha\mathbf{a}_n].$$

Every element is multiplied by $\alpha$: $c_{ij} = \alpha\, a_{ij}, \quad 1 \le i \le m, 1 \le j \le n.$

## Example

$$\text{If } \ \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \ \text{ then } \ 3\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 9 & 12 \\ 15 & 18 \end{bmatrix}.$$

# Transpose of a Matrix

## Definition

The **transpose** of an $n \times m$ matrix $\mathbf{A}$ is the $m \times n$ matrix whose columns are the rows of $\mathbf{A}$.

We use the notation $\mathbf{A}^T$ to specify the transpose of $\mathbf{A}$.

## Example

If

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 4 & -1 & 7 \\ 0 & 2 & 3 \end{bmatrix}$$

then

$$\mathbf{A}^T = \begin{bmatrix} 3 & -1 & 1 \\ 0 & 2 & 5 \end{bmatrix} \text{ and } \mathbf{B}^T = \begin{bmatrix} 4 & 0 \\ -1 & 2 \\ 7 & 3 \end{bmatrix}.$$

# Arithmetic Properties

- $$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$$
- $$\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$$
- $$\overbrace{\mathbf{A}+\mathbf{A}+\cdots+\mathbf{A}}^{n \text{ times}}=n\mathbf{A}$$
- $$(\alpha+\beta)\mathbf{A}=\alpha\mathbf{A}+\beta\mathbf{A}$$
- $$\alpha(\mathbf{A}+\mathbf{B})=\alpha\mathbf{A}+\alpha\mathbf{B}$$
- $$(\mathbf{A}+\mathbf{B})^T=\mathbf{A}^T+\mathbf{B}^T$$
- $$(\alpha\mathbf{A})^T=\alpha\mathbf{A}^T$$

# Matrix Multiplication I

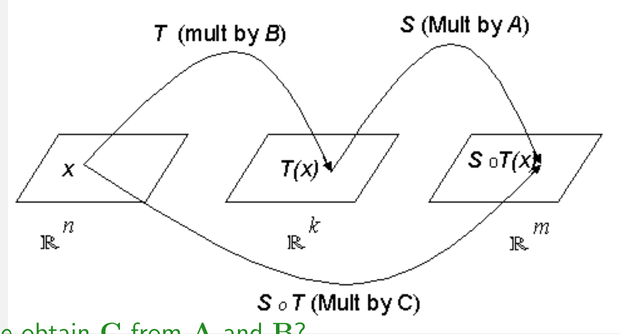
Suppose we have linear transformations

- $T : \mathbb{R}^n \to \mathbb{R}^k$ associated with a $k \times n$ matrix $\mathbf{B}$
- $S : \mathbb{R}^k \to \mathbb{R}^m$ associated with an $m \times k$ matrix $\mathbf{A}$.

Then there will be a composite linear transformation

- $S \circ T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $(S \circ T)(\mathbf{x}) = S(T(\mathbf{x}))$ for every $\mathbf{x} \in \mathbb{R}^n$.

Transformation $(S \circ T)$ must have an associated $m \times n$ matrix, $\mathbf{C}$ say.




How does one obtain $\mathbf{C}$ from $\mathbf{A}$ and $\mathbf{B}$?

# Matrix Multiplication II

Let $\mathbf{b}_1, \ldots, \mathbf{b}_n \in R^k$ be the columns of $\mathbf{B}$. Then

$$T(\mathbf{x}) = \mathbf{B}\mathbf{x} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} \mathbf{x} = x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n.$$

Then, using linearity where appropriate,

$$\begin{aligned}(S \circ T)(\mathbf{x}) = S(T(\mathbf{x})) = S(\mathbf{B}\mathbf{x}) &= \mathbf{A}(\mathbf{B}\mathbf{x}) \\ &= \mathbf{A}(x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n) \\ &= x_1\mathbf{A}\mathbf{b}_1 + \cdots + x_n\mathbf{A}\mathbf{b}_n \\ &= \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \cdots & \mathbf{A}\mathbf{b}_n \end{bmatrix} \mathbf{x}.\end{aligned}$$

Thus

$$\mathbf{C} = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \cdots & \mathbf{A}\mathbf{b}_n \end{bmatrix}.$$

So the columns of $\mathbf{C}$ are obtained by multiplying the columns of $\mathbf{B}$ by $\mathbf{A}$.

## Definition

If $\mathbf{A}$ is an $m \times k$ matrix and $\mathbf{B}$ a $k \times n$ matrix then the **product** of $\mathbf{A}$ and $\mathbf{B}$ (in that order) is defined to be the matrix $\mathbf{C}$ obtained above, *viz.*

$$\mathbf{A}\mathbf{B} = \mathbf{A}\begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \cdots & \mathbf{A}\mathbf{b}_n \end{bmatrix}.$$

# Matrix Multiplication Example

## Example

Calculate $\mathbf{AB}$ for $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 4 & -1 & 7 \\ 0 & 2 & 3 \end{bmatrix}$.

$\mathbf{A}$ is $3 \times 2$, $\mathbf{B}$ is $2 \times 3$, so $\mathbf{AB}$ will be a $3 \times 3$ matrix.

$$\mathbf{AB} = \left[ \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 7 \\ 3 \end{bmatrix} \right]$$

$$= \begin{bmatrix} (3 \times 4) + (0 \times 0) & (3 \times -1) + (0 \times 2) & (3 \times 7) + (0 \times 3) \\ (-1 \times 4) + (2 \times 0) & (-1 \times -1) + (2 \times 2) & (-1 \times 7) + (2 \times 3) \\ (1 \times 4) + (5 \times 0) & (1 \times -1) + (5 \times 2) & (1 \times 7) + (5 \times 3) \end{bmatrix}$$

$$= \begin{bmatrix} 12 & -3 & 21 \\ -4 & 5 & -1 \\ 4 & 9 & 22 \end{bmatrix}.$$

# Direct Calculation of Product Entries

Look at any particular entry in the product matrix $\mathbf{AB}$ in the previous example; say the entry in row 2, column 3:

$$(-1 \times 7) + (2 \times 3) = \begin{bmatrix} -1 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 3 \end{bmatrix} = \text{2nd row of } \mathbf{A} \times \text{3rd column of } \mathbf{B}.$$

The same thing works in general:

For an $m \times k$ matrix $\mathbf{A}$ with $i$-th row $\begin{bmatrix} a_{i1} & \dots & a_{ik} \end{bmatrix}$

and a $k \times n$ matrix $\mathbf{B}$ with $j$-th column $\begin{bmatrix} b_{1j} \\ \vdots \\ b_{kj} \end{bmatrix}$

the $ij$-entry in $\mathbf{AB}$ is given by

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{l=1}^{k} a_{il}b_{lj}.$$

# Remarks on Matrix Multiplication

1. Suppose $\mathbf{A}$ is an $m \times k$ matrix and $\mathbf{B}$ is an $l \times n$ matrix.
   - The product matrix $\mathbf{AB}$ is defined if and only if $k = l$.
   - When defined, $\mathbf{AB}$ is an $m \times n$ matrix.
   - In summary $m \times k$ by $k \times n \to m \times n$.
   - The product matrix $\mathbf{BA}$ is defined if and only if $m = n$.
   - When defined, $\mathbf{BA}$ is an $n \times m$ matrix.
   - Even when they are both defined and have the same shape, $\mathbf{AB}$ and $\mathbf{BA}$ are not normally equal.
2. The sizes really come from the corresponding linear transformations:

$$\overset{k\times n \text{ matrix } \mathbf{B}}{\mathbb{R}^n \xrightarrow{T} \mathbb{R}^k} \quad \text{followed by} \quad \overset{m\times k \text{ matrix } \mathbf{A}}{\mathbb{R}^k \xrightarrow{S} \mathbb{R}^m} \quad \text{produces} \quad \overset{m\times n \text{ matrix } \mathbf{AB}}{\mathbb{R}^n \xrightarrow{S\circ T} \mathbb{R}^m}$$

3. Note the order here: $S \circ T$ means first apply $T$, then $S$.

# Matrix Multiplication Example Revisited

## Example

Compare $\mathbf{AB}$ and $\mathbf{BA}$ when $\mathbf{A} = \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 4 & -1 & 7 \\ 0 & 2 & 3 \end{bmatrix}$.

$\mathbf{A}$ is $3 \times 2$ and $\mathbf{B}$ is $2 \times 3$, so $\mathbf{AB}$ and $\mathbf{BA}$ are both defined but have different shapes; $3 \times 3$ and $2 \times 2$ respectively.

$\mathbf{AB}$ was calculated previously; its value is $\begin{bmatrix} 12 & -3 & 21 \\ -4 & 5 & -1 \\ 4 & 9 & 22 \end{bmatrix}$.

$$\mathbf{BA} = \begin{bmatrix} 4 \times 3 + -1 \times -1 + 7 \times 1 & 4 \times 0 + -1 \times 2 + 7 \times 5 \\ 0 \times 3 + 2 \times -1 + 3 \times 1 & 0 \times 0 + 2 \times 2 + 3 \times 5 \end{bmatrix} = \begin{bmatrix} 20 & 33 \\ 1 & 19 \end{bmatrix}.$$

# Another Matrix Multiplication Example

## Example

For $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 8 & -6 \\ 4 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -3 & 9 & 1 \\ 0 & 2 & 2 \\ 5 & -5 & 4 \end{bmatrix}$ calculate those of the products $\mathbf{AB}$, $\mathbf{BA}$ that are defined.

$\mathbf{A}$ is $3 \times 2$, $\mathbf{B}$ is $3 \times 3$, so $\mathbf{AB}$ is not defined.
$\mathbf{B}$ is $3 \times 3$, $\mathbf{A}$ is $3 \times 2$, so $\mathbf{BA}$ is defined and is $3 \times 2$.

$$\mathbf{BA} = \begin{bmatrix} 70 & -57 \\ 24 & -12 \\ -14 & 35 \end{bmatrix}.$$

# Properties of Matrix Multiplication

Provided the stated operations are defined for matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, we have the following:

Associative law for multiplication

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}.$$

Left distributive law

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}.$$

Right distributive law

$$(\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}.$$

Scalar multiplication

$$r(\mathbf{AB}) = (r\mathbf{A})\mathbf{B} = \mathbf{A}(r\mathbf{B}) \text{ , } (r \in \mathbb{R}).$$

# Remarks on Properties of Matrix Multiplication

1. The associative laws for sums and products enable us to avoid worrying about bracketing in multiple sums and products.

   *E.g.* We can write $\mathbf{A} + \mathbf{BCD} + \mathbf{E}$ without fear of ambiguity.

2. The net effect of all the listed properties of matrix operations is that the usual laws of arithmetic apply *provided* the operations are defined, *except* that multiplication is **not commutative** in general.

# Identity matrix

## Definition

The $n \times n$ matrix with entries $\begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ is denoted $\mathbf{I}_n$ (or just $\mathbf{I}$ if $n$ can be inferred), and is called the $n \times n$ **identity matrix**.

## Illustration

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

## Theorem

*Given any $n \times n$ (square) matrix $\mathbf{A}$;* $\quad \mathbf{I}_n\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$.
*(This justifies the name 'identity' matrix for $\mathbf{I}_n$.)*

# Powers of Matrices

Square matrices are the only matrices that can be multiplied by themselves. (Why?)

If $\mathbf{A}$ is square then $\mathbf{A}^k$ is defined for every natural number $k$:

$$\mathbf{A}^k = \overbrace{\mathbf{A}\mathbf{A}\cdots\mathbf{A}}^{n \text{ copies of } \mathbf{A}} \;; \qquad \mathbf{A}^0 = \mathbf{I}.$$

## Example

Calculate $\mathbf{A}^2$ and $\mathbf{A}^3$ for $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$.

Answer:

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

$$\mathbf{A}^3 = \mathbf{A}^2\mathbf{A} = \mathbf{I}\mathbf{A} = \mathbf{A}.$$

# A Surprise

## Example

Calculate $\mathbf{B}^2$ and $\mathbf{B}^3$ for $\mathbf{B} = \begin{bmatrix} 0 & -1 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

Answer:

$$\mathbf{B}^2 = \mathbf{B}\mathbf{B} = \begin{bmatrix} 0 & -1 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

$$\mathbf{B}^3 = \mathbf{B}^2\mathbf{B} = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So a power of a non-zero matrix can give the zero matrix. This is a phenomenon with no parallel amongst scalars (real numbers).

# 11. The Inverse of A Matrix (Lay 2.2)

Case Study: `http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/dominance.pdf`

Lay's Summary: `http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_2ov.pdf`

# Matrix Inverses

For a real number $x \neq 0$, the multiplicative inverse of $x$ is $1/x$.
Alternatively we can say that the multiplicative inverse of $x$ is the real number $y$ for which $xy = yx = 1$.

We use this alternative definition as a model for defining the (multiplicative) inverse of a matrix:

## Definition

Suppose that matrices $\mathbf{A}$ and $\mathbf{B}$ are such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$.
Then $\mathbf{A}$ is **invertible** or **nonsingular**, with **inverse** $\mathbf{B}$.

## Remarks

- The equation $\mathbf{AB} = \mathbf{BA}$ implies that $\mathbf{A}$ and $\mathbf{B}$ must be square and of the same size.
- No all-zero matrix $\mathbf{0}$ can be invertible since whenever $\mathbf{0B}$ is defined it will also be all-zero and hence cannot be $\mathbf{I}$.
- Unlike the situation with real numbers, there are lots of non-zero matrices, including plenty of square ones, that have no inverse.

# Examples

## Example

Show that $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and $\mathbf{B} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ are inverses of each other.

Answer:

$$\mathbf{AB} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I},$$

$$\mathbf{BA} = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$$

## Example

Show that $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ is not invertible.

Answer: Since $\mathbf{A}$ is $2 \times 2$ any inverse $\mathbf{B}$ must also be $2 \times 2$, say $\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Then $\mathbf{AB} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ 0 & 0 \end{bmatrix} \neq \mathbf{I}.$

# Uniqueness

## Theorem

*A matrix $\mathbf{A}$ can have at most one inverse.*

Proof: Suppose that $\mathbf{B}$ and $\mathbf{C}$ are both inverses of $\mathbf{A}$.
So, in particular, $\mathbf{BA} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{I}$.
Then, using the associative law, we have

$$\mathbf{C} = \mathbf{IC} = (\mathbf{BA})\mathbf{C} = \mathbf{B}(\mathbf{AC}) = \mathbf{BI} = \mathbf{B} \qquad (\text{so } \mathbf{B} = \mathbf{C}).$$

## Definition

The unique inverse of $\mathbf{A}$, if it exists, is denoted by $\mathbf{A}^{-1}$.

