INTRODUCTORY LINEAR ALGEBRA

VLADIMIR V. KISIL

ABSTRACT. This is lecture notes for the course **Introductory Linear Algebra** at School of Mathematics of University of Leeds. They are based on the notes of Dr Reg B. J. T. Allenby used in the previous years. However all misprints, omissions, and errors are only my responsibility. Please let me know if you find any.

The notes are available also for download in PDF format.

The suggested textbooks are [1,2]

BOOKLIST:

[1] R. B. J. T. Allenby. *Linear Algebra*. Edward Arnold, 1995. ↑1

[2] H. Anton and C. Rorres. *Elementary Linear Algebra: applications version*. Wiley, Sixth, 1991. ↑1

NOTATIONS

To avoid possible confusion, it is usual to insert commas between the components of a vector or a $1 \times n$ matrix, thus: $[5, 2]$, $[1, 11, 1, 111]$.

1. GENERAL SYSTEMS OF LINEAR EQUATIONS

1.1. **Introduction.** The subject of Linear Algebra is based on the study of systems of simultaneous linear equations. As far as we are concerned there is one basic technique — that of reducing a matrix to echelon form. I will not assume any prior knowledge of such reduction nor even of matrices. The subject has ramifications in many areas of science, engineering etc. We shall look at only a minuscule part of it. We shall have no time for "real" applications.

Why the name *Algebra*, by the way?

Very often, in dealing with "real life" problems we find it easier—or even necessary! —to simplify the problem so that it becomes mathematically tractable. This often means "linearising" it so that it reduces to a system of (simultaneous) linear equations.

We shall deal with *specific* systems of linear equations in a moment.

1.2. **The different possibilities.** There are essentially *three* different possibilities which can arise when we solve systems of linear equations. These may be illustrated by the following example each part of which involves two equations in two unknowns. Since there are only two unknowns it is more convenient to label them x and y—instead of x_1 and x_2.

We will solve the following system of linear equations by the method of Gauss elimination.

Example 1.1. (i) The first case illustrated by the system:

$$\begin{array}{lrcl} (\alpha) & 2x+5y & = & 3 \\ (\beta) & 3x-2y & = & 14 \end{array}$$

To solve this system eliminate x from equation (β) by replacing (β) by $2\cdot(\beta)-3\cdot(\alpha)$, that is

$$(6x-4y)-(6x+15y)=$$

$$\begin{array}{lrcl} (\alpha) & 2x+5y & = & 3 \\ (\gamma) & 0x-19y & = & \end{array}$$

So the given pair of equations are changed to (α)–(γ). Equation (γ) shows that $y=$ and then (α) shows that $x=$. This system is *consistent* and has the *unique* solution.

(ii) The second case:

$$\begin{array}{lrcl} (\alpha) & 6.8x+10.2y & = & 2.72 \\ (\beta) & 7.8x+11.7y & = & 3.11 \end{array}$$

Replace here (β) by $6.8\cdot(\beta)-$ α gives the following system:

$$\begin{array}{lrcl} (\alpha) & 6.8x+10.2y & = & 2.72 \\ (\gamma) & 0x+0y & = & \end{array}$$

Then (γ) shows

(iii) The third case:

$$\begin{array}{lrcl} (\alpha) & 6.8x+10.2y & = & 2.72 \\ (\beta) & 7.8x+11.7y & = & 3.12 \end{array}$$

Replace here (β) again by $6.8\cdot(\beta)-$ α gives the following system:

$$\begin{array}{lrcl} (\alpha) & 6.8x+10.2y & = & 2.72 \\ (\gamma) & 0x+0y & = & \end{array}$$

Here (γ) imposes no restriction on possible values of x and y, so the given system of equations reduces to the single equation (α).

Taking y to be *any real number you wish*, say $y=c$, then (α) determines a corresponding value of x, namely

$$x=\frac{2.72-}{6.8}.$$

Thus the given system is *consistent* and has *infinitely many* solutions.

This three cases could be seen geometrically if we do three drawings:

When there are more than two equations we try to eliminate even more unknowns. A typical example would proceed as follows:

Example 1.2. (i) We get, successfully, the solution of the following system:

$$\begin{array}{rcll} 3x+y-z & = & 2 & (\alpha) \\ x+y+z & = & 2 & (\beta) \\ x+2y+3z & = & 5 & (\gamma) \end{array} \qquad \begin{array}{rcll} 3x+y-z & = & 2 & (\alpha) \\ 0x-2y-4z & = & -4 & (\delta)=(\alpha)-3(\beta) \\ 0x+y+2z & = & & (\epsilon)= \end{array}$$

$$\begin{array}{rcll} 3x+y-z & = & 2 & (\alpha) \\ 0x+y+2z & = & & (\zeta)= \\ 0x+y+2z & = & 3 & (\epsilon) \end{array} \qquad \begin{array}{rcll} 3x+y-z & = & 2 & (\alpha) \\ 0x+y+2z & = & 2 & (\zeta) \\ 0x+0y+0z & = & & (\eta)= \end{array}$$

The last equation η shows that the given system of equations *has no solutions*. Could you imagine it graphically in a space?

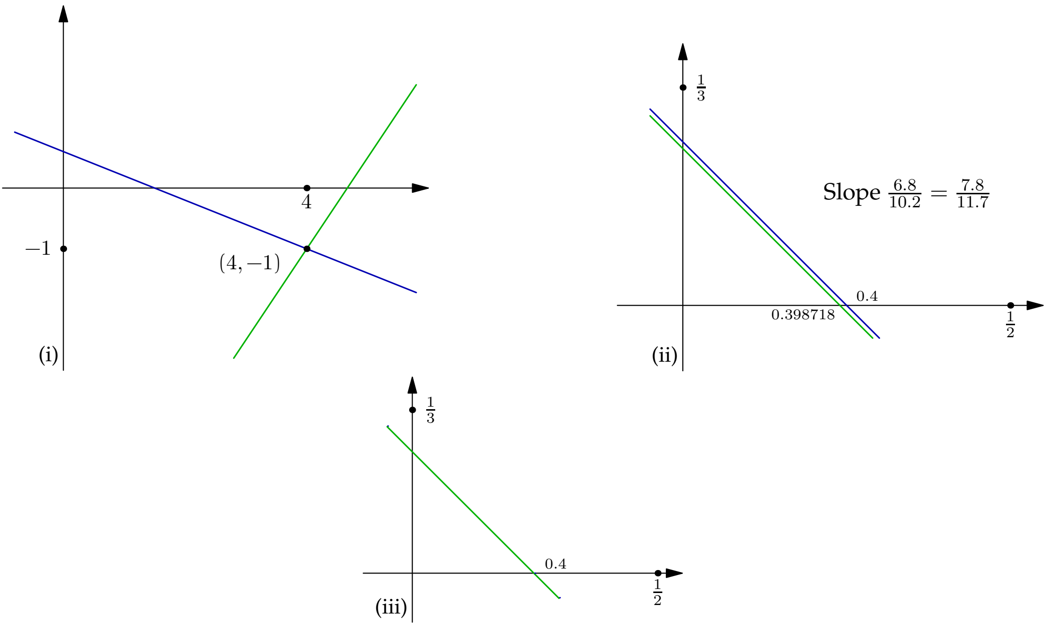


FIGURE 1. Three cases of linear systems considered in Example 1.1.
(i) Lines are in *generic position* and intersect in one point—the unique solution.
(ii) Lines are *parallel* and distinct—there is no solution.
(iii) Lines coinside—all points are solutions.

(ii) Let consider the same system but change only the very last number:

$$\begin{array}{rcll} 3x+y-z &=& 2 & (\alpha) \\ x+y+z &=& 2 & (\beta) \\ x+2y+3z &=& \mathbf{4} & (\gamma) \end{array} \qquad \begin{array}{rcll} 3x+y-z &=& 2 & (\alpha) \\ 0x-2y-4z &=& -4 & (\delta)=(\alpha)-3(\beta) \\ 0x+y+2z &=& & (\epsilon)= \end{array}$$

$$\begin{array}{rcll} 3x+y-z &=& 2 & (\alpha) \\ 0x+y+2z &=& & (\zeta)= \\ 0x+y+2z &=& 2 & (\epsilon) \end{array} \qquad \begin{array}{rcll} 3x+y-z &=& 2 & (\alpha) \\ 0x+y+2z &=& 2 & (\zeta) \\ 0x+0y+0z &=& & (\eta)= \end{array}$$

This time equation (η) places no restrictions whatsoever on x, y and z and so can be ignored. Equation (ζ) tells us that if we take z to have the "arbitrary" value c, say, then y **must** take the value $2-2z = 2-2c$ and then (β) tells us that x **must** take the value $2-y-z = 2-(2-2c)-c = c$. That is, the **general solution** of the given system of equations is: $x = c, y = 2-2c, z = c$, with c being any real number. So solutions include, for example, $(x, y, z) = (1, 0, 1)$, $($ $\qquad$ $, (-7\pi, 2-14\pi, -7\pi), \ldots$. The method of determining y from z and then x from y and z is the method of **back substitution.**

The **general system of** m **linear equations in** n **"unknowns"** takes the form:

$$
\text{(L)}\qquad
\begin{array}{ccccccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1j}x_j & + & \cdots & + & a_{1n}x_n & = & b_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2j}x_j & + & \cdots & + & a_{2n}x_n & = & b_2 \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
a_{i1}x_1 & + & a_{i2}x_2 & + & \cdots & + & a_{ij}x_j & + & \cdots & + & a_{in}x_n & = & b_i \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mj}x_j & + & \cdots & + & a_{mn}x_n & = & b_m
\end{array}
$$

Remark 1.3. (i) The a_{ij} and b_i are *given* real numbers and the **PROBLEM** is to find **all** n-tuples $(c_1, c_2, \ldots, c_j, \ldots, c_n)$ of real numbers such that *when the* $c_1, c_2, \ldots, c_n$ *are substituted for the* $x_1, x_2, \ldots, x_n$, *each of the equalities* in (L) is satisfied. Each such n-tuple is called a **solution** of (the system) (L).

(ii) If $b_1 = b_2 = \ldots = b_n = 0$ we say that the system (L) is **homogeneous.**

(iii) Notice the useful **double suffix notation** in which the symbol a_{ij} denotes the *coefficient of* x_j *in the* **i**-th *equation*.

(iv) In this module the a_{ij} and the b_j will always be **real** numbers.

(v) All the equations are *linear*. That is, in each term $a_{ij}x_j$, each x_j occurs to the power exactly 1. (E.g.: no $\sqrt{x_j}$ nor products such as $x_j^2 x_k$ are allowed.)

(vi) It is **not** assumed that the number of equations is equal to the number of "unknowns".

1.3. **Introduction of Matrices.** Consider the system of equations $\begin{array}{rcl} 2u + 5v & = & 3 \\ 3u - 2v & = & 14 \end{array}$ (cf. Example 1.1(i)). We easily obtain the answer $u = 4$, $v = -1$ (cf. $x = 4$, $y = -1$ in Example 1.1(i)).

This shows that *it is not important which letters are used for the unknowns.*

The important facts are what values the $m \times n$ coefficients a_{ij} and the m coefficients b_j have. Thus we can abbreviate the equations of Example 1.1(i) to the **arrays** $\boxed{\begin{array}{ccc} 2 & 5 & 3 \\ 3 & -2 & 14 \end{array}}$ and $\boxed{\begin{array}{ccc} 2 & 5 & 3 \\ 0 & -19 & -19 \end{array}}$ and those in Example 1.1(ii) to the arrays $\boxed{\begin{array}{ccc} 6.8 & 10.2 & 2.72 \\ 7.8 & 11.7 & 3.11 \end{array}}$ and $\boxed{\begin{array}{ccc} 6.8 & 10.2 & 2.72 \\ 0 & 0 & 0.068 \end{array}}$ correspondingly.

Any such (rectangular) array (usually enclosed in brackets instead of a box) is called a *matrix*. More formally we give the following definition:

Definition 1.4. An array A of $m \times n$ numbers arranged in m rows and n columns is called an m **by** n **matrix** (written "$m \times n$ matrix").

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,j} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \dots & a_{2,j} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & a_{i,3} & \dots & a_{i,j} & \dots & a_{i,n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \dots & a_{m,j} & \dots & a_{m,n} \end{pmatrix}$$

Remark 1.5. We often write the above matrix A briefly as $A = (a_{ij})$ using only the *general term* a_{ij}, which is called the $(\mathbf{i},\mathbf{j})^{\text{th}}$ **entry of A** or the **element in the** $(\mathbf{i},\mathbf{j})^{\text{th}}$ **position (in A)**. Note that the first suffix tells you the *row* which a_{ij} lies in and the second suffix which *column* it belongs to. (Cf. Remark 1.3(iii) above.)

Example 1.6. $\begin{pmatrix} \pi & -\frac{99}{23} & 8.1 \\ 0 & e^2 & -700 \end{pmatrix}$ is a 2×3 matrix with $a_{1,2} =$ and $a_{2,1} =$. What are $a_{2,3}$ and $a_{3,2}$?

1.4. **Linear Equations and Matrices.** With the system of equations (L) as given above we associate two matrices:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \qquad \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} & \vdots & b_1 \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} & \vdots & b_2 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} & \vdots & b_n \\ \vdots & \vdots & & \vdots & & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} & \vdots & b_m \end{pmatrix}$$

The first is the **coefficient matrix of** (the system) (L), the second is the **augmented matrix of** (the system) (L).

Example 1.7. We re-solve the system of equations of Example 1.2(i), noting, at each stage the corresponding augmented matrix.

$$\begin{aligned} 3x + y - z &= 2 \\ x + y + z &= 2 \\ x + 2y + 3z &= 5 \end{aligned} \qquad \begin{pmatrix} 3 & 1 & -1 & 2 \\ 1 & 1 & 1 & 2 \\ 1 & 2 & 3 & 5 \end{pmatrix}$$

We passed from this system to the equivalent by interchanging the first two equations;

$$\begin{aligned} x + y + z &= 2 \\ 3x + y - z &= 2 \\ x + 2y + 3z &= 5 \end{aligned} \qquad \begin{pmatrix} 1 & 1 & 1 & 2 \\ 3 & & & \\ 1 & 2 & 3 & 5 \end{pmatrix}$$

from this system we get the next by subtracting three times the first equation from the second and then the first equation from the third;

$$\begin{aligned} x + y + z &= 2 \\ 0x - 2y - 4z &= -4 \\ 0x + y + 2z &= \end{aligned} \qquad \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & -2 & -4 & -4 \\ 0 & & & \end{pmatrix}$$

we get the next system by multiplying the second equation by $-\frac{1}{2}$;

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

Finally we get the last system by subtracting the second equation from the third.

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

The last equation again demonstrates

1.5. **Reduction by Elementary Row Operations to Echelon Form (Equations and Matrices).** Clearly it is possible to operate with just the augmented matrix; we need not retain the unknowns. Note that in Example 1.7 the augmented matrices were altered by making corresponding **row** changes. These types of changes are called *elementary row operations.*

Definition 1.8. On an $m \times n$ matrix an **elementary row operation** is one of the following kind:

(i) An interchange of two rows;
(ii) The multiplying of one row by a non-zero real number;
(iii) The adding (subtracting) of a multiple of one row to (from) another.

We do an example to introduce some more notation

Example 1.9. Solve the system of equation:

$$\begin{aligned} -x_2+x_3+2x_4 &= 2 \\ x_1+2x_3-x_4 &= 3 \\ -x_1+2x_2+4x_3-5x_4 &= 1. \end{aligned}$$

We successfully reduce the augmented matrix:

$$\begin{pmatrix} 0 & -1 & 1 & 2 & 2 \\ 1 & 0 & 2 & -1 & 3 \\ -1 & 2 & 4 & -5 & 1 \end{pmatrix} \qquad \rho_1 \leftrightarrow \rho_2$$

We change the first and second row of the matrix.

$$\begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & -1 & 1 & 2 & 2 \\ -1 & & & & \end{pmatrix} \qquad \rho_3' = \rho_3 + \rho_1$$

The "new" row 3 is the sum of the "old" row 3 and the "old" row 1.

$$\begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & -1 & 1 & 2 & 2 \\ 0 & 2 & & & \end{pmatrix} \qquad \rho_3' = \rho_3 + 2\rho_2$$

Now third row is added by the twice of the second.

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

Finally we multiple the second row by -1 and the third by $1/8$.

$$\begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & 1 & -1 & -2 & -2 \\ 0 & 0 & 1 & -1/4 & 1 \end{pmatrix}$$

This correspond to the system:

$$\begin{aligned} x_1 + 2x_3 - x_4 &= \\ x_2 - x_3 - \quad &= -2 \\ x_3 - \frac{1}{4}x_4 &= 1 \end{aligned}$$

Here, if we take x_3 as having arbitrary value c, say, then we find $x_4 =$ then $x_2 =$ and $x_1 =$. Hence the most general solution is

$$(x_1, x_2, x_3, x_4) =$$

where c is an arbitrary real number. A variable, such as x_3 here, is called a **free variable** or, sometimes, a **disposable unknown.**

1.6. **Echelon Form (of Equations and Matrices).** In successively eliminating unknowns from a system of linear equations we end up with a so-called *echelon matrix*.