## Remarks

- The proof above actually establishes that if $\mathbf{A}$ has both a ‘left’ and a ‘right’ inverse then these are the same so are *the* inverse of $\mathbf{A}$.
- It turns out that in fact the existence of either such ‘one-sided’ inverse is sufficient for existence of *the* inverse, but this is harder to prove.

# Use of Inverses to Solve Matrix Equations

## Theorem

*If $\mathbf{A}$ is invertible, then the equation $\mathbf{Ax} = \mathbf{b}$ has the unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.*

Proof:
(i) $\mathbf{A}^{-1}\mathbf{b}$ is a solution because $\mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = (\mathbf{A}\mathbf{A}^{-1})\mathbf{b} = \mathbf{I}\mathbf{b} = \mathbf{b}$.
(ii) If $\mathbf{y}$ is any solution then $\mathbf{Ay} = \mathbf{b}$.
It follows that $\mathbf{y} = \mathbf{I}\mathbf{y} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{y} = \mathbf{A}^{-1}(\mathbf{Ay}) = \mathbf{A}^{-1}\mathbf{b}$.

## Remarks

The theorem is not of as much practical use as one might imagine because:

- Determining invertibility or otherwise is not an easy task in general.
- Even if $\mathbf{A}$ is known to be invertible, calculating the inverse can be a computationally intense activity. Solving $\mathbf{Ax} = \mathbf{b}$ directly by row operations is generally more efficient, especially for very large systems.

# A Formula for the Inverse of a $2 \times 2$ Matrix

Suppose $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has inverse $\mathbf{B} = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$.

Then $\mathbf{AB} = \mathbf{I}$, leading to the equations $\begin{aligned} ae + bg &= 1 \\ af + bh &= 0 \\ ce + dg &= 0 \\ cf + dh &= 1 \end{aligned}$.

Taking $e, f, g, h$ to be the unknowns, this system has augmented matrix

$$\begin{bmatrix} a & 0 & b & 0 & 1 \\ 0 & a & 0 & b & 0 \\ c & 0 & d & 0 & 0 \\ 0 & c & 0 & d & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{d}{ad-bc} \\ 0 & 1 & 0 & 0 & -\frac{b}{ad-bc} \\ 0 & 0 & 1 & 0 & -\frac{c}{ad-bc} \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{bmatrix}.$$

Reading off the solution values for $e, f, g, h$ gives

$$\mathbf{B} = \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \boxed{\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.}$$

# The Determinant of a 2 by 2 Matrix

The derivation of the formula for $\mathbf{A}^{-1}$ was a bit of a cheat, since it did not check what happens when $ad - bc = 0$. But it turns out that the system is inconsistent in that case, meaning that $\mathbf{A}^{-1}$ does not exist.

So the value of $ad - bc$ *determines* whether $\mathbf{A}$ is invertible. Hence

## Definition

For $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ the quantity $\det(\mathbf{A}) = ad - bc$ is the **determinant** of $\mathbf{A}$.

We have proved

## Theorem

*A $2 \times 2$ matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. In that case*

$$\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{2}$$

# Examples of a $2 \times 2$ inverses

## Example

Use the $2 \times 2$ inverse formula to find the inverses of each of the following:

(a) $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (b) $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ (c) $\mathbf{C} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

In each case comment on the geometrical significance of the answer.

Answer(a): $\mathbf{A}^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} = \frac{1}{1\times1-1\times0}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$.

$\mathbf{A}$ and $\mathbf{A}^{-1}$ represent horizontal shears that nullify each other.

Answer(b): $\mathbf{B}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \frac{1}{0\times0-1\times1}\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{B}$.

$\mathbf{B}$ is its own inverse. It represents reflection in the line $y = x$.

Answer(c): $\mathbf{C}^{-1} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} = \frac{1}{\cos^2\theta+\sin^2\theta}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} =$

$\begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$. $\mathbf{C}$ and $\mathbf{C}^{-1}$ represent rotations that nullify each other.

# Some Further Properties of Matrix Inverses

We need a process that works for any size (square) matrix, and the above approach is clearly hopeless for this. To this end we start with three further simple observations about matrix inverses.

1. If $\mathbf{A}$ has inverse $\mathbf{A}^{-1}$, then $\mathbf{A}^{-1}$ is invertible with inverse $\mathbf{A}$.

   Proof: This is a consequence of symmetry in the definition of inverse.

2. If $\mathbf{A}, \mathbf{B}$ are invertible and of the same size, then $\mathbf{AB}$ is invertible with $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

   Proof:
$$\begin{aligned}\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right)(\mathbf{AB}) &= \mathbf{B}^{-1}\left(\mathbf{A}^{-1}\mathbf{A}\right)\mathbf{B} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{I} \\ (\mathbf{AB})\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right) &= \mathbf{A}\left(\mathbf{BB}^{-1}\right)\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{I}\end{aligned}$$

3. If $\mathbf{A}$ is invertible then $\mathbf{A}^T$ is invertible with inverse $(\mathbf{A}^{-1})^T$.

   Proof:
$$\begin{aligned}\mathbf{AA}^{-1} &= \mathbf{I} = \mathbf{I}^T = (\mathbf{A}^{-1})^T\mathbf{A}^T \\ \mathbf{A}^{-1}\mathbf{A} &= \mathbf{I} = \mathbf{I}^T = \mathbf{A}^T(\mathbf{A}^{-1})^T\end{aligned}$$

# Elementary Matrices

Recall the definition of elementary row operations:

- Add a multiple of one row to another row.
- Interchange two rows.
- Multiply a row by a non-zero constant.

## Definition

A matrix obtained from an identity matrix by a *single* elementary row operation is an **elementary matrix**.

## Illustration

$$\begin{bmatrix}1 & 0 & 0\\ 3 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}, \begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix} \text{ and } \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 5\end{bmatrix} \text{ are elementary matrices.}$$

They correspond to the row operations
$R2 \leftarrow R2 + 3R1, \quad R1 \leftrightarrow R3$ and $R3 \leftarrow 5R3$ respectively.

# Elementary Matrices and Invertibility

The importance of elementary matrices is the following result which enables us to efficiently determine invertibility or otherwise.

## Theorem

*The effect on an $m \times n$-matrix $\mathbf{A}$ of an elementary row operation is the same as first performing the operation on the identity matrix $\mathbf{I}_m$ to get an elementary matrix $\mathbf{E}$, then multiplying $\mathbf{A}$ on the left by $\mathbf{E}$.*

## Remark

A consequence of this theorem is that elementary matrices are necessarily invertible, with inverses also elementary matrices.

# Illustration I

## Illustration

To subtract $2R1$ from $R2$ in the matrix $\begin{bmatrix} 2 & 3 & 2 \\ 1 & 4 & 7 \\ 3 & 5 & 4 \end{bmatrix}$

first perform the operation on $\mathbf{I}_3$: $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Then multiply:

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

# Illustration II

## Illustration

To swap $R1$ with $R3$ in any $3 \times 3$ matrix $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

first perform the operation on $\mathbf{I}_3$: $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

Then multiply:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} g & h & i \\ d & e & f \\ a & b & c \end{bmatrix}.$$

# Row Equivalence

Finding the reduced row echelon form of a matrix consists of applying a suitable series of elementary row operations, that is, repeatedly multiplying on the left by suitable elementary matrices.

The reduced row echelon form $\mathbf{R}$ of a matrix $\mathbf{A}$ can be written

$$\mathbf{R} = \mathbf{E}_k \mathbf{E}_{k-1} \cdots \mathbf{E}_1 \mathbf{A}$$

where the operation corresponding to $\mathbf{E}_1$ is the first applied, then that for $\mathbf{E}_2$, etcetera.

From $\mathbf{R}$ we can reconstruct $\mathbf{A}$ by a similar sequence. Note the change of order of the suffices:

$$\mathbf{A} = \mathbf{E}_1^{-1} \cdots \mathbf{E}_{k-1}^{-1} \mathbf{E}_k^{-1} \mathbf{R}.$$

## Definition

Two matrices that can be transformed into each other by elementary row operations are called **row equivalent**.

# Theorems on Inverses

## Theorem

*Any matrix $\mathbf{A}$ is row equivalent to its reduced row echelon form $\mathbf{R}$.*

Proof: As we have seen, $\mathbf{R} = \mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1\mathbf{A}$ and $\mathbf{A} = \mathbf{E}_1^{-1}\cdots\mathbf{E}_{k-1}^{-1}\mathbf{E}_k^{-1}\mathbf{R}$.

## Theorem

*A square matrix $\mathbf{A}$ is invertible if, and only if, its row reduced echelon form $\mathbf{R}$ is invertible.*

Proof $\Leftarrow$: $\mathbf{A}^{-1} = \left(\mathbf{E}_1^{-1}\cdots\mathbf{E}_{k-1}^{-1}\mathbf{E}_k^{-1}\mathbf{R}.\right)^{-1} = \mathbf{R}^{-1}\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1.$

$\Rightarrow$: $\mathbf{R}^{-1} = \left(\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1\mathbf{A}\right)^{-1} = \mathbf{A}^{-1}\mathbf{E}_1^{-1}\cdots\mathbf{E}_{k-1}^{-1}\mathbf{E}_k^{-1}.$

## Theorem

*An $n \times n$ rref matrix $\mathbf{R}$ is invertible if and only if it is the identity matrix $\mathbf{I}_n$.*

Proof: If $\mathbf{R} = \mathbf{I}_n$ then $\mathbf{R}^{-1} = \mathbf{I}_n^{-1} = \mathbf{I}_n$, so $\mathbf{R}$ is invertible.
Conversely if $\mathbf{R} \neq \mathbf{I}_n$ then $\mathbf{R}$ must have an all-zero last row.
So $\mathbf{RB}$ must have an all-zero last row for every $n \times n$ matrix $\mathbf{B}$.
Hence no $\mathbf{B}$ is the inverse of $\mathbf{R}$; *i.e.* $\mathbf{R}$ is **not** invertible.

# Construction of $n \times n$ Inverses

From the theorems, the row reduced echelon form $\mathbf{R}$ of an invertible $n \times n$ matrix $\mathbf{A}$ must be $I_n$.
So there must be elementary matrices $\mathbf{E}_1, \ldots, \mathbf{E}_k$ such that

$$\begin{aligned} \mathbf{A}^{-1} &= \mathbf{R}^{-1}\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1 = \mathbf{I}_n^{-1}\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1 \\ &= \mathbf{I}_n\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1 = \mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1 = \mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_1\mathbf{I}_n. \end{aligned}$$

The final right-hand side here is what is obtained by applying *to* $\mathbf{I}_n$ exactly the elementary row operations which, applied to $\mathbf{A}$, *produce* $\mathbf{I}_n$.

Thus, if $\mathbf{A}$ is invertible, the row reduction algorithm applied to the augmented matrix $[\mathbf{A}\ \mathbf{I}_n]$ will produce $\left[\mathbf{I}_n\ \mathbf{A}^{-1}\right]$.

If $\mathbf{A}$ is not invertible, the row reduction algorithm applied to $[\mathbf{A}\ \mathbf{I}_n]$ will result in an all-zero last row in the first half of the reduced matrix.

## Row Reduction Algorithm for Matrix Inverses

For any $n \times n$ matrix $\mathbf{A}$ reduce $[\mathbf{A}\ \mathbf{I}_n] \xrightarrow{\text{rref}} [\mathbf{R}\ \mathbf{B}]$.
If $\mathbf{R} = \mathbf{I}_n$ then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1} = \mathbf{B}$; otherwise $\mathbf{A}$ is not invertible.

# Examples

## Example

Let $\mathbf{A} = \begin{bmatrix} -1 & 2 & -3 \\ 2 & 1 & 0 \\ 4 & -2 & 5 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -2 & 0 \\ 3 & 0 & 1 \end{bmatrix}$. Find inverses where they exist.

$$[\mathbf{A}\ \mathbf{I}_3] = \begin{bmatrix} -1 & 2 & -3 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 4 & -2 & 5 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & -5 & 4 & -3 \\ 0 & 1 & 0 & 10 & -7 & 6 \\ 0 & 0 & 1 & 8 & -6 & 5 \end{bmatrix}.$$

So $\mathbf{A}$ is invertible and $\mathbf{A}^{-1} = \begin{bmatrix} -5 & 4 & -3 \\ 10 & -7 & 6 \\ 8 & -6 & 5 \end{bmatrix}$.

$$[\mathbf{B}\ \mathbf{I}_3] = \begin{bmatrix} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & -2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1 & 1/3 & 1/3 & -1/6 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}.$$

So $\mathbf{B}$ is not invertible.

# 12. Characterisation of the Invertibility of Matrix (Lay 2.3)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/condition.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_3ov.pdf

# Some Characterisations of Matrix Invertibility

As we have seen, by definition, a (square) matrix $\mathbf{A}$ is **invertible** if it has an inverse, *i.e.* if there is a matrix $\mathbf{B}$ for which $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$.

The purpose of this section is to establish six alternative ways (numbered 1-6 in what follows) to define, or *characterise*, invertibility.

For convenience, the task will be spread between three theorems, labeled A, B and C. Here is the first:

**Theorem (A: Invertibility Characterisations 1 and 2)**

*For an $n \times n$ matrix $\mathbf{A}$ the following are equivalent:*

0. $\mathbf{A}$ *is invertible.*
1. $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution. (The columns of $\mathbf{A}$ are linearly independent.)*
2. $\mathbf{A}$ *is row equivalent to $\mathbf{I}_n$. (The reduced row echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.)*

# Proof of Theorem A: $(0) \Rightarrow (1) \Rightarrow (2) \Rightarrow (0)$

If $\mathbf{A}$ is invertible then $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

Proof: Suppose $\mathbf{A}$ is invertible (so $\mathbf{A}^{-1}$ exists) and that $\mathbf{Ax} = \mathbf{0}$.
Multiplying both sides on the left by $\mathbf{A}^{-1}$ yields $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{0}$. So

$$\mathbf{x} = \mathbf{Ix} = \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0}.$$

That is, the solution $\mathbf{x}$ is trivial.

If $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution then $\mathbf{A}$ is row equivalent to $\mathbf{I}$.

Proof: If $\mathbf{Ax} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$ then there are no free variables and no inconsistency. This implies that $[\mathbf{A}\ \mathbf{0}] \xrightarrow{\text{rref}} [\mathbf{I}\ \mathbf{0}]$.
Omitting the last column shows that $\mathbf{A}$ is row equivalent to $\mathbf{I}$.

If $\mathbf{A}$ is row equivalent to $\mathbf{I}$ then $\mathbf{A}$ is invertible.

Proof: This was proved in the previous section, using elementary matrices.

# Example

Let $\mathbf{A} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 \\ 4 & 3 & 2 & 0 \\ 4 & 3 & 2 & 1 \end{bmatrix}$.

(a) Without first calculating an inverse, show that $\mathbf{A}$ is invertible.