Definition 1.10. An $m \times n$ matrix is in **echelon form** if and only if each of its *non-zero* rows begins with more zeros than does any previous row.

Example 1.11. The first two of following two matrices are in echelon form:

$$\begin{pmatrix} 1 & 5 & 13 \\ 0 & -\pi & -2 \\ 0 & 0 & 37 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 1 & 1 & 7 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

the last matrix is **not** in echelon form

Exercise* **1.12.** Prove that if an i-th row of a matrix in echelon form consists only of zeros then all subsequent rows also consist only of zeros.

SUMMARY In solving a system of simultaneous linear equations

(i) Replace the equations by the corresponding augmented matrix,
(ii) Apply elementary row operations to the matrices in order to reduce the original matrix to one in echelon form.
(iii) Read off the solution (or lack of one) from the echelon form.

Example 1.13. Solve for x, y, z the system (note: there are 4 equations and only 3 unknowns):

$$\begin{aligned} x + 2y + 3z &= 1 \\ 2x - y - 9z &= 2 \\ x + y - z &= 1 \\ 3y + 10z &= 0 \end{aligned}$$

First we construct the augmented matrix and then do the reduction to the echelon form:

$$\begin{pmatrix}1&2&3&1\\0&1&3&0\\0&0&1&0\\0&0&0&0\end{pmatrix}\quad \begin{matrix}\rho_3'=-\rho_3\\ \rho_4'=\rho_4+\rho_3\end{matrix}$$

Then the system
$$\begin{aligned}x+2y+3z&=1\\ y+3z&=0\\ z&=0\end{aligned}$$
has the solution $z=0, y=0, x=1$.

Thus equations were *not* independent, otherwise 4 equation in 3 unknowns does not have any solution at all.

Example 1.14. Find the full solution (if it has one!) of the system:

$$\begin{aligned}2x+2y+z-t&=0\\ x+y+2z+4t&=3\\ 3x+3y+z-3t&=-1\\ x+y+z+t&=1\end{aligned}$$

The augmented matrix is:
$$\begin{pmatrix}2&2&1&-1&0\\1&1&2&4&3\\3&3&1&-3&-1\\1&1&1&1&1\end{pmatrix}$$

The successive transformations are:

$$\begin{pmatrix}1&1&1&1&1\\1&1&2&4&3\\3&3&1&-3&-1\\2&&&&0\end{pmatrix}\quad \begin{pmatrix}1&1&1&1&1\\0&0&1&3&2\\0&0&-2&-6&-4\\0&0&&&\end{pmatrix}$$

$$\begin{pmatrix}1&1&1&1&1\\0&0&1&3&2\\0&0&0&0&0\\0&0&&&\end{pmatrix}$$

so that the original system has been reduced to the system
$$\begin{aligned}x+y+z+t&=1\\ z+3t&=2\end{aligned}$$

with the general solution $t=c, z=2-3c, y=d, x=2c-d-1$. Two particular solutions are $(x,y,z,t)=(-4,-1,8,-2)$ and $(x,y,z,t)=\left(\frac{1}{3},0,0,\frac{2}{3}\right)$.]

Example 1.15. Discuss the system reducing the augmented matrix

$$\begin{aligned}x-y+z&=2\\ 2x+3y-2z&=-1\\ x-6y+5z&=5\end{aligned}\qquad \begin{pmatrix}1&-1&1&2\\2&3&-2&-1\\1&-6&5&5\end{pmatrix}$$

Since the last row corresponds to the equation $0x+0y+0z=-2$, which clearly has no solution, we may deduce that the original system of equations has no solution.

1.7. **Aside on Reduced Echelon Form.** The above examples illustrate that every matrix can, by using row operations, be changed to *reduced echelon form.* We first make the following

Definition 1.16. In a matrix, the first non-zero element in a non-zero row is called the **pivot** of that row.

Now we define a useful variant of echelon form

Definition 1.17. The $m \times n$ matrix is in **reduced echelon form** if and only if

(i) It is in echelon form;
(ii) each pivot is equal to 1;
(iii) each pivot is the *only non-zero element* of its *column*.

Example 1.18. Here is only the second matrix in reduced echelon form:

$$\begin{pmatrix} 1 & 0 & 3 & 0 & 11 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 3 & 0 & 11 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 3 & -4 & 11 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Which are in echelon form?

Why the reduced echelon form is useful? A system of equations which gives rise to the second of the above matrices is equivalent to the system

$$\begin{aligned} x + 3z &= 11 \\ y - 2z &= 3 \\ t &= -1 \end{aligned}$$

Note that, if we take z as the free variable, then each of x and y is immediately expressible (without any extra work) in terms of z. Indeed we get (immediately) $x = 11 - 3z$, $y = 3 + 2z$, $t = -1$.

We shall be content to solve systems of equations using the ordinary echelon version. However reduced echelon form will be used later in the matrix algebra to find inverses.

1.8. **Equations with Variable Coefficients.** The above example are rather simple and all are solved in a routine way. In applications it is sometime required to consider more interesting cases of linear systems with variable coefficients. For different values of parameters we could get all three situations illustrated on Figure 1.

Example 1.19. Find the values of k for which the following system is consistent and solve the system for these values of

$$\begin{aligned} x + y - 2 &= \mathbf{k} \\ 2x + y - 3z &= \mathbf{k^2} \\ x - 2y + z &= -2 \end{aligned}$$

The augmented matrix is

$$\begin{pmatrix} 1 & 1 & -2 & k \\ 2 & 1 & -3 & k^2 \\ 1 & -2 & 1 & -2 \end{pmatrix}$$

The successive transformations are:

$$\begin{pmatrix} 1 & 1 & -2 & k \\ 0 & -1 & 1 & k^2 - 2k \\ 0 & -3 & 3 & -2 - k \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & -2 & k \\ 0 & -1 & 1 & k^2 - 2k \\ & & & \end{pmatrix}$$

Consequently, for the given system to be consistent we require $3k^2 - 5k + 2$ to be 0. But $3k^2 - 5k + 2 = (3k - 2)(k - 1)$. Hence the system is consistent if and only if $k = \frac{2}{3}$ or 1.

In the former case we have

$$\begin{pmatrix} 1 & 1 & -2 & \frac{2}{3} \\ 0 & -1 & 1 & -\frac{8}{9} \end{pmatrix} \quad \text{or} \quad \begin{array}{rcl} x+y-2z & = & \frac{2}{3} \\ -y+z & = & -\frac{8}{9} \end{array}$$

giving $y = z + \frac{8}{9}$, $x = z - \frac{2}{3}$.

So the general solution is $(x, y, z) = \left(c - \frac{2}{3}, c + \frac{8}{9}, c\right)$ for each real number c.

Corresponding to $k = 1$ we likewise get $(x, y, z) = (c, (c+1), c)$ for each real c.

Answer: if $k \neq \frac{2}{3}$ and $k \neq 1$ then there is *no solution*;
if $k = \frac{2}{3}$ then $(x, y, z) = \left(c - \frac{2}{3}, c + \frac{8}{9}, c\right)$ for each real c;
if $k = 1$ then $(x, y, z) = (c, (c+1), c)$ for each real c.

Example 1.20. Discuss the system with a parameter k:

$$\begin{array}{rcl} x+2y+3z & = & 1 \\ x-z & = & 1 \\ 8x+4y+\mathbf{k}z & = & 4 \end{array}$$

The augmented matrix is:

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 0 & -1 & 1 \\ 8 & 4 & k & 4 \end{pmatrix}$$

Its successive transformations are:

The final equation, $kz = -4$, has solution $z = -\frac{4}{k}$ *provided that* $k \neq 0$. In that case from the second equation $y = -2z =$ and from the first equation $x = 1 - 2y - 3z =$.

Answer: If $k = 0$ then there is *no solution*;
otherwise $(x, y, z) =$.

Example 1.21. What condition on a, b, c, d makes the followingsystem consistent

$$\begin{array}{rcl} x_1 + 2x_3 - 6x_4 - 7x_5 & = & a \\ 2x_1 + x_2 + x_4 & = & b \\ x_2 - x_3 + x_4 + 5x_5 & = & c \\ -x_1 - 2x_2 + x_3 - 6x_5 & = & d \end{array}$$

The augment matrix is:

$$\begin{pmatrix} 1 & 0 & 2 & -6 & -7 & a \\ 2 & 1 & 0 & 1 & 0 & b \\ 0 & 1 & -1 & 1 & 5 & c \\ -1 & -2 & 1 & 0 & -6 & d \end{pmatrix}$$

Its transformations are:

$$\begin{pmatrix} 1 & 0 & 2 & -6 & -7 & \\ 0 & 1 & -4 & 13 & 14 & \\ 0 & 1 & -1 & 1 & 5 & \\ 0 & -2 & 3 & -6 & -13 & \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 2 & -6 & -7 & \\ 0 & 1 & -4 & 13 & 14 & \\ 0 & 0 & 3 & -12 & -9 & \\ 0 & 0 & -5 & 20 & 15 & \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 2 & -6 & -7 & \\ 0 & 1 & -4 & 13 & 14 & \\ 0 & 0 & 1 & -4 & -3 & \\ 0 & 0 & -1 & 4 & 3 & \end{pmatrix}$$

Omitting the last (obvious) step we see that the condition for consistency is:

$$(c - b + 2a)/3 \quad = \quad -(d - 3a + 2b)/5$$

that is,
$5(c - b + 2a) = -3(d - 3a + 2b)$,
or $a + b + 5c + 3d = 0$.

2. Matrices and Matrix Algebra

Matrices are made out of numbers. In some sense they also are "like number", i.e. we could equate, add, and multiply them by number or by other matrix (under certain assumptions). Shortly we define all *algebraic operation* on matrices, that is rules of *matrix algebra*.

Historically, matrix *multiplication* appeared first but we begin with a trio of simpler notions.

2.1. Equality.

The most fundamental question one can ask if one is wishing to develop an arithmetic of matrices is: when should two matrices be regarded as *equal*? The only (?) sensible answer seems to be given by

Definition 2.1. Matrices $A = [a_{ij}]_{m\times n}$ and $B = [b_{kl}]_{r\times s}$ are **equal** if and only if $m = r$ and $n = s$ and $a_{uv} = b_{uv}$ for all u, v $(1 \leqslant u \leqslant m\{= r\}, 1 \leqslant v \leqslant n\{= s\})$.

That is, two matrices are equal when and only when they *"have the same shape"* and elements in corresponding positions are equal.

Example 2.2. Given matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $B = \begin{pmatrix} 5 & 13 \\ 3 & 8 \end{pmatrix}$ and $C = \begin{pmatrix} r & s & t \\ u & v & w \end{pmatrix}$ we see that neither A nor B can be equal to C (because C is the "wrong shape") and that A and B are equal if, and only if, $a = 5$, $b = 13$, *and* .

2.2. Addition.

How should we define the *sum* of two matrices? The following has always seemed most appropriate.

Definition 2.3. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ both be $m \times n$ matrices (so that *they have the same shape*). Their **sum** $A \oplus B$ is the $m \times n$ matrix

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \oplus \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{pmatrix}.$$

That is, *addition is componentwise.*

We use the symbol $\oplus$ (rather than $+$) to remind us that, whilst we are *not* actually adding *numbers*, we are doing something very similar — namely, adding *arrays* of numbers.

Example 2.4. We could calculate:

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

Example 2.5. A sum is not defined for the matrices (due to different sizes):

$$\begin{pmatrix} 2 & 1 & 6 \\ 1 & -3 & 2 \\ 0 & 31 & -7 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & 5 \\ 4 & 9 \\ -1 & 6 \end{pmatrix}$$

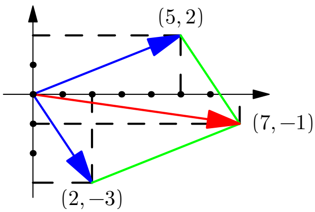


Figure 2. Vector addition

In particular, the sum of the two 1×2 matrices $[5, 2]$ and $[2, -3]$ is the 1×2 matrix $[7, -1]$. The reader who is familiar with the idea of vectors in the plane will see from Figure 2 that, in this case, matrix addition coincides with the usual parallelogram law for vector addition of vectors in the plane.

A similar correspondence likewise exists between 1×3 matrices and vectors in three-dimensional space. It then becomes natural to speak of the $1 \times n$ matrix $[a_1 a_2 \dots a_n]$ as being a *vector in* n*-dimensional space* — even though few of us can "picture" n-dimensional space geometrically for $n \geqslant 4$. Thus, for $n \geqslant 4$, the *geometry* of n-dimensional space seems hard but its corresponding *algebraic version* is equally easy for all n.

Since it is the *order* in which the components of an n-dimensional vector occur which is important, we could equally represent such an n-vector by an $n \times 1$ matrix $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ — rather than $(a_1 a_2 \dots a_n)$— and on many occasions we shall do just that. Later, we shall readily swap between the vector notation $\mathbf{v} = (a_1, a_2, \dots, a_n)$ and either of the above matrix forms, as we see fit and, in particular, usually use bold letters to represent $n \times 1$ and $1 \times n$ matrices.

2.3. **Scalar Multiplication.** Next we introduce multiplication into matrices. There are two types. To motivate the first consider the matrix sums $A \oplus A$ and $A \oplus A \oplus A$ where A is the matrix $\begin{pmatrix} a & b & c & d \\ p & q & r & s \\ x & y & z & t \end{pmatrix}$. Clearly $A \oplus A = \begin{pmatrix} 2a & 2b & 2c & 2d \\ 2p & 2q & 2r & 2s \\ 2x & 2y & 2z & 2t \end{pmatrix}$ whilst $A \oplus A \oplus A = \begin{pmatrix} 3a & 3b & 3c & 3d \\ 3p & 3q & 3r & 3s \\ 3x & 3y & 3z & 3t \end{pmatrix}$. If, as is natural, we write the sum $A \oplus A \oplus \cdots \oplus A$ of k copies of A briefly as kA, we see that kA is the matrix each of whose elements is k times the corresponding element of A. There seems no reason why we shouldn't extend this to any rational or even real value of k, as in

Definition 2.6. *Scalar Multiplication:* If α is a number (in this context often called a **scalar**) and if A is the $m \times n$ matrix above then αA is defined to be the $m \times n$ matrix $\begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \dots & \alpha a_{mn} \end{pmatrix}$ (briefly $[\alpha a_{ij}]$).

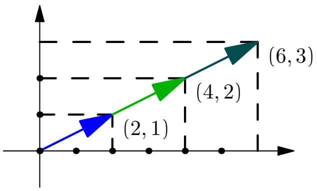


Figure 3. Multiplication by a scalar.

Thus, multiplying a $1\times n$ (or $n\times 1$) matrix by a scalar corresponds, for $n = 2$ and 3, to the usual multiplication of a vector by a scalar. See Figure 3.

2.4. **Multiplication.** To motivate the definition of the multiplication of two matrices we follow the historical path. Indeed, suppose that we have two systems of equations:

$$\begin{aligned} z_1 &= a_{11}y_1 + a_{12}y_2 + a_{13}y_3 \\ z_2 &= a_{21}y_1 + a_{22}y_2 + a_{23}y_3 \end{aligned} \quad \text{and} \quad \begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 \\ y_2 &= b_{12}x_1 + b_{22}x_2 \\ y_3 &= b_{13}x_1 + b_{32}x_2 \end{aligned}$$

We associate, with these systems, the matrices of coefficients namely

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

Clearly we may substitute the y's from the second system of equations into the first system and obtain the z's in terms of the x's. If we do this what matrix of coefficients do we get? It is fairly easy to check that the resulting matrix is the 2×2 matrix $C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$ where, for example, $c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31}$—and, generally, $c_{ij} =$ where i and j are either of the integers 1 and 2. We call C the **product** of the matrices A and B. Notice how, for each i, j, the element c_{ij} of C is determined by the elements a_{i1}, a_{i2}, a_{i3} of the i-th *row* of A and those, b_{1j}, b_{2j}, b_{3j}, of the j-th *column* of B. Notice, too, how this definition requires that the number of *columns* of A must be equal to the number of *rows* of B and that the number of rows (columns) of C is the same as the number of rows of A (columns of B).

We adopt the above definition for the product of two general matrices in

Definition 2.7. *Multiplication of two matrices:* Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. (Note the positions of the two ns). Then the **product** $A \odot B$ is the $m \times p$ matrix $[c_{ij}]_{m\times p}$ where for each i, j, we set $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$.

It might be worthwhile noting that each "outside" pair of numbers in each product $a_{ik}b_{kj}$ is i, j whereas each "inside" pair is a pair of equal integers ranging from 1 to n. We remind the reader that the very definition of c_{ij} explains why we insist that the number of *columns* of A must be equal to the number of *rows* of B.

Some examples should make things even clearer.