(b) Calculate $\mathbf{A}^{-1}$.

(a) Consider the equation $\mathbf{Ax} = \mathbf{0}$. The first row gives $4\mathbf{x}_1 = 0$, so $\mathbf{x}_1 = 0$.
The second row gives $4\mathbf{x}_1 + 3\mathbf{x}_2 = 0$ so $\mathbf{x}_2 = 0$.
Continuing, we find $\mathbf{x}_3 = \mathbf{x}_4 = 0$, so $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
Hence $\mathbf{A}$ is invertible by Theorem A.

(b) Using the row reduction method from the last section, $[\mathbf{A}\ \mathbf{I}_4] =$

$$\begin{bmatrix} 4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 1 & 0 & 0 \\ 4 & 3 & 2 & 0 & 0 & 0 & 1 & 0 \\ 4 & 3 & 2 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{rref}} \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -\frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$= \left[\mathbf{I}_4\ \mathbf{A}^{-1}\right]$, from which we read off $\mathbf{A}^{-1}$.

# Triangular Matrices

## Definition

A square matrix is called **lower triangular** when all the entries *above* its diagonal are zero, and called **upper triangular** when all the entries *below* its diagonal are zero.

Both the matrix $\mathbf{A}$ of the last example, and its inverse, are lower triangular.

A study of that example should encourage you to believe the following theorem, whose proof is omitted.

## Theorem

- *A triangular matrix is invertible if and only if there are no zero entries on the diagonal.*
- *The inverse of an invertible lower triangular matrix is lower triangular.*
- *The inverse of an invertible upper triangular matrix is upper triangular.*

# One-Sided Inverses

## Theorem (B: Invertibility Characterisations 3 and 4)

*For an $n \times n$ matrix $\mathbf{A}$ the following are equivalent:*

0. $\mathbf{A}$ *is invertible.*
3. *There is an $n \times n$ matrix $\mathbf{B}$ with $\mathbf{BA} = \mathbf{I}_n$.*
4. *There is an $n \times n$ matrix $\mathbf{C}$ with $\mathbf{AC} = \mathbf{I}_n$.*

The proof is on the next slide. First we deal with 'one-sided' inverses.

## Definition

A matrix $\mathbf{B}$ with property (3) of Theorem B is called a left inverse of $\mathbf{A}$.
A matrix $\mathbf{C}$ with property (4) of Theorem B is called a right inverse of $\mathbf{A}$.

Any one-sided inverse of a square matrix $\mathbf{A}$ is a genuine inverse of $\mathbf{A}$.

Proof: From Theorem B if $\mathbf{A}$ has a left inverse $\mathbf{B}$ then it also has a right inverse $\mathbf{C}$ and vice-versa. These one-sided inverses $\mathbf{B}$ and $\mathbf{C}$ must be equal because $\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C}$.
So $\mathbf{B} = \mathbf{C}$ is both a left and a right inverse; *i.e.* it is an inverse.

# Proof of Theorem B: $(0) \Rightarrow (3) \Rightarrow (0) \Rightarrow (4) \Rightarrow (0)$

If $n \times n$ matrix $\mathbf{A}$ is invertible then there is an $n \times n$ matrix $\mathbf{B}$ with $\mathbf{BA} = \mathbf{I}_n$.

Proof: Take $\mathbf{B} = \mathbf{A}^{-1}$.

If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices with $\mathbf{BA} = \mathbf{I}_n$ then $\mathbf{A}$ is invertible.

Proof: By Theorem A it suffices to show that $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution. If $\mathbf{x}$ is *any* solution, $\mathbf{x} = \mathbf{Ix} = (\mathbf{BA})\mathbf{x} = \mathbf{B}(\mathbf{Ax}) = \mathbf{B0} = \mathbf{0}$.
That is, $\mathbf{x}$ is a trivial solution, as required.

If $n \times n$ matrix $\mathbf{A}$ is invertible then there is a $n \times n$ matrix $\mathbf{C}$ with $\mathbf{AC} = \mathbf{I}_n$.

Proof: Take $\mathbf{C} = \mathbf{A}^{-1}$.

If $\mathbf{A}$ and $\mathbf{C}$ are $n \times n$ matrices with $\mathbf{AC} = \mathbf{I}_n$ then $\mathbf{A}$ is invertible.

Proof: Taking transposes gives $\mathbf{C}^T\mathbf{A}^T = \mathbf{I}^T = \mathbf{I}$.
By $(3) \Rightarrow (0)$ this implies that $\mathbf{A}^T$ is invertible.
But $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$, implying that $\mathbf{A}$ is invertible.

# Invertibility and Solutions to $\mathbf{Ax} = \mathbf{b}$

Our final characterisation of invertibility concerns solving non-homogeneous linear systems with as many equations as unknowns.

**Theorem (C: Invertibility Characterisation 5)**

*For an $n \times n$ matrix $\mathbf{A}$ the following are equivalent:*

0. $\mathbf{A}$ *is invertible.*
5. $\mathbf{Ax} = \mathbf{b}$ *has a unique solution for each* $\mathbf{b} \in \mathbb{R}^n$.

The proof is on the next two slides. First an immediate implication:

If $\mathbf{A}$ is $n \times n$ matrix then:

- If $\mathbf{A}$ is not invertible, then $\exists \mathbf{b} \in \mathbb{R}^n$ for which $\mathbf{Ax} = \mathbf{b}$ has no solution.

# Proof of Theorem C: $(0) \Rightarrow (5)$

On this slide and the next $\mathbf{A}$ is always an $n \times n$ matrix.

If $\mathbf{A}$ is invertible then $\mathbf{Ax} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.

Proof: In fact, given *any* $\mathbf{b} \in \mathbb{R}^n$ we proved in the previous section that the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is unique.

# Proof of Theorem C: $(5) \Rightarrow (0)$

If $\mathbf{Ax} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$ then $\mathbf{A}$ is invertible.

Proof: First recall the definition of matrix multiplication:
Given an $n \times n$ matrix $\mathbf{X}$, with columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the column matrices $\mathbf{Ax}_1, \ldots, \mathbf{Ax}_n$ are exactly the columns of $\mathbf{AX}$.

Now let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the columns of $\mathbf{I}_n$.
Since $\mathbf{Ax} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$, in particular there is a unique vector $\mathbf{x}_j \in \mathbb{R}^n$ such that $\mathbf{Ax}_j = \mathbf{e}_j$ for each $\mathbf{e}_j,\ 1 \leq j \leq n$.

Setting $\mathbf{X} = [\mathbf{x}_1 \ \ldots \ \mathbf{x}_n]$ we get

$$\mathbf{AX} = [\mathbf{Ax}_1 \ \ldots \ \mathbf{Ax}_n] = [\mathbf{e}_1 \ \ldots \ \mathbf{e}_n] = \mathbf{I}_n.$$

This means that $\mathbf{X}$ is a right inverse of $\mathbf{A}$.
But by the one-sided inverse properties from Theorem B this in turn means that $\mathbf{X} = \mathbf{A}^{-1}$, and so $\mathbf{A}$ is invertible.

# 13. Partition Matrices (Lay 2.4)
# Matrix Factorisations (Lay 2.5)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/luqr.pdf

Lay's Summaries: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_4ov.pdf

http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_5ov.pdf

# Partitioned Matrices

A matrix can be partitioned into sub-matrices or blocks in many ways. We will now look at some ways of working with these.
We have already partitioned matrices into columns, and we can use rows in a similar fashion.

$$\mathbf{A} = \begin{bmatrix} \mathsf{col}_1(\mathbf{A}) & \mathsf{col}_2(\mathbf{A}) & \cdots & \mathsf{col}_n(\mathbf{A}) \end{bmatrix} = \begin{bmatrix} \mathsf{row}_1(\mathbf{A}) \\ \mathsf{row}_2(\mathbf{A}) \\ \vdots \\ \mathsf{row}_m(\mathbf{A}) \end{bmatrix}$$

With $\mathbf{b}_j = \mathsf{col}_j(\mathbf{B})$, $j = 1, \ldots, n$, the definition of matrix multiplication gives

$$\mathbf{AB} = \mathbf{A}\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \ldots & \mathbf{Ab}_n \end{bmatrix}.$$

It can also be verified that

$$\mathbf{AB} = \begin{bmatrix} \mathsf{row}_1(\mathbf{A}) \\ \mathsf{row}_2(\mathbf{A}) \\ \vdots \\ \mathsf{row}_n(\mathbf{A}) \end{bmatrix} \mathbf{B} = \begin{bmatrix} \mathsf{row}_1(\mathbf{A})\mathbf{B} \\ \mathsf{row}_2(\mathbf{A})\mathbf{B} \\ \vdots \\ \mathsf{row}_n(\mathbf{A})\mathbf{B} \end{bmatrix}.$$

# General Partitions

More generally an $m \times n$ matrix $\mathbf{A}$ can be partitioned into an $r \times s$ matrix of submatrices (or 'blocks') $\mathbf{A}_{ij}$, $i = 1, \ldots, r \leq m$, $j = 1, \ldots, s \leq n$.

## Illustration

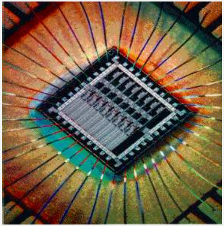
The $3 \times 6$ matrix $\mathbf{A}$ below has been partitioned into a $2 \times 3$ 'block matrix':

$$\mathbf{A} = \left[\begin{array}{cc|cc|cc} 3 & 0 & -1 & 5 & 9 & -2 \\ -5 & 2 & 4 & 0 & -3 & 1 \\ \hline -8 & -6 & 3 & 1 & 7 & -4 \end{array}\right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{bmatrix},$$

where $\mathbf{A}_{11} = \begin{bmatrix} 3 & 0 \\ -5 & 2 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} -1 & 5 \\ 4 & 0 \end{bmatrix}$, $\mathbf{A}_{13} = \begin{bmatrix} 9 & -2 \\ -3 & 1 \end{bmatrix}$

$\mathbf{A}_{21} = \begin{bmatrix} -8 & -6 \end{bmatrix}$, $\mathbf{A}_{22} = \begin{bmatrix} 3 & 1 \end{bmatrix}$, $\mathbf{A}_{23} = \begin{bmatrix} 7 & -4 \end{bmatrix}$.

**Note:** Block structures often match an underlying physical system. Typical systems are integrated circuits in which the different blocks are associated with different functional components of the integrated circuit (memory, arithmetic units . . . ) .



# Multiplication of Partitioned Matrices

Multiplication of partitioned matrices works just like ordinary matrix multiplication.

## Illustration

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}.$$

If the sizes of the sub matrices 'match up', we have

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 \\ \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 \end{bmatrix}.$$

Care is needed when checking the 'match up' conditions. For example if $\mathbf{A}$ and $\mathbf{B}$ are partitioned as $\mathbf{A} = \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 2 & 1 & 1 \\ \hline 0 & 1 & 2 \end{array}\right]$, $\mathbf{B} = \left[\begin{array}{cc} 2 & 1 \\ \hline 0 & 1 \\ 2 & 2 \end{array}\right]$ then even though $\mathbf{A}$ and $\mathbf{B}$ have the same block patterns as in the illustration above, $\mathbf{A}_{11}$ and $\mathbf{B}_1$ do not match up because $\mathbf{A}_{11}\mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}$

is undefined.

# Multiplication Example

Calculate $\mathbf{AB}$ where $\mathbf{A} = \left[\begin{array}{cc|c} 1 & 0 & 3 \\ 2 & 1 & 1 \\ \hline 0 & 1 & 2 \end{array}\right]$ and $\mathbf{B} = \left[\begin{array}{cc} 2 & 1 \\ 0 & 1 \\ \hline 2 & 2 \end{array}\right]$.

Set $\mathbf{A}_{11} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$, $\mathbf{B}_2 = \begin{bmatrix} 2 & 2 \end{bmatrix}$,

$\mathbf{A}_{21} = \begin{bmatrix} 0 & 1 \end{bmatrix}$, $\mathbf{A}_{22} = \begin{bmatrix} 2 \end{bmatrix}$.

Then we have $\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_1 + \mathbf{A}_{12}\mathbf{B}_2 \\ \mathbf{A}_{21}\mathbf{B}_1 + \mathbf{A}_{22}\mathbf{B}_2 \end{bmatrix}$, with

$$\mathbf{A}_{11}\mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \qquad \mathbf{A}_{12}\mathbf{B}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 2 & 2 \end{bmatrix},$$

$$\mathbf{A}_{21}\mathbf{B}_1 = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \end{bmatrix}, \qquad \mathbf{A}_{22}\mathbf{B}_2 = \begin{bmatrix} 2 \end{bmatrix} \begin{bmatrix} 2 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \end{bmatrix}.$$

$$\text{Therefore } \mathbf{AB} = \begin{bmatrix} \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix} + \begin{bmatrix} 6 & 6 \\ 2 & 2 \end{bmatrix} \\ \begin{bmatrix} 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 4 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 8 & 7 \\ 6 & 5 \\ 4 & 5 \end{bmatrix}.$$

# Column – Row Expansion

An interesting application of this technique is the decomposition of the product $\mathbf{AB}$ into a sum of matrices.
In particular for an $n$ by $k$ matrix $\mathbf{A}$ and a $k$ by $m$ matrix $\mathbf{B}$ we have

$$\mathbf{AB} = \begin{bmatrix} \mathrm{col}_1(\mathbf{A}) & \cdots & \mathrm{col}_k(\mathbf{A}) \end{bmatrix} \begin{bmatrix} \mathrm{row}_1(\mathbf{B}) \\ \vdots \\ \mathrm{row}_k(\mathbf{B}) \end{bmatrix} .$$

$$= \mathrm{col}_1(\mathbf{A})\mathrm{row}_1(\mathbf{B}) + \cdots + \mathrm{col}_k(\mathbf{A})\mathrm{row}_k(\mathbf{B})$$

## Definition

Each term $\mathrm{col}_j(\mathbf{A})\mathrm{row}_j(\mathbf{B})$ is a full matrix and is known as an **outer product** of two vectors.

This column-row expansion is particularly useful if the columns and rows of $\mathbf{A}$ and $\mathbf{B}$ correspond to data which is obtained sequentially. We
can create and update the product of the matrices as the data comes in.

# Elimination Method

Consider solving the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ partitioned as

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}. \quad \textit{i.e.} \quad \begin{aligned} \mathbf{A}_{11}\mathbf{X}_1 + \mathbf{A}_{12}\mathbf{X}_2 &= \mathbf{B}_1 \qquad (1) \\ \mathbf{A}_{21}\mathbf{X}_1 + \mathbf{A}_{22}\mathbf{X}_2 &= \mathbf{B}_2 \qquad (2) \end{aligned}$$

Suppose we know that $\mathbf{A}_{11}^{-1}$ exists. In this case we can eliminate $\mathbf{X}_1$ from equation (2) as follows:

Subtract $\mathbf{A}_{21}\mathbf{A}_{11}^{-1}(1)$ from (2) to yield an equation for $\mathbf{X}_2$, namely

$$\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)\mathbf{X}_2 = \mathbf{B}_2 - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{B}_1.$$

This is a smaller system of equations and we need only solve for $\mathbf{X}_2$, while $\mathbf{X}_1$ can be calculated as $\mathbf{X}_1 = \mathbf{A}_{11}^{-1}(\mathbf{B}_1 - \mathbf{A}_{12}\mathbf{X}_2)$.

Often the blocks correspond to very large matrices, and so halving the size of the problem may improve the solution time by an order of magnitude.

# The LU Factorisation

Recall from Section 12 that a square matrix is lower triangular when all entries above the diagonal are zero, and upper triangular when all entries below the diagonal are zero.

The definition can be extended to non-square matrices by the convention that "diagonal" entries are those of the form $a_{ii}$; *i.e.* entries whose row number is the same as the column number.

An important special case of upper triangular is row echelon form.

## Definition

An $m \times n$ matrix $\mathbf{A}$ is said to have an **LU Factorization** if there exist matrices $\mathbf{L}$ and $\mathbf{U}$ such that $\mathbf{A} = \mathbf{L}\mathbf{U}$ where

- $\mathbf{L}$ is an $m \times m$ lower triangular matrix with 1's on the diagonal. (In this case $\mathbf{L}$ is called a unit lower triangular matrix.)

  For example, when $m = 3$, $\mathbf{L}$ has the the form $\begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & * & 1 \end{bmatrix}$.
- $\mathbf{U}$ is (an upper triangular) row echelon form for $\mathbf{A}$.

# Solving Equations using an LU factorisation

If a square matrix $\mathbf{A}$ has an LU factorisation $\mathbf{A} = \mathbf{LU}$ then an equation $\mathbf{Ax} = \mathbf{b}$, can be solved by first solving $\mathbf{Ly} = \mathbf{b}$ and then solving $\mathbf{Ux} = \mathbf{y}$.

The solution to each these equations is achieved by simple substitution; forward then backward.

## Example

Use this LU factorisation

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

to solve $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{b} = \begin{bmatrix} -9 \\ 5 \\ 7 \\ 11 \end{bmatrix}$.

# Example: Solving $\mathbf{Ly} = \mathbf{b}$

The augmented matrix for $\mathbf{Ly} = \mathbf{b}$ is $\begin{bmatrix} 1 & 0 & 0 & 0 & -9 \\ -1 & 1 & 0 & 0 & 5 \\ 2 & -5 & 1 & 0 & 7 \\ -3 & 8 & 3 & 1 & 11 \end{bmatrix}$.

Starting at the top and working down (forward substitution) gives:

$$\begin{aligned}
y_1 &= -9 \\
-y_1 + y_2 &= 5 \\
&\text{so } y_2 = 5 + y_1 \\
&\qquad\;\; = 5 + (-9) = -4 \\
2y_1 - 5y_2 + y_3 &= 7 \\
&\text{so } y_3 = 7 - 2y_1 + 5y_2 \\
&\qquad\;\; = 7 - 2(-9) + 5(-4) = 5 \\
-3y_1 + 8y_2 + 3y_3 + y_4 &= 11 \\
&\text{so } y_4 = 11 + 3(y_1) - 8(y_2) - 3(y_3) \\
&\qquad\;\; = 11 + 3(-9) - 8(-4) - 3(5) = 1
\end{aligned}$$

# Example: Solving $\mathbf{Ux} = \mathbf{y}$

The augmented matrix for $\mathbf{Ux} = \mathbf{y}$ is $\begin{bmatrix} 3 & -7 & -2 & 2 & -9 \\ 0 & -2 & -1 & 2 & -4 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$.

Starting at the bottom and working up (back substitution) gives:

$$\begin{aligned}
-x_4 &= 1 \text{ so } x_4 = -1 \\
-x_3 + x_4 &= 5 \\
&\quad \text{so } x_3 = -5 + x_4 \\
&\qquad\quad = -5 + (-1) = -6 \\
-2x_2 - x_3 + 2x_4 &= -4 \\
&\quad \text{so } x_2 = \tfrac{1}{2}(4 - x_3 + 2x_4) \\
&\qquad\quad = \tfrac{1}{2}\Big(4 - (-6) + 2(-1)\Big) = 4 \\
3x_1 - 7x_2 - 2x_3 + 2x_4 &= -9 \\
&\quad \text{so } x_1 = \tfrac{1}{3}(-9 + 7x_2 + 2x_3 - 2x_4) \\
&\qquad\quad = \tfrac{1}{3}\Big(-9 + 7(4) + 2(-6) - 2(-1)\Big) = 3
\end{aligned}$$

# Factorising $\mathbf{A}$ as $\mathbf{LU}$

We saw in Section 11 that applying row operations to $\mathbf{A}$ is equivalent to multiplying $\mathbf{A}$ by elementary matrices.In particular, if $\mathbf{U}$ is an echelon form for $\mathbf{A}$ then there are elementary matrices $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_p$ such that

$$\mathbf{E}_p \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{U}$$

then

$$\mathbf{A} = (\mathbf{E}_p \cdots \mathbf{E}_1)^{-1}\mathbf{U} = \mathbf{LU} \text{ with } \mathbf{L} = (\mathbf{E}_p \cdots \mathbf{E}_1)^{-1}.$$

Observe that

$$\mathbf{E}_p \cdots \mathbf{E}_1 \mathbf{L} = (\mathbf{E}_p \cdots \mathbf{E}_1)(\mathbf{E}_p \cdots \mathbf{E}_1)^{-1} = \mathbf{I},$$

so the row operations represented by $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_p$ reduce $\mathbf{L}$ to $\mathbf{I}$. Hence:

## Algorithm for LU Factorisation

1. Reduce $\mathbf{A}$ to an echelon form $\mathbf{U}$ by a sequence of only type 1 (row replacement) ops. (If impossible, $\mathbf{A}$ doesn't have an LU factorisation.)
2. Place entries in $\mathbf{L}$ such that the same sequence of row operations reduces $\mathbf{L}$ to $\mathbf{I}$. The off-diagonal entries of $\mathbf{L}$ are the negatives of the corresponding entries in the relevant elementary matrices.

# LU Factorisation Example

## Example

Find the LU factorisation of $\mathbf{A} = \begin{bmatrix} 2 & 4 & -1 & 5 \\ -4 & -5 & 3 & -8 \\ 2 & -5 & -4 & 1 \\ -6 & 0 & 7 & -3 \end{bmatrix}$.

$$\begin{bmatrix} 2 & 4 & -1 & 5 \\ -4 & -5 & 3 & -8 \\ 2 & -5 & -4 & 1 \\ -6 & 0 & 7 & -3 \end{bmatrix} \to \begin{bmatrix} 2 & 4 & -1 & 5 \\ 0 & 3 & 1 & 2 \\ 0 & -9 & -3 & -4 \\ 0 & 12 & 4 & 12 \end{bmatrix} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & & 1 & 0 \\ -3 & & & 1 \end{bmatrix};$$

$$\to \begin{bmatrix} 2 & 4 & -1 & 5 \\ 0 & 3 & 1 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 4 \end{bmatrix} \to \begin{bmatrix} 2 & 4 & -1 & 5 \\ 0 & 3 & 1 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ -3 & 4 & 2 & 1 \end{bmatrix}.$$

# LU Factorisation in Practice

1. To solve a sequence of matrix equations

$$\mathbf{Ax} = \mathbf{b}_1,\ \mathbf{Ax} = \mathbf{b}_2,\ \ldots\ \mathbf{Ax} = \mathbf{b}_p$$

   the standard way is to
   - first solve $\mathbf{Ax} = \mathbf{b}_1$ by row reduction, generating the LU factorisation of $\mathbf{A}$ in the process (as in the previous example),
   - then solve all of the other equations by forward-backward substitution using the LU factorisation of $\mathbf{A}$ (as in the earlier example).
2. The above method is more efficient, and often more accurate, than first calculating $\mathbf{A}^{-1}$ and then calculating the products $\mathbf{A}^{-1}\mathbf{b}_1, \ldots, \mathbf{A}^{-1}\mathbf{b}_p$.
3. Even when $\mathbf{A}$ is invertible, it does not necessarily **have** an LU factorisation because row swaps (type 3 row ops) might be necessary to reduce it to echelon form.
   However a modified LU method can still be used in this case.
   See Lay p127(4ed) or p146(3ed), but this topic is not part of MATH1013.

# 14. Subspaces (Lay 2.8)

Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c02/pdf_files/sec2_8ov.pdf

# Definition: Subspace

## Definition

A **subspace** of $\mathbb{R}^n$ is a subset $H$ of $\mathbb{R}^n$ that has the following properties

1. $\mathbf{0} \in H$.
2. For any $\mathbf{u}, \mathbf{v} \in H$, the sum $\mathbf{u} + \mathbf{v} \in H$.
3. For any $\mathbf{u} \in H$, and $c \in \mathbb{R}$, $c\mathbf{u} \in H$.

## Remarks

1. Condition (1) ensures that $H$ is non-empty.
2. Conditions (2) and (3) are often jointly expressed as '$H$ is closed under the operations of vector addition and scalar multiplication'.
3. The idea behind the subspace concept is 'space within a space'.
4. Lines, planes and hyperplanes in $\mathbb{R}^n$ are subspaces if and only if they satisfy condition (1); *i.e* if and only if they contain the origin. See next slide.

# Examples of Subspaces

1. The set Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a subspace of $\mathbb{R}^n$ for any collection of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$.
   Proof: An arbitrary member of $S =$Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ has the form $\left(\sum_{i=1}^k \alpha_i \mathbf{v}_i\right)$. So:
   (1) $\mathbf{0} = \left(\sum_{i=1}^k 0\mathbf{v}_i\right) \in S.$
   (2) $\left(\sum_{i=1}^k \alpha_i \mathbf{v}_i\right) + \left(\sum_{i=1}^k \beta_i \mathbf{v}_i\right) = \left(\sum_{i=1}^k (\alpha_i + \beta_i)\mathbf{v}_i\right) \in S.$
   (3) $c\left(\sum_{i=1}^k \alpha_i \mathbf{v}_i\right) = \left(\sum_{i=1}^k c\alpha_i \mathbf{v}_i\right) \in S.$
2. Lines, planes and hyperplanes through the origin are subspaces of $\mathbb{R}^n$.
   Proof: Any such linear object has a parametric equation of the the form $\mathbf{x} = \sum_{i=1}^k p_i \mathbf{v}_i$ and so the set of all vectors $\mathbf{x}$ in the object is Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.
3. Linear objects which do **not** go through the origin are **not** subspaces because they fail condition (1) of the definition.

# The Two Fundamental Subspaces of a Matrix

## Definition

The **column space** of a matrix $\mathbf{A}$ is the subspace Col $\mathbf{A} = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ of all linear combinations of the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of $\mathbf{A}$.

## Definition

The **null space** of a matrix $\mathbf{A}$ is the subspace Nul$(\mathbf{A})$ of all solutions of the homogeneous equation $\mathbf{Ax} = \mathbf{0}$.

## Remarks

Suppose $\mathbf{A}$ is an $m \times n$ matrix. Then

1. Col $\mathbf{A}$ is a subspace because it is a span set by definition.
2. Col $\mathbf{A} \subseteq \mathbb{R}^m$ because the columns of $\mathbf{A}$ are vectors in $\mathbb{R}^m$.
3. Nul $\mathbf{A}$ is a subspace because it is a hyperplane.
4. Nul $\mathbf{A} \subseteq \mathbb{R}^n$ because solutions $\mathbf{Ax} = \mathbf{b}$ are vectors in $\mathbb{R}^n$.

# Basis for a Subspace

## Definition

A **basis** for a subspace $H$ of $\mathbb{R}^n$ is a linearly independent subset $B$ of $H$ that spans $H$.

## Remarks

1. In other words, if $H = \operatorname{Span} B$, then we call $B$ a basis for $H$ if, and only if, $B$ is also a linearly independent set of vectors.
2. Recall that a set of vectors $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is linearly independent if, and only if, the homogeneous equation $\mathbf{Bx} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$, where $\mathbf{B}$ is the matrix whose columns are the members of $B$; *i.e.* $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \ldots\ \mathbf{b}_n]$.

# Examples of Bases

## Example

Show that $B = \left\{ \mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$.

Let $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2]$. Then $\mathbf{B} \xrightarrow{\text{rref}} \mathbf{I}_2$. So:
$B$ spans $\mathbb{R}^2$ because $\mathbf{B}$ has a pivot position in every row.
$B$ is lin.ind. because $\mathbf{B}$ has a pivot position in every column.

## Illustration: The Standard Basis for $\mathbb{R}^n$

The subset $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$ given by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

is a linearly independent set that spans $\mathbb{R}^n$, and so is a basis for it, known as the **standard basis** for $\mathbb{R}^n$.

# Finding a Basis for a Null Space

**Example**

Find a basis for the null space of the matrix
$$\mathbf{A} = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}.$$

We need to first find the solution set of $\mathbf{Ax} = \mathbf{0}$:

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

$$\longrightarrow \begin{array}{rcl} x_1 & = & 2x_2 + x_4 - 3x_5 \\ x_2 & = & x_2 \text{ (free)} \\ x_3 & = & -2x_4 + 2x_5 \\ x_4 & = & x_4 \text{ (free)} \\ x_5 & = & x_5 \text{ (free)} \end{array} \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \quad \text{(say)}.$$

The set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is independent and spans Nul $\mathbf{A}$.
So $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a basis for Nul $\mathbf{A}$.

# Finding a Basis for a Column Space

## Theorem

*Suppose* $\mathbf{R}$ *is an echelon form for* $\mathbf{A}$*. Then:*

1. *The columns of* $\mathbf{R}$ *have exactly the same linear relationships as the columns of* $\mathbf{A}$.
2. *The set of pivot columns of* $\mathbf{A}$ *is a basis for Col* $\mathbf{A}$.

## Illustration

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

$\mathbf{c}_2 = -2\mathbf{c}_1$ $\quad \mathbf{c}_5 = 3\mathbf{c}_1 - 2\mathbf{c}_3$

$\mathbf{c}_4 = -\mathbf{c}_1 + 2\mathbf{c}_3$

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

$\mathbf{c}_2 = -2\mathbf{c}_1$ $\quad \mathbf{c}_5 = 3\mathbf{c}_1 - 2\mathbf{c}_3$

$\mathbf{c}_4 = -\mathbf{c}_1 + 2\mathbf{c}_3$

Cols 1 and 3 have the pivots, so a basis for Col $\mathbf{A}$ is $\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$.