Example 2.8. Let $A = \begin{pmatrix} 3 & 1 & 7 \\ 2 & -5 & 4 \end{pmatrix}$; $B = \begin{pmatrix} x & \alpha \\ y & \beta \\ z & \gamma \end{pmatrix}$; $C = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$; $D = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then $A \odot B$ exists and the result $\begin{pmatrix} 3x + y + 7z & 3\alpha + \beta + 7\gamma \\ 2x - 5y + 4z & 2\alpha \end{pmatrix}$ *is* 2×2 matrix.

Also $B \odot A$ exists, the result, $\begin{pmatrix} x3+\alpha 2 & x1+\alpha(-5) & x7+\alpha 4 \\ y3+\beta 2 & y1+\beta(-5) & y7+\beta 4 \\ z3+\gamma 2 & & \end{pmatrix}$ being 3×3 matrix.

Finally observe that $D \odot A$ is the 2×3 matrix $\begin{pmatrix} 1\cdot 3+2\cdot 2 & 1\cdot 1+2\cdot(-5) & 1\cdot 7+2\cdot 4 \\ 3\cdot 3+4\cdot 2 & & \end{pmatrix} =$ $\begin{pmatrix} 7 & -9 & 15 \\ & & \end{pmatrix}$ and yet $A \odot D$ doesn't exist (since A is $2 \times \underline{3}$ and D is $\underline{2} \times 2$).

We have seen that $D \odot A$ exists and yet $A \odot D$ doesn't, and that $A \odot B$ and $B \odot A$ *both* exist and yet are (very much!) unequal (since they are not even the same shape!)

Example 2.9. Even if both products $A \odot B$ and $B \odot A$ are of the same shape they are likely to be different, for example:

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad A \odot B = \qquad , \quad B \odot A = \qquad .$$

Exercise* 2.10. Show that if both products $A \odot B$ and $B \odot A$ are defined then the both products $A \odot B$ and $B \odot A$ are square matrices (but may be of different sizes).

Remark 2.11. Each system of equations is now expressible in form $A \odot \mathbf{x} = \mathbf{b}$.

Indeed, the system

$$\begin{aligned} 3x+y-5z+8t &= 2 \\ x-7y-4t &= 2 \\ -4x+2y+3z-t &= 5 \end{aligned}$$

can be written in matrix form as

$$\begin{pmatrix} 3 & 1 & -5 & 8 \\ 1 & -7 & 0 & -4 \\ -4 & 2 & 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix},$$

that is, as $A \odot \mathbf{x} = \mathbf{b}$ where $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix}$. Note that multiplication of 3×4 matrix A by 4×1 matrix (vector) $\mathbf{x}$ gives a 3×1 matrix (vector) $\mathbf{b}$.

Question: Could we then solve the above system $A \odot \mathbf{x} = \mathbf{b}$ just by the formula $\mathbf{x} = A^{-1} \odot \mathbf{b}$ with some suitable A^{-1}?

Before making some useful observations it is helpful to look at some **Comments and Definitions**

(i) Two matrices are **equal** if both are $m \times n$ **AND** if the elements in corresponding positions are equal.

Example 2.12. (i) $\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$; (ii) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$.

(iii) $\begin{pmatrix} 1 & x \\ 3 & -7 \end{pmatrix} = \begin{pmatrix} y & 4 \\ z & t \end{pmatrix}$ if, and only if, $x = 4$, $y = 1$, $z = 3$ and $t = -7$.

Exercise 2.13. Given that $\begin{pmatrix} 4 & x & y \\ t & 2 & 7 \end{pmatrix} = \begin{pmatrix} x & x & x \\ z & z & w \end{pmatrix}$ find x, y, z, w and t.

(ii) If $m = n$ then the $m \times n$ matrix is said to be **square**.

(iii) If we interchange the rows and columns in the $m \times n$ matrix A we obtain an $n \times m$ matrix which we call the **transpose** of A. We denote it by A^T or A'.

Example 2.14. $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ is 2×3 and $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$ is 3×2.Note how the i-th *row* of A becomes the i-th **column** of A^T and the j-th*column* of A becomes the j-th **row** of A^T.

2.5. **Identity Matrices, Matrix Inverses.** Let $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{12} & a_{22} & a_{23} & a_{24} \\ a_{13} & a_{32} & a_{33} & a_{34} \\ a_{14} & a_{42} & a_{43} & a_{44} \end{pmatrix}$ be any 4×4 matrix. Consider the matrix $I_{4,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. The products $A \odot I_{4,4}$ and $I_{4,4} \odot A$ are both equal to A—as is easily checked. $I_{4,4}$ is called the **4×4 identity matrix**. There is a similar matrix for each "size". For example $I_{3,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is the corresponding 3×3 matrix. The identity matrices are, to matrix theory, what the number 1 is to number theory, namely multipliers which don't change things. And just as the number $\frac{7}{88}$ is the *multiplicative inverse* of the number $\frac{88}{7}$ (since their product $\frac{88}{7} \times \frac{7}{88}$ is equal to 1) so we make the following definition:

Definition 2.15. Let A and B be two $n \times n$ matrices. If $A \odot B = B \odot A = I_{n \times n}$ (the identity $n \times n$ matrix) then B is said to be the **multiplicative inverse of** A — and (of course) A is the multiplicative inverse of B.
Notation: $B = A^{-1}$ and likewise $A = B^{-1}$.

Example 2.16. $\begin{pmatrix} 2 & 0 & -3 \\ -1 & 4 & -8 \\ 0 & -5 & 12 \end{pmatrix}$ and are multiplicative inverses.

Remark 2.17. We shall soon replace the rather pompous signs $\oplus$ and $\odot$ by $+$ and $\times$. Indeed we may drop the multiplication sign altogether writing AB rather than $A \odot B$ or $A \times B$. In particular $A \odot \mathbf{x} = \mathbf{b}$ becomes $A\mathbf{x} = \mathbf{b}$.

If an $n \times n$ matrix A has a multiplicative inverse we say that A is **invertible** or **non-singular**. **Not all matrices are invertible**. For example, neither $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ nor $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ have multiplicative inverses. This is scarcely surprising since each is a matrix corresponding to the number 0. But neither has $\begin{pmatrix} -1 & 1 & 2 \\ -1 & 1 & 2 \\ 3 & 7 & 9 \end{pmatrix}$ a (mult.) inverse. Later we shall give tests for determining which matrices *have* inverses and which do not. First we give a *method* of determining whether or not a given (square) matrix has a (multiplicative) inverse and, if it has, of finding it.

Example 2.18. Determine if the matrix $A = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 4 & 1 \\ 1 & 3 & 0 \end{pmatrix}$ has an inverse — and if it has, find it.

METHOD Form the 3×6 matrix $B = \begin{pmatrix} 1 & 0 & 3 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{pmatrix}$ in which the given matrix A is followed by the 3×3 identity matrix $I_{3,3}$. We now apply row operations to change B into *reduced echelon form*. We therefore obtain

$$\begin{pmatrix} 1 & 0 & 3 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{pmatrix} \rightarrow \qquad \rightarrow$$

$$\rightarrow \qquad \rightarrow \qquad \rightarrow$$

Note that the "left hand half" of thus matrix is $I_{3,3}$. It turns out that $\begin{pmatrix} -1 & 3 & -4 \\ 1/3 & -1 & 5/3 \\ 2/3 & -1 & 4/3 \end{pmatrix}$ is the required (multiplicative) inverse of A, check this yourself!

Example 2.19. Find the multiplicative inverse of $C = \begin{pmatrix} 1 & 2 & 3 \\ -1 & -1 & 1 \\ -1 & -1 & 1 \end{pmatrix}$.

METHOD Form the 3×6 matrix and aim to row reduce it.

We get

$$\rightarrow \qquad \rightarrow \qquad .$$

There is clearly no way in which the "left hand half" of this matrix will be $I_{3,3}$ when we have row reduced it. Accordingly C *does not have* a multiplicative inverse.

The same method applies to (square) matrices of any size:

Example 2.20. Find the multiplicative inverse of $A = \begin{pmatrix} 7 & 8 \\ -2 & 3 \end{pmatrix}$ (if it exists!)

METHOD Form $B = \begin{pmatrix} 7 & 8 & 1 & 0 \\ -2 & 3 & 0 & 1 \end{pmatrix}$.Row operations change B to $\begin{pmatrix} 1 & 17 & 1 \\ -2 & 3 & 0 \end{pmatrix}$ then by to $\begin{pmatrix} 1 & 17 & 1 & 3 \\ 0 & 37 & 2 & 7 \end{pmatrix}$,by to $\begin{pmatrix} 1 & 17 & 1 & 3 \\ 0 & 1 & 2/37 & 7/37 \end{pmatrix}$ and, finally by $\rho_1' = \rho_1 - 17\rho_2$ to $\begin{pmatrix} 1 & 0 & 3/37 & -8/37 \\ 0 & 1 & 2/37 & 7/37 \end{pmatrix}$. In fact, if $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has a multiplicative then it is $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, as you may easily check!