# Explanation I

**Proof of (1):**

There is a linear relationship $\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{0}$ between the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of $\mathbf{A}$ if and only if $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Now an echelon form $\mathbf{R}$ of $\mathbf{A}$ is row equivalent to $\mathbf{A}$ and so $\mathbf{A}\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{R}\mathbf{x} = \mathbf{0}$.

Thus $\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{0}$ if and only if $\sum_{j=1}^{n} x_j \mathbf{r}_j = \mathbf{0}$, where $\mathbf{r}_1, \ldots, \mathbf{r}_n$ are the columns of $\mathbf{R}$. That is, $\mathbf{A}$ has the same linear relationships between its columns as does $\mathbf{R}$.

## Illustration

Amongst the columns of $\mathbf{R}$ in the previous illustration we observed that
$$\mathbf{r}_5 = 3\mathbf{r}_1 - 2\mathbf{r}_3.$$
This can be written $3\mathbf{r}_1 + 0\mathbf{r}_2 - 2\mathbf{r}_3 + 0\mathbf{r}_4 - 1\mathbf{r}_5 = \mathbf{0}$.
As explained above, it follows that $3\mathbf{a}_1 + 0\mathbf{a}_2 - 2\mathbf{a}_3 + 0\mathbf{a}_4 - 1\mathbf{a}_5 = \mathbf{0}$, *i.e.*
$$\mathbf{a}_5 = 3\mathbf{a}_1 - 2\mathbf{a}_3$$
The relationship amongst the columns of $\mathbf{R}$ is matched by the same relationhsip amongst the corresponding columns of $\mathbf{A}$.

# Explanation II

**Proof of (2):**

The pivot columns of $\mathbf{R}$ are each columns from an identity matrix, and so are linearly independent.

Hence there is no non-trivial linear relationship between these columns.

By (1) this implies no linear relationship amongst the corresponding columns of $\mathbf{A}$, *i.e.* the pivot columns of $\mathbf{A}$ are also linearly independent.

---

Since $\mathbf{R}$ is in echelon form, any non-pivot column can be written as a linear combination of preceding pivot columns, and by (1) the same applies to $\mathbf{A}$.

It follows that any linear combination of the columns of $\mathbf{A}$ can be rewritten as a linear combination of just the pivot columns of $\mathbf{A}$.
That is, $\operatorname{Col} \mathbf{A}$ is spanned by the pivot columns of $\mathbf{A}$.

---

Thus the set of pivot columns of $\mathbf{A}$ is a linearly independent spanning set for $\operatorname{Col} \mathbf{A}$. By definition, this is a basis for $\operatorname{Col} \mathbf{A}$.

# Coordinate Vector Relative to a Basis.

## Definition

Consider a basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ for a subspace $H$.
For each $\mathbf{x} \in H$ there is a unique set of weights $c_1, \dots c_p$ such that

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

The vector $[\mathbf{x}]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \in \mathbb{R}^p$ is called the **coordinate vector** of $\mathbf{x}$ relative to the basis $B$.

## Example

The set of vectors $B = \left\{ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is linearly independent and $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} \in \text{span}\ \{\mathbf{v}_1, \mathbf{v}_2\}$. Find $[\mathbf{x}]_B$.

Answer: $[\mathbf{x}]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, because $\mathbf{x} = 2\mathbf{v}_1 + 3\mathbf{v}_2$.

# Dimension and Rank

## Definition

The **dimension** of a non-zero subspace $H$, denoted $\dim H$ is the number of vectors in any basis for $H$. By convention $\dim \mathbf{0} = 0$.

## Definition

The **rank** of a matrix $\mathbf{A}$, denoted by $\operatorname{rank} \mathbf{A}$ is the dimension of the column space of $\mathbf{A}$; $\operatorname{rank} \mathbf{A} = \dim \operatorname{Col} \mathbf{A}$.

## Theorem

*If $\mathbf{A}$ has $n$ columns then $\operatorname{rank} \mathbf{A} + \dim \operatorname{Nul} \mathbf{A} = n$.*

## Remarks

1. $\operatorname{rank} \mathbf{A}$ is the number of pivot positions of $\mathbf{A}$.
2. $\dim \operatorname{Nul} \mathbf{A}$ is the number of free variables of $\mathbf{A}$.

# 15. Determinants (Lay 3.1)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/lay03_03_cs.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c03/pdf_files/sec3_1_and_2ov.pdf

# Introduction to Determinants

## $2 \times 2$ Determinants

Recall from Section 11 (Inverses) that for a $2 \times 2$ matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$:

1. The quantity $\det \mathbf{A} = ad - bc$ is called the **determinant** of $\mathbf{A}$.
2. $\det \mathbf{A}$ 'deternines' whether $\mathbf{A}$ is invertible *via* testing for $\det \mathbf{A} \neq 0$.

As we will see shortly, every square matrix has a determinant, but for larger matrices the definition/formula is considerably more complicated.

## Properties of Determinants

Determinants are:

- used for testing whether a matrix is invertible *via*:
  Theorem: $\det \mathbf{A} = 0$ if and only if $\mathbf{A}$ is singular (*i.e.* not invertible);
- useful for theoretical results;
- used to measure areas and volumes;
- not so useful for computational problems involving large matrices.

# Justification of 2×2 Determinant Formula

The matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $\mathbf{A} \xrightarrow{\text{rref}} \mathbf{I}_2$.

Let's verify that this happens if and only if $\det \mathbf{A}$ is non-zero.

<u>Case (i) $a \neq 0$:</u>

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \to \begin{bmatrix} 1 & b/a \\ c & d \end{bmatrix} \to \begin{bmatrix} 1 & b/a \\ 0 & \frac{ad-bc}{a} \end{bmatrix} \to \mathbf{I}_2 \text{ iff}^1 \ ad - bc \neq 0.$$

<u>Case (ii) $a = 0, c \neq 0$:</u>

$$\mathbf{A} = \begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \to \begin{bmatrix} c & d \\ 0 & b \end{bmatrix} \to \begin{bmatrix} 1 & d/c \\ 0 & b \end{bmatrix} = \begin{bmatrix} 1 & d/c \\ 0 & \frac{bc-ad}{c} \end{bmatrix} \to \mathbf{I}_2 \text{ iff } ad-bc \neq 0.$$

<u>Case (iii) $a = 0, c = 0$:</u>

$\mathbf{A} = \begin{bmatrix} 0 & b \\ 0 & d. \end{bmatrix}$ In this case $\mathbf{A} \not\to \mathbf{I}_2$ and $ad-cb = 0-0 = 0$.

[1] "iff" = "if and only if"

# Determinant for 3 by 3 Matrix I

Assuming certain terms are non-zero, the matrix $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ can be reduced to

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where
$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$.

Clearly, if $\Delta = 0$ then $\mathbf{A}$ is singular. By a careful analysis of a number of cases (as for $2 \times 2$ matrices) it can be shown that, whatever the values of individual entries of $\mathbf{A}$, the matrix is singular if and only if $\Delta = 0$.

**Definition**

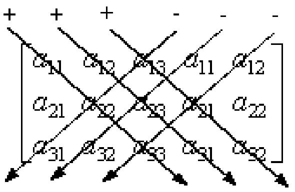
With $\mathbf{A}$ and $\Delta$ as above, $\det \mathbf{A} = \Delta$.

# Determinant for 3 by 3 Matrix II

One way to remember the intimidating formula

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

is via the diagram at right.
From the sum of the 'right' pointing products subtract the sum of the 'left' pointing products.




Another way to think of $\Delta$ is to observe that it can be written

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

If we now denote by $\mathbf{A}_{ij}$ the matrix obtained from $\mathbf{A}$ by removing its $i$-th row and $j$-th column then the above formula can be abbreviated to

$$\Delta = a_{11} \det \mathbf{A}_{11} - a_{12} \det \mathbf{A}_{12} + a_{13} \det \mathbf{A}_{13}.$$

This compact and symmetrical expression leads to a recursive definition for the determinant of any sized square matrix.

# Determinant of a Square Matrix

## Definition

Let $\mathbf{A}$ be any $n \times n$ matrix. The **determinant** $\det \mathbf{A}$ of $\mathbf{A}$ is defined as

$a_{11}$ for $n = 1$ and as

$a_{11} \det \mathbf{A}_{11} - a_{12} \det \mathbf{A}_{12} + \cdots + (-1)^{1+n} a_{1n} \det \mathbf{A}_{1n}$ for $n > 1$.

**Notation:** $\det \mathbf{A}$ may be written out by replacing brackets by vertical lines.

## Illustration

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

$$1 \times \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 2 \end{vmatrix} - 0 \times \det \mathbf{A}_{12} + (-1) \times \begin{vmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & -2 \end{vmatrix} - 0 \times \det \mathbf{A}_{14}.$$

# Determinant Example

## Example

Complete the evaluation of $\det \mathbf{A}$ for the matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 2 & 1 & 0 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & -1 & 1 & -2 \end{bmatrix}$ of the previous illustration.

Answer:

$$\begin{aligned}
\det \mathbf{A} &= \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & -2 \end{vmatrix} - \begin{vmatrix} 2 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & -2 \end{vmatrix} \\
&= \left( 1 \times \begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} + 1 \times \begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix} \right) \\
&\qquad - \left( 2 \times \begin{vmatrix} 1 & -1 \\ -1 & -2 \end{vmatrix} - 1 \times \begin{vmatrix} 0 & -1 \\ 1 & -2 \end{vmatrix} + 1 \times \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} \right) \\
&= \left( 1 \times -1 \ + \ 1 \times 2 \right) - \left( 2 \times -3 \ - \ 1 \times 1 \ + \ 1 \times -1 \right) \\
&= (-1 + 2) - (-6 - 1 - 1) \ = \ 1 + 8 \ = \ 9.
\end{aligned}$$

# Cofactor Expansion

The definition of $\det \mathbf{A}$ involves 'expansion along the first row'; we multiply each entry in the row by the appropriately signed determinant of the matrix obtained eliminating the row and column containing the entry.

It turns out that we can do the expansion using *any row*, or indeed *any column*, provided we modify "appropriately signed" correctly. We need:

## Definition

Let $\mathbf{A}$ be an $n \times n$ matrix. For$1 \leq i, j \leq n$ the **cofactor** $\mathbf{C}_{ij}$ is defined by

$$\mathbf{C}_{ij} = (-1)^{i+j} \det \mathbf{A}_{ij},$$

where $\mathbf{A}_{ij}$ is obtained by omitting the $i^{\text{th}}$ row and $j^{\text{th}}$ column of $\mathbf{A}$.

## Theorem

*For any $n \times n$ matrix $\mathbf{A}$, $\det \mathbf{A}$ can be calculated using cofactor expansion:*

*along the $i^{th}$ row:* $\quad \det \mathbf{A} = a_{i1}\mathbf{C}_{i1} + a_{i2}\mathbf{C}_{i2} + \cdots + a_{in}\mathbf{C}_{in}$

*or down the $j^{th}$ col:* $\quad \det \mathbf{A} = a_{1j}\mathbf{C}_{1j} + a_{2j}\mathbf{C}_{2j} + \cdots + a_{nj}\mathbf{C}_{nj}.$

# Cofactor Examples

## Illustration

$$\begin{vmatrix} 1 & 3 & 4 \\ 0 & 2 & 1 \\ 0 & -1 & 1 \end{vmatrix} = a_{11}\mathbf{C}_{11} + a_{21}\mathbf{C}_{21} + a_{31}\mathbf{C}_{31} = 1\times 3 \ + \ 0 \ + \ 0 = 3$$

(by cofactor expansion down the first column).

## Example

Use cofactors to evaluate the determinant of $\mathbf{A} = \begin{bmatrix} 1 & 0 & -4 & 2 \\ 3 & 2 & -1 & 0 \\ 7 & 0 & 5 & -2 \\ 2 & -1 & 0 & 3 \end{bmatrix}$.

Answer: Using the two zeros in the $2^{\text{nd}}$ column, cofactor expansion gives

$$2\times\begin{vmatrix} 1 & -4 & 2 \\ 7 & 5 & -2 \\ 2 & 0 & 3 \end{vmatrix} + (-1)\times\begin{vmatrix} 1 & -4 & 2 \\ 3 & -1 & 0 \\ 7 & 5 & -2 \end{vmatrix}.$$

Now using the rows containing zeros, further cofactor expansion gives

$$2\left(2\begin{vmatrix} -4 & 2 \\ 5 & -2 \end{vmatrix} + 3\begin{vmatrix} 1 & -4 \\ 7 & 5 \end{vmatrix}\right) - \left(-3\begin{vmatrix} -4 & 2 \\ 5 & -2 \end{vmatrix} + (-1)\begin{vmatrix} 1 & 2 \\ 7 & -2 \end{vmatrix}\right)$$

$$= \ 2(2\times -2 \ + \ 3\times 33) \ - \ (-3\times -2 \ - \ (-16)) = 168.$$

# Numerical Note

Calculating determinants via cofactors requires on the order of $n!$ operations for an $n$ by $n$ matrix.

Thus the calculation of the determinant of a 25 by 25 matrix requires on the order of 25! multiplications. That is approximately $1.5 \times 10^{25}$ multiplications.

Fast computers today can execute approximately $10^{15}$ floating point operations per second (1 petaflops). So such a machine would take approximately 500 years to do the calculation!

So we need a more efficient method. We can use row operations; see next section.

# 16. Properties of Determinants (Lay 3.2) Cramer's Rule, Volume and Linear Transformations (Lay 3.3)

Case Study: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/cs_apps/jacobian.pdf
Lay's Summary: http://media.pearsoncmg.com/aw/aw_lay_linearalg_3/parachute/tm/c03/pdf_files/sec3_3ov.pdf

# The Effect of Row Operations on a Determinant

## Theorem

*Let $\mathbf{A}$ be a square matrix and let $\mathbf{B}$ be obtained from $\mathbf{A}$ by applying an elementary row operation $E$. The relationship between $\det \mathbf{A}$ and $\det \mathbf{B}$ is:*

1. *If $E$ is type 1, 'row replacement' $R_i \leftarrow R_i + cR_j$, then $\det \mathbf{B} = \det \mathbf{A}$.*
2. *If $E$ is type 2, 'row interchange' $R_i \leftrightarrow R_j$, then $\det \mathbf{B} = -\det \mathbf{A}$.*
3. *If $E$ is type 3, 'row scale' $R_i \leftarrow kR_i$, then $\det \mathbf{B} = k \det \mathbf{A}$.*

## Illustrations

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} 7 & 10 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B}_2 = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{B}_3 = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix}$. Then

1. $\det \mathbf{B}_1 = \det \mathbf{A} = -2$ since $\mathbf{B}$ is obtained from $\mathbf{A}$ by $R_1 \leftarrow R_1 + 2R_2$.
2. $\det \mathbf{B}_2 = -\det \mathbf{A} = 2$ since $\mathbf{B}$ is obtained from $\mathbf{A}$ by $R_1 \leftrightarrow R_2$.
3. $\det \mathbf{B}_3 = 2 \det \mathbf{A} = -4$ since $\mathbf{B}$ is obtained from $\mathbf{A}$ by $R_2 \leftarrow 2R_2$

# Using Row Operations to Calculate a Determinant

## Example

Calculate the determinant of $\begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$.

The row operation $R_2 \leftarrow R_2 + 2R_1$ will not change the determinant so

$$\begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix}.$$

The value in this is that we can now use the zeros created in row 2 to simplify the calculation using cofactor expansion:

$$\begin{aligned} &= -(-5)\begin{vmatrix} 1 & -4 \\ -1 & 7 \end{vmatrix} \\ &= 5(7-4) = 15. \end{aligned}$$

# Numerical Note

The calculation of a 25 by 25 determinant using row operations requires approximately 10,000 floating point operations. This takes less than a millisecond on any modern computer.

The world's current fastest computer (since 2015), the Tianhe-2 supercomputer (China)achieves 34 Pflops.

(1 petaflops $= 10^{15}$ floating point operations) per second.

At that rate, Tianhe-2 would take less than one picosecond for the task.
(1 picosecond $= 10^{-12}$ seconds.)
By comparison, Tianhe-2 would take more than 10 years without row operations (*i.e.* using the raw definition of determinant).

# Echelon Form

A square matrix in echelon form is upper triangular, and iterated cofactor expansion down first columns shows that the determinant of such a matrix is just the product of its diagonal entries. (True for all triangular matrices.)
Combining this with the known effects of row operations gives

### Theorem

*Let $\mathbf{U}$ be an echelon form matrix obtained from an $n \times n$ matrix $\mathbf{A}$ via a sequence row interchanges and row replacements but no row scaling. (So pivots have not been scaled to 1.) Then*

$$\det \mathbf{A} = (-1)^r u_{11} u_{22} \cdots u_{nn},$$

*where $r$ is the number of interchanges and $u_{ii}$ is the $i^{th}$ diagonal entry in $\mathbf{U}$.*

The diagonal entries of an echelon form matrix are all non zero if and only if they are all pivots. Together with the theorem above this leads to

### Theorem

*A square matrix $\mathbf{A}$ is invertible if and only if $\det \mathbf{A} \neq 0$.*

# Some Properties of Determinants

**Transpose Property:** $\det \mathbf{A}^T = \det \mathbf{A}.$

This is due to the equivalence of row and column cofactor expansions.

**Multiplication Property:** $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}.$

Proof by expansion as products of elementary matrices.

**Addition Non-Property:** $\det(\mathbf{A} + \mathbf{B}) \neq \det \mathbf{A} + \det \mathbf{B}$ (in general).

## Illustration of Addition Non-Property

$$\det \left( \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \right) = \begin{vmatrix} 5 & 5 \\ 5 & 5 \end{vmatrix} = 25 - 25 = 0.$$

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \det \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} = (4 - 6) + (4 - 6) = -4 \neq 0.$$

# Linearity Property of Determinants

## Notation

For a given $n \times n$ matrix $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \ldots \ \mathbf{a}_n]$ and vector $\mathbf{b} \in \mathbb{R}^n$ denote by $\mathbf{A}_i(\mathbf{b})$ the matrix obtained by replacing the $i^{\text{th}}$ column of $\mathbf{A}$ by $\mathbf{b}$:

$$\mathbf{A}_i(\mathbf{b}) = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_{i-1} \ \mathbf{b} \ \mathbf{a}_{i+1} \ \ldots \ \mathbf{a}_n].$$

## Illustration

Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$. Then $\mathbf{A}_1(\mathbf{b}) = \begin{bmatrix} 5 & 2 \\ 6 & 4 \end{bmatrix}$, $\mathbf{A}_2(\mathbf{b}) = \begin{bmatrix} 1 & 5 \\ 3 & 6 \end{bmatrix}$.

## Theorem

*For any $n \times n$ matrix $\mathbf{A}$ and any $1 \leq i \leq n$ the transformation $T : \mathbb{R}^n \to \mathbb{R}$ given by $T(\mathbf{x}) = \det \mathbf{A}_i(\mathbf{x})$ is linear. That is*

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \textit{and} \quad T(k\mathbf{u}) = kT(\mathbf{u}).$$

These properties follow from a cofactor expansion down the $i^{\text{th}}$ column.

# Cramer's Rule

## Theorem

*If $\mathbf{A}$ is an invertible $n \times n$ matrix then the solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is given by*

$$x_i = \frac{\det \mathbf{A}_i(\mathbf{b})}{\det \mathbf{A}} \qquad i = 1, \ldots, n.$$

## Example

Use Cramer's rule to find $y$ given that:

$$\begin{array}{rcrcrcr} 2x & - & y & + & 3z & = & 4 \\ x & + & 2y & - & z & = & -3 \\ 4x & + & y & + & 2z & = & 2 \end{array}.$$

$$y = \frac{\begin{vmatrix} 2 & \mathbf{4} & 3 \\ 1 & \mathbf{-3} & -1 \\ 4 & \mathbf{2} & 2 \end{vmatrix}}{\begin{vmatrix} 2 & -1 & 3 \\ 1 & 2 & -1 \\ 4 & 1 & 2 \end{vmatrix}} = \frac{-\begin{vmatrix} 1 & -3 & -1 \\ 0 & 10 & 5 \\ 0 & 14 & 6 \end{vmatrix}}{-\begin{vmatrix} 1 & 2 & -1 \\ 0 & -5 & 5 \\ 0 & -7 & 6 \end{vmatrix}} = \frac{-(60-70)}{-(-30+35)} = \frac{10}{-5} = -2.$$

# Proof of Cramer's Rule

Let $\mathbf{I}$ be the $n \times n$ identity matrix $\mathbf{I} = \mathbf{I}_n = [\mathbf{e}_1 \ \mathbf{e}_2 \ \ldots \ \mathbf{e}_n]$. Then

$$\begin{aligned} \mathbf{A}\mathbf{I}_i(\mathbf{x}) &= \mathbf{A}\left[\begin{array}{ccccccc} \mathbf{e}_1 & \cdots & \mathbf{e}_{i-1} & \mathbf{x} & \mathbf{e}_{i+1} & \cdots & \mathbf{e}_n \end{array}\right] \\ &= \left[\begin{array}{ccccccc} \mathbf{A}\mathbf{e}_1 & \cdots & \mathbf{A}\mathbf{e}_{i-1} & \mathbf{A}\mathbf{x} & \mathbf{A}\mathbf{e}_{i+1} & \cdots & \mathbf{A}\mathbf{e}_n \end{array}\right] \\ &= \left[\begin{array}{ccccccc} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{array}\right] \\ &= \mathbf{A}_i(\mathbf{b}). \end{aligned}$$

Taking determinants and using the multiplication property yields

$$\det \mathbf{A} \det \mathbf{I}_i(\mathbf{x}) = \det \mathbf{A}_i(\mathbf{b}).$$

Provided $\det \mathbf{A} \neq 0$ we conclude that

$$x_i = \frac{\det \mathbf{A}_i(\mathbf{b})}{\det \mathbf{A}}$$

because $\det \mathbf{I}_i(\mathbf{x}) = x_i$ by cofactor expansion along row $i$.

# Laplace Transform Example I

In the solution of differential equations we often use a technique called Laplace Transforms. (You will see this method in MATH2503 next year).

Consider the system of differential equations

$$\begin{aligned} 3\frac{dx_1}{dt} &- 2x_2 &= 0 \\ \frac{dx_2}{dt} &- 6x_1 &= 0 \end{aligned}$$

with initial conditions $\quad x_1(0) = 4/3,\ x_2(0) = 1.$

This *differential* system in $x_1 = x_1(t),\ x_2 = x_2(t)$ can be transformed (via the Laplace transform) to an *algebraic* system in $X_1 = X_1(s),\ X_2 = X_2(s)$:

$$\begin{array}{c} 3sX_1 - 2X_2 = 3x_1(0) \\ sX_2 - 6X_1 = x_2(0) \end{array} \qquad \textit{i.e.} \qquad \begin{array}{c} 3sX_1 - 2X_2 = 4 \\ -6X_1 + sX_2 = 1 \end{array} .$$

# Laplace Transform Example II

The system of equations for $X_1, X_2$ has the form $\mathbf{A}\mathbf{x} = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \text{ and } \quad \mathbf{b} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}.$$

Solving using Cramer's rule we first note that

$$\mathbf{A}_1(\mathbf{b}) = \begin{bmatrix} 4 & -2 \\ 1 & s \end{bmatrix}, \qquad \mathbf{A}_2(\mathbf{b}) = \begin{bmatrix} 3s & 4 \\ -6 & 1 \end{bmatrix}$$

and $\qquad \det \mathbf{A} = 3s^2 - 12 = 3(s+2)(s-2) \qquad$ so

$$\begin{aligned} X_1 &= \frac{\det(\mathbf{A}_1(\mathbf{b}))}{\det(\mathbf{A})} = \frac{4s+2}{3(s+2)(s-2)} \\ X_2 &= \frac{\det(\mathbf{A}_2(\mathbf{b}))}{\det(\mathbf{A})} = \frac{s+8}{(s+2)(s-2)}. \end{aligned}$$

We can then apply the inverse Laplace transform to conclude that

$$\begin{aligned} x_1 &= \tfrac{1}{2}e^{-2t} + \tfrac{5}{6}e^{2t}, \\ x_2 &= -\tfrac{3}{2}e^{-2t} + \tfrac{5}{2}e^{2t}. \end{aligned}$$

# A Formula for the Inverse of $\mathbf{A}$

## Theorem

*Let $\mathbf{A}$ be an invertible $n \times n$ matrix.*
*Then the row $i$, column $j$ entry of $\mathbf{B} = \mathbf{A}^{-1}$ is given by*

$$b_{ij} = \frac{\det \mathbf{A}_i(\mathbf{e}_j)}{\det \mathbf{A}} = \frac{C_{ji}}{\det \mathbf{A}}$$

*where $\mathbf{e}_j$ is the $j^{th}$ column of $\mathbf{I}_n$ and $C_{ji}$ is the $ji^{th}$ cofactor of $\mathbf{A}$.*
*(Note the change in order of $i$ and $j$.)*

**Proof:** Suppose $\mathbf{A}\mathbf{x} = \mathbf{e}_j$. Then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{e}_j = \text{col}_j(\mathbf{A}^{-1})$.
That is $\mathbf{x}$ will be the $j$th column of the inverse of $\mathbf{A}$.

By Cramer's rule the $i^{\text{th}}$ component of $\mathbf{x}$, and so the $ij^{\text{th}}$ component of $\mathbf{B} = \mathbf{A}^{-1}$, will be given by

$$b_{ij} = x_i = \frac{\det \mathbf{A}_i(\mathbf{e}_j)}{\det \mathbf{A}}.$$

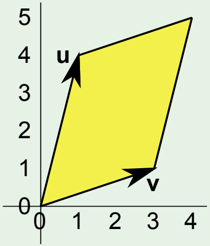
Finally, cofactor expansion along the $j^{\text{th}}$ row shows that $\det \mathbf{A}_i(\mathbf{e}_j) = C_{ji}$.

# Determinants as Area and Volumes

## Theorem

*If $\mathbf{A}$ is a $2 \times 2$ matrix, the area of the parallelogram determined by the columns of $\mathbf{A}$ is $|\det \mathbf{A}|$.*

## Illustration




$$\mathbf{A} = [\mathbf{u}\ \mathbf{v}] = \begin{bmatrix} 1 & 3 \\ 4 & 1 \end{bmatrix}.$$

$$\begin{aligned} \text{Area of parallelogram} &= |\det \mathbf{A}| \\ &= |1 - 12| = |-11| = 11. \end{aligned}$$

## Theorem

*If $\mathbf{A}$ is a $3 \times 3$ matrix, the volume of the parallelepiped determined by the columns of $\mathbf{A}$ is $|\det \mathbf{A}|$.*

# Scaling Factor of a Linear Transformation

## Theorem

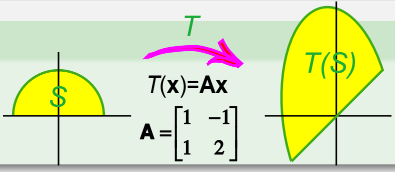
*Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation determined by a $2 \times 2$ matrix $\mathbf{A}$. If $S$ is any region in $\mathbb{R}^2$ with finite area, then*

$$\{\textit{area of } T(S)\} = |\det \mathbf{A}| \cdot \{\textit{area of } S\}\,.$$

*(i.e. the area of the region is 'scaled up' by a factor of $|\det \mathbf{A}|$.)*

## Illustration

For the illustrated transformation $T$, $|\det \mathbf{A}| = 3$. So $T(S)$ has three times the area of $S$.




## Theorem

*Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation determined by a $3 \times 3$ matrix $\mathbf{A}$. If $S$ is any region in $\mathbb{R}^3$ with finite volume, then*

$$\{\textit{volume of } T(S)\} = |\det \mathbf{A}| \cdot \{\textit{volume of } S\}\,.$$

# 17. Complex Numbers (Lay Appendix B and Adams Appendix I)

# Standard Form

Basic facts about the set $\mathbb{C}$ of complex numbers are:

1. The set $\mathbb{C}$ contains an element, usually denoted by $i$, which is not in $\mathbb{R}$.
2. Every member of $\mathbb{C}$ can be written uniquely in the **standard form**

$$x + iy$$

   where $x$ and $y$ are real numbers. 'Uniqueness' means that, if $x_1 + iy_1$ and $x_2 + iy_2$ represent the same complex number, then $x_1 = x_2$ and $y_1 = y_2$ . In other words, if they look different, they are different.
3. The operations of addition, negation and multiplication are defined:

$$\begin{aligned}(x_1 + iy_1) + (x_2 + iy_2) &= (x_1 + x_2) + i(y_1 + y_2),\\ -(x_1 + iy_1) &= (-x_1) + i(-y_1)\\ (x_1 + iy_1)(x_2 + iy_2) &= (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1).\end{aligned}$$

## Illustration

$$\left[2 + i(-3)\right]\left[-1 + i\tfrac{1}{2}\right] = \left(2(-1) - (-3)\tfrac{1}{2}\right) + i\left(2(\tfrac{1}{2}) + (-1)(-3)\right) = -\tfrac{1}{2} + i4$$

From these basic facts all else follows.