The question arises as to why this works! To explain it we need a definition.

Definition 2.21. Any matrix obtained from an identity matrix by means of *one elementary row operation* is called an **elementary matrix.**

Example 2.22. (i) $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, (ii) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -(1/3\pi) \end{pmatrix}$, (iii) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \pi/31 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ are elementary and

(iv) $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, (v) $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -(1/3\pi) \end{pmatrix}$, (vi) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \pi/31 \\ 0 & 0 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{pmatrix}$ are not.

Remark 2.23. Each elementary matrix has an inverse. The inverses of (i), (ii), (iii) are

, , and respectively. Each "undoes"

the effect of the corresponding matrix in the previous Example.

We can (but will not) prove

Theorem 2.24. *Let* E *be an elementary* $n \times n$ *matrix and let* A *be* any $n \times r$ *matrix. Put* $B = EA$. *Then* B *is the* $n \times r$ *matrix obtained from* A *by applying* ***exactly the same row operation*** *which produced* E *from* $I_{n\times n}$.

We omit the proof, but here is an example.

Example 2.25. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 5/3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c & d \\ j & k & l & m \\ w & x & y & z \end{pmatrix} =$

We now use Theorem 2.24 repeatedly. Suppose that the $n \times n$ matrix A can be reduced to $I_{n\times n}$ by a succession of elementary row operations $e_1, e_2, \ldots, e_n$, say. Let $E_1, E_2, \ldots, E_n$ be the $n\times n$ elementary matrices corresponding to $e_1, e_2, \ldots, e_n$. Then $I_{n\times n} = E_n E_{n-1} \ldots E_2 E_1$ (note the order of matrices!). It follows that $A^{-1} = E_n E_{n-1} \ldots E_2 E_1$ is the multiplicative inverse for A, since $A^{-1}A = I_{n\times n}$. In other words, we have

Theorem 2.26. *If* A *is an invertible matrix then* A^{-1} *is obtained from* $I_{n\times n}$ *by applying* exactly the same sequence of elementary row operations *which will convert* A *into* $I_{n\times n}$.

So if we place A and $I_{n\times n}$ side by side and apply the same elementary transformations which convert A to $I_{n\times n}$ they make A^{-1} out of $I_{n\times n}$.

Example 2.27. Find, if it exists, the inverse of $A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 2 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$.

SOLUTION Form $B =$. Row reduction successively

gives

$\rightarrow$ $\rightarrow$

Hence A *is invertible* and $A^{-1} =$ (as you may CHECK!!! ALWAYS incorporate a CHECK into your solutions.)

Remark 2.28 (Historic).
(i) The concept of matrix inverses was introduced in Western mathematics by Arthur Cayley in 1855.
(ii) Matrix algebra was rediscovered around 1925 by physicists working on Quantum Mechanics.

3. Determinants

Determinants used to be central to equation solving. Nowadays somewhat peripheral in that context—but important in other areas.

Solving $\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 & = & b_1 \\ a_{21}x_1 + a_{22}x_2 & = & b_2 \end{array}$ we get $x_1 = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11}a_{22} - a_{21}a_{12}}$ and $x_2 = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{21}a_{12}}$, **provided** $a_{11}a_{22} - a_{12}a_{21} \neq 0$. So this denominator is an important number. To remind us where it has come from, namely the matrix of coefficients of the given system of equations, we denote it by $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$.

We call this NUMBER the **determinant** of A where A is the matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. We also denote it by $|A|$ and by $\det(A)$.

Example 3.1. $\begin{vmatrix} 7 & 8 \\ 9 & 10 \end{vmatrix} = 7 \cdot 10 - 9 \cdot 8 = -2$. The system $\begin{array}{rcl} 7x + 8y & = & -1 \\ 9x + 10y & = & 2 \end{array}$ has solution $x = \frac{\begin{vmatrix} -1 & 8 \\ 2 & 10 \end{vmatrix}}{\begin{vmatrix} 7 & 8 \\ 9 & 10 \end{vmatrix}}$, $y = \frac{}{\begin{vmatrix} 7 & 8 \\ 9 & 10 \end{vmatrix}}$, i.e. $x =$, $y =$.

For each $n \times n$ system of equations there is a similar formula. In the case of the 3×3 system

$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 & = & b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 & = & b_3 \end{array}$ we find that $x_1 = \dfrac{b_1 G_{11} - b_2 G_{21} + b_3 G_{31}}{a_{11}G_{11} - a_{21}G_{21} + a_{31}G_{31}}$ etc.[*provided that* $a_{11}G_{11} - a_{21}G_{21} + a_{31}G_{31} \neq 0$] where $G_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$, $G_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}$, $G_{31} = \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$.

Note that each G_{i1} is the 2×2 determinant obtained from the 3×3 matrix $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ by striking out the row and column in which a_{i1} (the multiplier of G_{i1}) lies. It can be shown that

$$a_{11}G_{11} - a_{21}G_{21} + a_{31}G_{31} = -a_{12}G_{12} + a_{22}G_{22} - a_{32}G_{32} = a_{13}G_{13} - a_{23}G_{23} + a_{33}G_{33}$$

and that these are also all equal to

$$a_{11}G_{11} - a_{12}G_{12} + a_{13}G_{13} = -a_{21}G_{21} + a_{22}G_{22} - a_{23}G_{23} = a_{31}G_{31} - a_{32}G_{32} + a_{33}G_{33}.$$

This common value of these six sums is the **determinant** of the matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, which we denote by $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$, by $\det A$, or by $|A|$.

Example 3.2. $\begin{vmatrix} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{vmatrix} = 2\cdot\mathbf{37} - 9\cdot\mathbf{38} + 4\cdot(-\mathbf{23}) =$. Use the (chessboard!) mnemonic $\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$ to memorise which sign $+$ or $-$ to use.

TO SUMMARISE

(i) We know how to evaluate determinants of order 2.
(ii) We have defined determinants of order 3 in terms of determinants of order 2.

3.1. **Definition by expansion.** We define determinants of order n in terms of determinants of order $n-1$ as follows. Given the $n \times n$ matrix $A = (a_{ij})$ the **minor** M_{ij} of the element a_{ij} is the determinant of the $(n-1)\times(n-1)$ matrix obtained from A by deleting both the i^{th} row and the j^{th} column of A. The **cofactor** of a_{ij} is $(-1)^{i+j} M_{ij}$.

The basic method of evaluating $\det A$ is as follows. Choose any one row or column — usually with as many zeros as possible to make evaluation easy! Multiply each element in that row (or column) by its corresponding cofactor. Add these results together to get $\det A$.

Example 3.3. Evaluate $\begin{vmatrix} 1 & -4 & 5 & -1 \\ -2 & 3 & -8 & 4 \\ 1 & 6 & 0 & -2 \\ 0 & 7 & 0 & 9 \end{vmatrix}$. Column 3 has two 0s in it so let us expand down column 3. We get

$$D = 5\cdot\begin{vmatrix} -2 & 3 & 4 \\ 1 & 6 & -2 \\ 0 & 7 & 9 \end{vmatrix} - (-8)\cdot\begin{vmatrix} 1 & -4 & -1 \\ 1 & 6 & -2 \\ 0 & 7 & 9 \end{vmatrix} + 0\cdot\begin{vmatrix} * & * & * \\ * & * & * \\ * & * & * \end{vmatrix} - 0\cdot\begin{vmatrix} * & * & * \\ * & * & * \\ * & * & * \end{vmatrix}$$

We know how to evaluate 3×3 determinants, so could finish calculation easily. However there exist rules which ease calculation even further.

3.2. **Effect of elementary operations, evaluation.** The following rules for determinants apply equally to *columns* as to *rows*. They are related to elementary row operations.

Rule 3.4. (i) *If one row (or column) of* A *is full of zeros then* $\det A = 0$. Demonstration for 3×3 matrices:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -0 \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix} + 0 \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix} - 0 \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix}$$

(ii) *If **one** row (or column) of* A *is multiplied by a constant* α *then so is* $\det A$. Demonstration for 3×3 matrices:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \alpha a_{31} & \alpha a_{32} & \alpha a_{33} \end{vmatrix} = \alpha a_{31} \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix} - \alpha a_{32} \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix} + \alpha a_{33} \cdot \begin{vmatrix} * & * \\ * & * \end{vmatrix} = \alpha \det A.$$

Example 3.5. $\begin{vmatrix} 77 & 12 \\ 91 & 14 \end{vmatrix} = \begin{vmatrix} 7 \cdot 11 & 2 \cdot 6 \\ 7 \cdot 13 & 2 \cdot 7 \end{vmatrix} = 7 \cdot 2 \cdot$

Remark 3.6. Beware: If A is $n \times n$ matrix then αA means that all entry (i.e. **each row**) is multiplied by α and then $\det(\alpha A) = \alpha^n \det(A)$.