# Alternative Notations

There are various fairly obvious variations on standard form. For example:

- $x + yi$ (the $i$ written after the $y$ instead of before it);
  — we almost always use this form when $x$ and $y$ are actual numbers
- $yi$ or $iy$ for $0 + iy$;
- $x$ for $x + i0$ (so the set $\mathbb{R}$ of real numbers is 'embedded' in $\mathbb{C}$);
- $x - iy$ for $x + i(-y)$;
- $i$ and $-i$ for $0 + 1i$ and $0 + (-1)i$;
- $x + i$ and $x - i$ for $x + 1i$ and $x - 1i$;
- other notational simplifications along the same lines;
- in engineering $j$ is sometimes used in place of $i$

## Example

Calculate the product $(1 - 2i)(2 - i)$.

Answer: $-5i$

# Arithmetic Operations

## Theorem

*The complex numbers form a 'field'. This means that all the normal 'rules of arithmetic/algebra' apply (see note below). In particular:*

- *addition and multiplication each have identities, (0 and 1 respec.);*
- *every complex number has an additive inverse (its negative);*
- *every non-zero complex number has a multiplicative inverse*

$$(x+iy)^{-1} = \left(\frac{x}{x^2+y^2}\right) - i\left(\frac{y}{x^2+y^2}\right).$$

To verify the above formula, check that $(x+iy)^{-1}(x+iy) = 1$.

**Note:** The other field-defining 'rules of arithmetic' satisfied by $\mathbb{C}$ are the commutativity and associativity of addition and multiplication and the distributivity of multiplication over addition. See the Adams Appendix if you need explanations of any of these terms.

# $i$ is a Square Root of -1

Complex numbers were invented to provide square roots for negative reals.

The theorem below shows that the mysterious $i$ is actually a square root of $-1$, and square roots for all negative reals can be easily given using $i$.

## Theorem

1. $i^2 = -1$ *and* $(-i)^2 = -1$; *i.e* $\pm i$ *are square roots of* $-1$.
2. *For any positive* $r \in \mathbb{R}$, $\pm i\sqrt{r}$ *are square roots of* $-r$.

Proof:

(1) $i^2 = (0 + i1)(0 + i1) = (0^2 - 1^2) + i(0 \times 1 + 0 \times 1) = -1 + i0 = -1.$
$(-i)^2$ is handled similarly.

(2) Using commutativity and associativity,

$$(i\sqrt{r})^2 = (i\sqrt{r})(i\sqrt{r}) = i^2(\sqrt{r})^2 = (-1)r = -r.$$

$(-i\sqrt{r})^2$ is handled similarly.

# Real and Imaginary Parts

The introduction of $\mathbb{C}$ was at first treated with suspicion, and indeed 'i' was used as standing for 'imaginary'. Despite the now widespread usage of complex numbers, the term is still used as follows:

**Definition**

The complex number $iy$ ($y$ real) is called **imaginary** or **purely imaginary**.

It is often convenient to use a single letter, such as $z$ or $w$, to denote a complex number $x + iy$. To extract $x$ and $y$ from $z$ we use:

**Definition**

If $z = x + iy$ then

- $x$ is the **real part**, $\mathrm{Re}(z)$, of $z$, and
- $y$ is the **imaginary part**, $\mathrm{Im}(z)$, of $z$.

For example, the rule for adding complex numbers can be remembered as 'first add the real parts, then add the imaginary parts'.

# Easy Multiplication

Instead of remembering the somewhat complicated definition for the product of two complex numbers we can make use the following fact:

Products of complex numbers may be calculated using ordinary 'expanding the brackets', $i^2 = -1$ and collecting real and imaginary parts.

This works because, together with $i^2 = -1$, the distributivity, commutativity and associativity properties of $\mathbb{C}$ justify each step below:

$$\begin{aligned}
(x_1 + iy_1)(x_2 + iy_2) &= x_1x_2 + ix_1y_2 + iy_1x_2 + i^2y_1y_2 && \text{(expanding)} \\
&= x_1x_2 + ix_1y_2 + iy_1x_2 - y_1y_2 && (i^2 = -1) \\
&= (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1) && \text{(collecting).}
\end{aligned}$$

This agrees with the definition of multiplication in $\mathbb{C}$.

# Easy Inverses

It is also not necessary to remember the formula for inverses. Instead remember the following trick:

To calculate $(x+iy)^{-1}$, multiply top and bottom of $\dfrac{1}{x+iy}$ by $x-iy$.

This works because, by the 'difference of two squares' formula,

$$(x+iy)(x-iy) = x^2 - (iy)^2 = x^2 - i^2y^2 = x^2 + y^2,$$

and so

$$\begin{aligned}(x+iy)^{-1} = \frac{1}{x+iy} &= \frac{1}{x+iy}\frac{x-iy}{x-iy} \\ &= \frac{x-iy}{x^2+y^2} \\ &= \left(\frac{x}{x^2+y^2}\right) - i\left(\frac{y}{x^2+y^2}\right).\end{aligned}$$

This is exactly the inverse formula given earlier.

# Complex Roots of Quadratics

The first advantage of $\mathbb{C}$ is that all negative reals have square roots.

A second advantage that follows on from this is that every (real) quadratic

$$ax^2 + bx + c, \qquad a, b, c \in \mathbb{R}$$

also now always has roots, namely the possibly non-real complex numbers given by the standard 'quadratic formula' $\dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

## Example

Factorize the quadratic $x^2 + x + 1$.

From the standard formula the roots of the quadratic are

$$\frac{-1 \pm \sqrt{1^2 - 4}}{2} = \frac{-1 \pm \sqrt{-3}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}\, i.$$

Calling these roots $z_1 = -\frac{1}{2} + \frac{\sqrt{3}}{2}\, i$ and $z_2 = -\frac{1}{2} - \frac{\sqrt{3}}{2}\, i$ we have

$$x^2 + x + 1 = (x - z_1)(x - z_2) = (x + \tfrac{1}{2} - \tfrac{\sqrt{3}}{2}\, i)(x + \tfrac{1}{2} + \tfrac{\sqrt{3}}{2}\, i).$$

# The Fundamental Theorem of Algebra

It is an amazing fact that not only quadratics, but **all** (real) polynomials, can be completely factorized in $\mathbb{C}$. Even more amazingly, this remains true even when the coefficients are allowed themselves to be complex:

## Theorem (Fundamental Theorem of Algebra)

*Let* $p(x) = x^n + c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + ... + c_1 x + c_0$ *be any polynomial of degree* $n \geq 1$ *with complex coefficients. Then*

$$p(x) = (x - z_1)(x - z_2)...(x - z_n)$$

*where the roots* $z_1, z_2, ..., z_n$ *are complex numbers.*

The proof is difficult and will not be covered in this course.

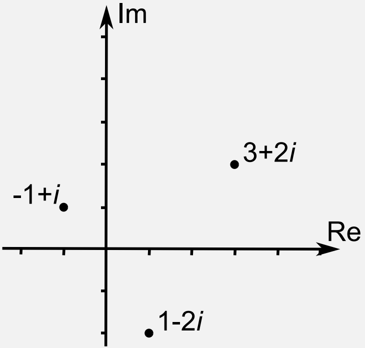
## Example

Verify that the cube roots of $i$ are $-i$, $\frac{\sqrt{3}}{2} + \frac{1}{2}\,i$ and $-\frac{\sqrt{3}}{2} + \frac{1}{2}\,i$; *i.e.* that

$$x^3 - i = (x + i)(x - \tfrac{\sqrt{3}}{2} - \tfrac{1}{2}\,i)(x + \tfrac{\sqrt{3}}{2} - \tfrac{1}{2}\,i).$$

# Argand diagrams

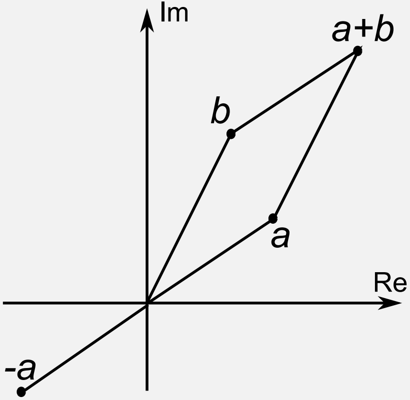
In order to have a picture of complex numbers, it is convenient to plot them in the plane, with their real and imaginary parts as the horizontal and vertical axes respectively. Here are some examples:




Such a diagram is called an **Argand diagram**. Complex numbers represented this way are said to be in the **complex plane**.

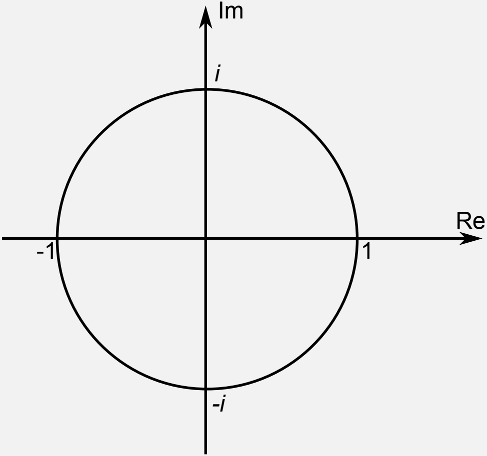
# Geometric Picture of Complex Addition

Since complex numbers add by adding their real and imaginary parts, they add just like 2-dimensional vectors, as on this diagram:

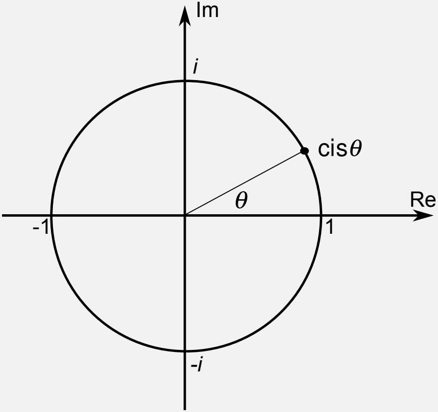


# Unit Circle

The numbers $1$ , $-1$ , $i$ and $-i$ all lie on a unit circle around 0 on the complex plane:



# Complex Numbers on the Unit Circle




What is the number which lies on the unit circle, at an angle of $\theta$ from the real axis?

This number is called **cis** $\theta$.

It can be seen that its real part is cos $\theta$ and its imaginary part is sin $\theta$.

Thus we have

$$\boxed{\text{cis } \theta = \cos\theta + i\sin\theta.}$$

Note that by substituting $\theta = 0, \pi/2, \pi, 3\pi/2$ in the definition gives

$$\text{cis } 0 = 1, \quad \text{cis } \pi/2 = i, \quad \text{cis } \pi = -1, \quad \text{cis } 3\pi/2 = -i.$$

# Multiplying Complex Numbers on the Unit Circle

There is a neat formula for the product of two numbers on the unit circle:

$$\begin{aligned}\text{cis }\theta.\text{cis }\varphi &= (\cos\theta + i\sin\theta).(\cos\varphi + i\sin\varphi)\\ &= \cos\theta\cos\varphi - \sin\theta\sin\varphi + i(\sin\theta\cos\varphi + \cos\theta\sin\varphi)\\ &= \cos(\theta+\varphi) + i\sin(\theta+\varphi).\end{aligned}$$

Thus: $\boxed{\text{cis }\theta.\text{cis }\varphi = \text{cis }(\theta+\varphi).}$

By substituting $\varphi = -\theta$ in this formula, we get cis $\theta$.cis $(-\theta) =$ cis $0 = 1$

Thus: $\boxed{(\text{cis }\theta)^{-1}\ \left[\text{ written cis}^{-1}\theta\ \right] = \text{cis }(-\theta).}$

Also, substituting $\varphi = \theta$, we get $(\text{cis }\theta)^2 = \text{cis }2\theta$ .
Similarly $(\text{cis }\theta)^3 = (\text{cis }\theta)^2.\text{cis }\theta = \text{cis }2\theta.\text{cis }\theta = \text{cis }3\theta$ and so on. Thus

## De Moivre's Theorem

$$(\text{cis }\theta)^n\ \left[\text{ written cis}^n\theta\ \right] = \text{cis }n\theta \quad n = 1, 2, \ldots\ .$$

# Example using De Moivre's Theorem

From De Moivre's Theorem we can get various trigonometric formulas.

For instance, expanding out

$$\text{cis } 3\theta = \text{cis}^3\theta$$

gives

$$\begin{aligned} \cos 3\theta + i \sin 3\theta &= (\cos\theta + i\sin\theta)^3 \\ &= \cos^3\theta + 3i\cos^2\theta\sin\theta - 3\cos\theta\sin^2\theta - i\sin^3\theta. \end{aligned}$$

Now, equating real and imaginary parts,

$$\boxed{\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta}$$

and

$$\boxed{\sin 3\theta = 3\cos^2\theta\sin\theta - \sin^3\theta.}$$

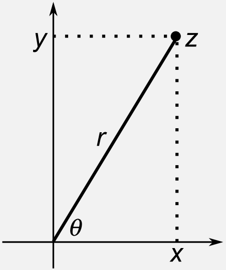
# Polar Form, Absolute Value and Arg

We can write any complex number in **polar form**, by making use of the Argand diagram at right.

A complex number $z$ is represented in standard form by its real and imaginary parts $x$ and $y$. It is also represented in polar form by $r$ and $\theta$.

We have $x = r\cos\theta$ and $y = r\sin\theta$ , so

$$z = x + iy = r\cos\theta + r\sin\theta\, i = r \text{ cis } \theta.$$




## Definition

The number $r$ is called the **absolute value** or **modulus** (**mod** for short) of $z$ and is written $|z|$.

## Definition

The angle $\theta$, when restricted to the range $-\pi < \theta \leq +\pi$ , is called the **argument** of $z$ and written arg $z$ .

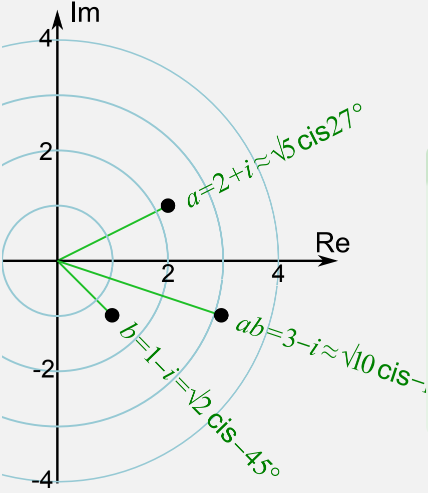
# Conversion between Polar and Standard Forms

We can convert back and forth between standard form and polar form using the formulas $z = x + iy = r \,\text{cis}\, \theta$:

$$
\begin{aligned}
x &= r\cos\theta, \\
y &= r\sin\theta, \\
r &= \sqrt{x^2+y^2}, \\
\theta &= \begin{cases}
\arctan(y/x) & \text{if } x > 0 \\
\arctan(y/x) + \pi & \text{if } x < 0, y \geq 0 \\
\arctan(y/x) - \pi & \text{if } x < 0, y < 0 \\
\pi/2 & \text{if } x = 0, y > 0 \\
-\pi/2 & \text{if } x = 0, y < 0 \\
\text{undefined} & \text{if } x = 0, y = 0
\end{cases} .
\end{aligned}
$$

# Geometric Picture of Complex Multiplication

For a geometrical picture of multiplication, it is best to use polar form.




If $a = r\,\mathrm{cis}\ \theta$ and $b = s\,\mathrm{cis}\ \varphi$, then

$$ab = rs\,\mathrm{cis}(\theta+\varphi);$$

**'mods multiply and args add'**.

## Illustration

The Argand diagram shows two complex numbers $a$ and $b$, together with their product $ab$.

$$|ab| = |a|\,|b| = \sqrt{5}\sqrt{2} = \sqrt{10}$$

$$\begin{aligned}\arg ab &= \arg a + \arg b\\ &\approx 27° - 45° \approx -18°\end{aligned}$$

# Conjugates

## Definition

The **conjugate** of $z = x + iy$ is the number $\overline{z}$ given by $\overline{z} = x - iy$.

Illustration: The conjugate of $\frac{1}{2} + \frac{\sqrt{3}}{2}i$ is $\frac{1}{2} - \frac{\sqrt{3}}{2}i$.

## Theorem

*Conjugation preserves the basic algebraic operations, viz.*

$$\begin{aligned} \overline{a+b} &= \overline{a} + \overline{b}, \\ \overline{-a} &= -\overline{a}, \\ \overline{a-b} &= \overline{a} - \overline{b}, \\ \overline{ab} &= \overline{a}\,\overline{b}, \\ \overline{(a/b)} &= \overline{a}/\overline{b}. \end{aligned}$$

*Also, from the above two properties,*

$$\overline{a^n} = \overline{a}^n \quad \textit{for any integer } n.$$

These are all easy to prove by substitution.

# Further Properties of Conjugation

The following are true for any complex number $z$. Again, they are all easily proved by substituting the definition of the conjugate.

1. $$\overline{\overline{z}} = z,$$

2. $$z\overline{z} = |z|^2,$$

3. $$z = \overline{z} \quad \text{if and only if } z \text{ is real,}$$

4. $$\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z}); \quad \mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z}),$$

5. $$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

# Complex $n^{\text{th}}$ Roots I

In this section we investigate

## Problem

Given a complex number $a = R\text{cis}\,\alpha$ and a positive integer $n$, find all the $n^{\text{th}}$ roots of $a$. That is, find all $n$ roots of the polynomial $x^n - a$.

Note that $a$ is real when $\alpha$ is 0 or a multiple of $\pi$, so the problem includes the case of finding $n^{\text{th}}$ roots of real numbers.

## Some simple cases:

The following examples were demonstrated earlier or can be easily checked.

$a = -2 = 2\text{cis}\,\pi,\ n = 2$: The two square roots of $-2$ are $\pm i\sqrt{2}$.

$a = i = 1\text{cis}\,\pi/2,\ n = 3$: The three cube roots of $i$ are $-i$ and $\pm\frac{\sqrt{3}}{2} + \frac{1}{2}\,i$.

$a = 1 = 1\text{cis}\,0,\quad n = 4$: The four fourth roots of 1 are $\pm 1$ and $\pm i$.

*Checking* that a given $z$ is an $n^{\text{th}}$ root of $a$ is relatively straightforward, but how do we *find* such a $z$ in the first place?

# Complex $n^{\text{th}}$ Roots II

When $a = 0$ all its $n^{\text{th}}$ roots are 0 because $x^n - a$ factorizes as $(x-0)^n$.

From now on we shall assume $a \neq 0$.
In this case, by the fundamental theorem of algebra,

$$x^n - a = (x - z_1)(x - z_2) \cdots (x - z_n)$$

and we shall see that the $n$ $n^{\text{th}}$ roots $z_1, \ldots, z_n$ are all different. Indeed:

## Theorem (Formula for the $n^{\text{th}}$ roots of a complex number)

*The $n^{th}$ roots of $a = R\,\mathit{cis}\,\alpha$, $R \geq 0$, $0 \leq \alpha < 2\pi$, are*

$$z_k = \sqrt[n]{R}\,\mathit{cis}\,\frac{\alpha + 2k\pi}{n}, \qquad k = 0, \ldots, n-1.$$

## Illustration

The square roots of $-1 + i = \sqrt{2}\,\text{cis}\,3\pi/4$ are:

$z_0 = \sqrt[2]{\sqrt{2}}\,\text{cis}\,\frac{3\pi/4}{2} = \sqrt[4]{2}\,\text{cis}\,\frac{3\pi}{8} = \sqrt[4]{2}\cos\frac{3\pi}{8} + i\sqrt[4]{2}\sin\frac{3\pi}{8})$

$z_1 = \sqrt[2]{\sqrt{2}}\,\text{cis}\,\frac{(3\pi/4+2\pi)}{2} = \sqrt[4]{2}\,\text{cis}\,\frac{11\pi}{8} = -\sqrt[4]{2}\cos\frac{3\pi}{8} - i\sqrt[4]{2}\sin\frac{3\pi}{8}) = -z_0$

# Complex $n^{\text{th}}$ Roots III

**Proof of the $n^{\text{th}}$ root formula:**
Suppose $z = r\,\text{cis}\,\theta$ is an $n^{\text{th}}$ root of $a = R\,\text{cis}\,\alpha$ with $r \geq 0$, $0 \leq \theta < 2\pi$. From $z^n = a$ we have $(r\,\text{cis}\,\theta)^n = R\,\text{cis}\,\alpha$, and so by De Moivre

$$r^n \text{cis}\, n\theta = R\,\text{cis}\,\alpha.$$

Taking the absolute value of both sides we find that

$$r^n = R \quad \text{and hence} \quad \text{cis}\, n\theta = \text{cis}\,\alpha.$$

Since $r$ and $R$ are both real and positive, we have $\boxed{r = \sqrt[n]{R}}$.

Since $\text{cis}\, n\theta$ and $\text{cis}\,\alpha$ are both complex numbers on the unit circle they are equal if and only if $n\theta$ and $\alpha$ are equal or differ by any number of exact complete rotations of the circle.

Thus $n\theta = \alpha + k(2\pi)$ for some integer $k$, and so $\boxed{\theta = \dfrac{\alpha + 2k\pi}{n}}$.

This gives different values for $z = r\text{cis}\,\theta$ for $k = 0, \ldots, n-1$
but no new values for $z$ outside this range.

# Complex $n^{\text{th}}$ Roots IV

From the root formula we see that all the roots of $a = R\text{cis}\,\alpha$ lie on a circle of radius $\sqrt[n]{R}$ , centre 0 (in an Argand diagram). They are equally spaced around the circle at angles of $\frac{2\pi}{n}$ .

A particularly important special case is that of finding the complex $n^{\text{th}}$ roots of 1. (These are called "roots of unity".)

In that case the solution given by the formula simplifies to

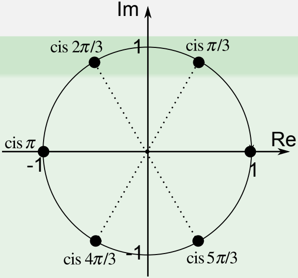
$$z = 1, \text{cis}\,\frac{2\pi}{n}, \text{cis}\,\frac{4\pi}{n}, \ldots, \text{cis}\,\frac{2(n-1)\pi}{n}.$$

# Complex $n^{\text{th}}$ Roots V

## Illustration: The sixth roots of 1

The six sixth roots of 1 are:

$z = 1,\ \text{cis}\,\frac{\pi}{3},\ \text{cis}\,\frac{2\pi}{3},\ \text{cis}\,\pi,\ \text{cis}\,\frac{4\pi}{3},\ \text{cis}\,\frac{5\pi}{3}.$




In standard notation the roots are

$$1,\ \frac{1}{2} + \frac{\sqrt{3}}{2}\mathrm{i},\ -\frac{1}{2} + \frac{\sqrt{3}}{2}\ i,\ -1,\ -\frac{1}{2} - \frac{\sqrt{3}}{2}\ i,\ \frac{1}{2} - \frac{\sqrt{3}}{2}\,i.$$

Using this we can factorise the polynomial $x^6 - 1$ thus

$$\begin{aligned} x^6 - 1 \quad &= \quad (x-1)(x - \frac{1}{2} - \frac{\sqrt{3}}{2}\,i)(x + \frac{1}{2} - \frac{\sqrt{3}}{2}i) \\ &\quad\quad (x+1)(x + \frac{1}{2} + \frac{\sqrt{3}}{2}\,i)(x - \frac{1}{2} + \frac{\sqrt{3}}{2}\,i). \end{aligned}$$

# Polynomials and Conjugates I

If a quadratic has real coefficients but non-real roots, then its two roots are conjugates of one another (what is called a 'conjugate pair'). This is a consequence of the usual formula for roots,

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

For example, the roots of $x^2 + x + 1$ are

$$z = -\frac{1}{2} + \frac{\sqrt{3}}{2}\,i \quad \text{and} \quad \overline{z} = -\frac{1}{2} - \frac{\sqrt{3}}{2}\,i.$$

But this goes both ways: if $z$ and $\overline{z}$ are a conjugate pair, then the quadratic of which they are the roots, namely

$$(x - z)(x - \overline{z}) = (x^2 - (z + \overline{z})x + z\overline{z}),$$

has real coefficients.
This follows from properties of conjugation.

# Polynomials and Conjugates II

## Theorem

*If $f(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$ is a polynomial with real coefficients and $z$ is a root, then so is $\overline{z}$ .*

**Proof:** Given that $z$ is a root, we have

$$c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n = 0.$$

Conjugate this whole equation

$$\overline{c_0 + c_1 z + c_2 z^2 + \cdots , + c_n z^n} = \overline{0}$$

which is the same as

$$\overline{c}_0 + \overline{c_1}\,\overline{z} + \overline{c_2}\,\overline{z}^2 + \cdots + \overline{c_n}\,\overline{z}^n = \overline{0}.$$

But here we have $\overline{0} = 0$ and $\overline{c_j} = c_j$ for each $j$ (since $c_j$ is real).
So the last equation is

$$c_0 + c_1 \overline{z} + c_2 \overline{z}^2 + \cdots , + c_n \overline{z}^n = 0$$

which tells us that $\overline{z}$ is a root of $f$.

# Polynomials and Conjugates III

## Illustration

The sixth roots of 1, calculated earlier, are the roots of the polynomial $x^6 - 1$ . This polynomial has real coefficients and so the conjugate of each root must also be a root. Indeed, the roots are:

- 1 (real, so self-conjugate),
- -1 (real, so self-conjugate),
- $\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $\frac{1}{2} - \frac{\sqrt{3}}{2}i$ (a conjugate pair),
- $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$ (another conjugate pair).

Now if $z$ and $\overline{z}$ are a conjugate pair of roots of a polynomial $f(x)$ then the quadratic $(x-z)(x-\overline{z})$, which has real coefficients, is part of the factorisation of $f(x)$. Hence the previous theorem leads to

## Theorem (Factorization of Real Polynomials)

*Any polynomial with real coefficients can be factorised into factors which are either linear or quadratic and in which all the coefficients are real.*

# Polynomials and Conjugates IV

## Illustration Continued

From the calculation of the sixth roots of unity we have the factorisation

$$\begin{aligned} x^6 - 1 \quad &= \quad (x-1)(x - \frac{1}{2} - \frac{\sqrt{3}}{2}i)(x + \frac{1}{2} - \frac{\sqrt{3}}{2}i) \\ &\quad\quad (x+1)(x + \frac{1}{2} + \frac{\sqrt{3}}{2}\,i)(x - \frac{1}{2} + \frac{\sqrt{3}}{2}\,i). \end{aligned}$$

Multiplying together factor pairs corresponding to conjugate roots gives

$$\begin{aligned} \left(x - \frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right)\left(x - \frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right) \quad &= \quad x^2 - x + 1 \\ \left(x + \frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right)\left(x + \frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right) \quad &= \quad x^2 + x + 1. \end{aligned}$$

This leads to the factorisation into **real** linear and quadratic factors:

$$x^6 - 1 = (x-1)(x+1)(x^2 - x + 1)(x^2 + x + 1).$$

Observe that we used complex numbers to get this result containing only real numbers!

# Connection with Exponential I

Recall these properties of the cis function and the arg function:

$$\begin{aligned} \mathsf{cis}(\theta_1 + \theta_2) &= \mathsf{cis}\,\theta_1 \mathsf{cis}\theta_2 \\ \arg(z_1 z_2) &= \arg z_1 + \arg z_2 \quad \text{(essentially).} \end{aligned}$$

So the cis function acts like the exponential function and the arg function acts like the ln function.

In fact it turns out that we can extend the definitions of the exp and ln functions in such a way that:

$$\begin{aligned} \mathsf{cis}\,\theta &= \exp(i\theta) = e^{i\theta} \\ \ln z &= \ln|z| + i \arg z. \end{aligned}$$

We will only deal here with the first of these properties.

# Connection with Exponential II

From Calculus we know that the exp function can be written as a series:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^{2n}}{(2n)!} + \frac{x^{2n+1}}{(2n+1)!} + \cdots$$

The same series can be used to define $\exp(z)$ for complex $z$. In particular:

$$\begin{aligned}
\exp(i\theta) &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \cdots + \frac{(i\theta)^{2n}}{(2n)!} + \frac{(i\theta)^{2n+1}}{(2n+1)!} + \cdots \\
&= \left(1 + \frac{(i\theta)^2}{2!} + \cdots + \frac{(i\theta)^{2n}}{(2n)!} + \cdots\right) \\
&\quad + \left(i\theta + \frac{(i\theta)^3}{3!} + \cdots + \frac{(i\theta)^{2n+1}}{(2n+1)!} + \cdots\right) \\
&= \left(1 + (-1)\frac{\theta^2}{2!} + \cdots + (-1)^n\frac{\theta^{2n}}{(2n)!} + \cdots\right) \\
&\quad + i\left(\theta + (-1)\frac{\theta^3}{3!} + \cdots + (-1)^n\frac{\theta^{2n+1}}{(2n+1)!} + \cdots\right)
\end{aligned}$$

# Connection with Exponential III

Using the series for cos and sin now gives:

## Theorem (Euler's Theorem)

$$e^{i\theta} = \exp(i\theta) = \cos\theta + i\sin\theta = \mathit{cis}\,\theta.$$

So you can use the cis $\theta$ notation or the $e^{i\theta}$ notation.

Finally, by substituting $\pi$ for $\theta$ and noting that cis $\pi = -1$, we get the 'magic' equation

$$\boxed{e^{i\pi} = -1.}$$