(iii) *If two rows of the matrix are interchanged then the value of determinant is multiplied by* -1. Demonstration:

2×2 case: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$, $\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - da$. This is used to demonstrate the 3×3 case:

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

$$= -\left(a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \right) = -\det A$$

(iv) *If two rows (or columns) of a matrix are identical then its determinant is* 0.
Proof: if two rows of A are identical then interchanging them we got the same matrix A. But the previous rule tells that $\det A = -\det A$, thus $\det A = 0$.

Example 3.7.
$$\begin{vmatrix} 1 & 1 & 2 & 2 \\ 3 & 6 & 8 & 12 \\ -4 & 7 & 8 & 14 \\ 5 & 3 & 7 & 6 \end{vmatrix} = 2 \begin{vmatrix} 1 & 1 & 2 & 1 \\ 3 & 6 & 8 & 6 \\ -4 & 7 & 8 & 7 \\ 5 & 3 & 7 & 3 \end{vmatrix} = 2 \cdot 0 = 0.$$

Remark 3.8 (Historic). In Western mathematics determinants was introduced by a great philosopher and mathematician Gottfried Wilhelm von Leibniz (1646–1716). At the same time a similar quantities were used by Japanese Takakazu Seki (1642–1708).

(v) *If a multiple of one row is added to another the value of a determinant is* ***unchanged***. Demonstration:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21}+\alpha a_{31} & a_{22}+\alpha a_{32} & a_{23}+\alpha a_{33} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -(a_{21}+\alpha a_{31})\,|X| + (a_{22}+\alpha a_{32})\,|Y| - (a_{23}+\alpha a_{33})\,|Z|$$

$$= -a_{21}\,|X| + a_{22}\,|Y| - a_{23}\,|Z| + \alpha(-a_{31}\,|X| + a_{32}\,|Y| - a_{33}$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \alpha \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Remark 3.9. Thus our aim is: in evaluating a determinant try by elementary row operations to get as many zeros as possible in some row or column and then expand over that row or column.

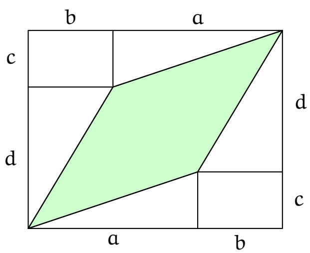
But before providing examples for this program we list few more useful rules.

(vi) *The determinant of a triangular matrix is the product of its diagonal elements.* Demonstration:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} = a_{11}\cdot \begin{vmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{vmatrix} = a_{11}\cdot a_{22}\cdot a_{33}, \text{ similarly } \begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\cdot a_{22}\cdot a_{33}.$$

(vii) *If* A *and* B *are* $n\times n$ *matrices then* $\det(AB) = \det(A)\det(B)$.On the first glance this is not obvious even for 2×2 matrices however there is a geometrical explanation. Any 2×2 matrix transform vectors by multiplication: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$. For any figure under such transformation we have:

$$\text{Area of the image} = \text{Area of the figure} \times \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$



Particularly for the image of the unit square with vertexes $(0,0)$, $(1,0)$, $(0,1)$, $(1,1)$ we have (see Figure 4):
$(a+b)(c+d) - 2bc - ac - bd = ad - bc$.

Figure 4. Area of the image.

Two such subsequent transformations with matrices A and B give the single transformation with matrix AB. Then property of determinants follows from the above formula for area. For 3×3 matrices we may consider volumes of solids in the space.

(viii) *The transpose matrix has the same determinant* $\det(A) = \det(A^T)$.This is easy to see for 2×2 matrices and could be extended for all matrices since we could equally expand determinants both by columns and rows.

Example 3.10. Evaluate $\begin{vmatrix} 4 & 0 & 2 \\ 3 & 2 & 1 \\ 7 & 5 & 1 \end{vmatrix}$. We have $a_{12} = 0$ and $\kappa_1' = \kappa_1 - 2\kappa_3$ makes $a_{11} = 0$ as well, then: $\begin{vmatrix} 0 & 0 & 2 \\ 1 & 2 & 1 \\ 5 & 5 & 1 \end{vmatrix} =$ .Another example of application of the Rule 3.8(v) about addition:

$$\begin{vmatrix} 1 & -1 & -1 & 5 \\ 2 & 0 & 1 & 4 \\ -3 & 2 & 7 & 3 \\ 4 & 1 & 4 & 6 \end{vmatrix} = \begin{vmatrix} 1 & -1 & -1 & 5 \\ 2 & 0 & 1 & 4 \\ -1 & 0 & 5 & 13 \\ & 0 & & \end{vmatrix} = -(-1) \cdot \begin{vmatrix} 2 & 1 & 4 \\ -1 & 5 & 13 \\ 5 & 3 & 11 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 0 \\ -11 & 5 & \\ -1 & 3 & \end{vmatrix} = -1 \cdot \begin{vmatrix} -11 & -7 \\ & \end{vmatrix}$$

The interchanging of rows or columns (Rule 3.6(iii)) might also be useful:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 \\ 3 & 3 & 4 & 5 \\ 4 & 5 & 6 & 7 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 1 & 2 & 3 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ & & & \end{vmatrix} = -1 \cdot 1 \cdot 1 \cdot 1.$$

3.3. **Some Applications.**

Example 3.11. Find all x such that $\begin{vmatrix} 0-x & 2 & 3 \\ 2 & 2-x & 4 \\ 3 & 4 & 4-x \end{vmatrix} = 0$. A direct expansion yields a cubic equation. To factorise it we observe that $x = -1$ implies $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 1 & \\ 2 & 1 & \\ 3 & 1 & \end{vmatrix} =$ 0, thus $x+1$ is a factor. *Alternatively* by : $\begin{vmatrix} 0-x & 2 & 3 \\ 2 & 2-x & 4 \\ 3 & 4 & 4-x \end{vmatrix} =$

$$\begin{vmatrix} -1-x & 2+2x & -1-x \\ 2 & 2-x & 4 \\ 3 & 4 & 4-x \end{vmatrix} = (1+x) \begin{vmatrix} -1 & 2 & -1 \\ 2 & 2-x & 4 \\ 3 & 4 & 4-x \end{vmatrix} = (1+x) \begin{vmatrix} -1 & & \\ 2 & 6-x & 2 \\ 3 & 10 & 1-x \end{vmatrix} =$$

$$-(1+x) \qquad = -(1+x)\left((6-x)(1-x) - 2\cdot 10\right) = -(1+x)(\qquad).$$

The last quadratic expression could be easily factorised.

Example 3.12. Find all values of λ which makes $\begin{vmatrix} 3-\lambda & 2 & 2 \\ 1 & 4-\lambda & 1 \\ -2 & -2 & -1-\lambda \end{vmatrix} = 0.$

The direct evaluation of determinants gives $-\lambda^3+6\lambda^2-9\lambda+4$. To factorise it we spot $\lambda = 1$ produces $\begin{vmatrix} 2 & 2 & 2 \\ 1 & 3 & 1 \\ & & \end{vmatrix} = 0$ (why?), i.e. $\lambda - 1$ is a factor which reduces the above cubic expression to quadratic. *Alternatively* by we have: $\begin{vmatrix} 3-\lambda & 2 & 2 \\ 1 & 4-\lambda & 1 \\ -2 & -2 & -1-\lambda \end{vmatrix} =$

$$\begin{vmatrix} & & \\ 1 & 4-\lambda & 1 \\ -2 & -2 & -1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} & & \\ 1 & 4-\lambda & 1 \\ -2 & -2 & -1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} 1 & 0 & \\ 1 & 4-\lambda & \\ -2 & -2 & \end{vmatrix} =$$

Remark 3.13. The last two determinants have the form $\det(A-\lambda I)$ for constant valued matrix A, indeed:

$$\begin{pmatrix} 3-\lambda & 2 & 2 \\ 1 & 4-\lambda & 1 \\ -2 & -2 & -1-\lambda \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -2 & -1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We will use them later to study *eigenvalues* of matrices.

Example 3.14. Factorise the determinant:

$$\begin{vmatrix} x & x^3 & x^5 \\ y & y^3 & y^5 \\ z & z^3 & z^5 \end{vmatrix} = xyz\begin{vmatrix} 1 & x^2 & x^4 \\ 1 & y^2 & y^4 \\ 1 & z^2 & z^4 \end{vmatrix} = xyz\begin{vmatrix} 1 & x^2 & x^4 \\ 0 & y^2-x^2 & y^4-x^4 \\ 0 & & \end{vmatrix} = xyz\begin{vmatrix} y^2-x^2 & (y^2-x^2)(y^2+x^2) \\ z^2-x^2 & \end{vmatrix}$$

$$= xyz(y^2-x^2)(z^2-x^2)\begin{vmatrix} 1 & y^2+x^2 \\ 1 & \end{vmatrix} = xyz(y^2-x^2)(z^2-x^2)\begin{vmatrix} 1 & y^2+x^2 \\ 0 & \end{vmatrix} =$$

If we swap any two columns in our matrices, then according to the Rule 3.6(iii) the determinant changes it sign. Could you see itfrom the last expression?

Example 3.15. Let us take an fancy approach to the well-known elementary result $x^2 - y^2 = (x-y)(x+y)$ through determinants:

$$x^2 - y^2 = \begin{vmatrix} x & y \\ y & x \end{vmatrix} = \begin{vmatrix} & \\ y & x \end{vmatrix} = (x+y)\begin{vmatrix} 1 & 1 \\ y & x \end{vmatrix} = (x+y)\begin{vmatrix} 1 & 0 \\ y & \end{vmatrix} =$$

We may use this approach to factorise $x^3+y^3+z^3-3xyz$, indeed:

$$x^3+y^3+z^3-3xyz = \begin{vmatrix} x & y & z \\ z & x & y \\ y & z & x \end{vmatrix} = \begin{vmatrix} & & \\ z & x & y \\ y & z & x \end{vmatrix} = (x+y+z)\begin{vmatrix} 1 & 1 & 1 \\ z & x & y \\ y & z & x \end{vmatrix}$$

$$= (x+y+z)\begin{vmatrix} 1 & 0 & 0 \\ z & x-z & y-z \\ y & & \end{vmatrix} = (x+y+z)\begin{vmatrix} x-z & y-z \\ z-y & x-y \end{vmatrix} = (x+y+z)((x-z)(x-y) - (y-z)(z-y))$$

$$=$$

Example 3.16. Show that a non-degenerate quadratic equation $px^2+qx+r=0$ could not have three *different* roots.

First restate the problem to make it *linear*: for given a, b, c find such p, q, r that we simultaneously have:

$$pa^2+qa+r=0,$$
$$pb^2+qb+r=0,$$
$$pc^2+qc+r=0.$$

That system has a non-zero solution only if its determinant is zero, but:
$$\begin{vmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{vmatrix} = \begin{vmatrix} & & 0 \\ b^2-c^2 & b-c & 0 \\ c^2 & c & 1 \end{vmatrix} = -\begin{vmatrix} a^2-c^2 & a-c \\ b^2-c^2 & b-c \end{vmatrix} = -\begin{vmatrix} & a-c \\ (b-c)(b+c) & b-c \end{vmatrix}$$
$$= -(a-c)(b-c)\begin{vmatrix} a+c & 1 \\ b+c & 1 \end{vmatrix} =$$
Thus the non-zero p, q, r could be found only if there at least two equal between numbers a, b, c.

Example 3.17. Let us evaluate the determinant of the special form:
$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \mathbf{i}(a_2b_3 - b_2a_3) - \mathbf{j}(a_1b_3 - b_1a_3) + \mathbf{k}(a_1b_2 - b_1a_2).$$
This is the well known *vector product*: $(\mathbf{i}a_1 + \mathbf{j}a_2 + \mathbf{k}a_3) \wedge (\mathbf{i}b_1 + \mathbf{j}b_2 + \mathbf{k}b_3)$ in three dimensional Euclidean space.

4. Real vector spaces and subspaces

4.1. **Examples and Definition.** There are many different objects in mathematics which have properties similar to *vectors* in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example 4.1. (i) Solutions (w, x, y, z) of the *homogeneous* system of linear equations,

for example:
$$\begin{aligned} w + 3x + 2y - z &= 0 \\ 2w + x + 4y + 3z &= 0 \; . \\ w + x + 2y + z &= 0 \end{aligned}$$

(ii) Functions which satisfy the *homogeneous* differential equation $\frac{d^2y}{dx^2} + 4\frac{dy}{dx} + 3y = 0$.
(iii) The set $M_{3\times 4}(\mathbb{R})$ of all 3×4 matrices.
(iv) A *real arithmetic progression*, i.e. a sequence of numbers $a_1, a_2, a_3, \ldots, a_n, \ldots$ with the constant difference $a_n - a_{n-1}$ for all n.
(v) A *real* 3×3 *magic squares*, i.e. a 3×3 array of real numbers such that the numbers in every row, every column and both diagonals add up to the same constant.
(vi) The set of all n-tuples of real numbers. The pairs of real numbers correspond to points on plane $\mathbb{R}^2$ and triples—to points in the space $\mathbb{R}^3$. The higher dimension vector spaces can be considered analytically and sometimes even visualised!

Vectors could be added and multiplied by a scalar according to the following rules, which are common for the all above (and many other) examples.

Axiom 4.2. We have the following properties of vector addition:

(i) $\mathbf{a} + \mathbf{b} \in V$, i.e. the *sum* of $\mathbf{a}$ and $\mathbf{b}$ is in V.
(ii) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$, i.e. the *commutative law* holds for the addition.
(iii) $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$, i.e. the *associative law* holds for the addition.
(iv) There is a special *null vector* $\mathbf{0} \in V$ (sometimes denoted by $\mathbf{z}$ as well) such that $\mathbf{a} + \mathbf{0} = \mathbf{a}$, i.e. V contains *zero vector*.
(v) For any vector $\mathbf{a}$ there exists the *inverse vector* $-\mathbf{a} \in V$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$, i.e. each element in V has an *additive inverse* or *negative*.

Axiom 4.3. There are the following properties involving multiplication by a scalar:

(i) $\lambda \cdot \mathbf{a} \in V$, i.e. *scalar multiples* are in V.
(ii) $\lambda \cdot (\mathbf{a} + \mathbf{b}) = \lambda \cdot \mathbf{a} + \lambda \cdot \mathbf{b}$, i.e. the *distributive law* holds for the vector addition.

(iii) $(\lambda + \mu) \cdot \mathbf{a} = \lambda \cdot \mathbf{a} + \mu \cdot \mathbf{a}$, i.e. the *distributive law* holds for the scalar addition.
(iv) $(\lambda\mu) \cdot \mathbf{a} = \lambda \cdot (\mu \cdot \mathbf{a}) = \mu \cdot (\lambda \cdot \mathbf{a})$, i.e. the *associative law* holds for multiplication by a scalar.
(v) $1 \cdot \mathbf{a} = \mathbf{a}$, i.e. the multiplication by 1 acts trivially as usual.

Remark 4.4. Although the above properties looks very simple and familiar we should not underestimate them. There is no any other ground to build up our theory: all further result should not be taken granted. Instead we should provide a *proof* for each statement based only on the above properties or other statements already proved in this way.

Because we need to avoid "chicken–egg" uncertainty some statements should be accepted without proof. They are called *axioms*. The axioms should be simple statements which

- do not *contradict* each other, more precisely we could not derive from them both a statement and its negation.
- be *independent*, i.e. any of them could not be derived from the rest.

Yet they should be sufficient for derivation of interesting theorems.

The famous example is the Fifth Euclidean postulate about parallel lines. It could be replaced by its negations and results in the beautiful *non-Euclidean geometry* (read about Lobachevsky and Gauss).

To demonstrate that choice of axioms is a delicate matter we will consider examples which violet these "obvious" properties.

On the set $\mathbb{R}^2$ of ordered pairs (x, y) define $+$ as usual, i.e. $(x, y) + (u, v) = (x + u, y + v)$, but define the multiplication by scalar as $\lambda \cdot (x, y) = (x, 0)$.

Since the addition is defined in the usual way we could easily check all corresponding axioms:

(i) $\mathbf{a} + \mathbf{b} \in V$.
(ii) $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.
(iii) $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$.
(iv) $\mathbf{a} + \mathbf{0} = \mathbf{a}$, where $\mathbf{0} =$.
(v) For any vector $\mathbf{a} = (x, y)$ there is $-\mathbf{a} =$ such that $\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$.

However for our strange multiplication $\lambda \cdot (x, y) = (x, 0)$ we should be more careful:

(i) $\lambda \cdot \mathbf{a} \in V$.
(ii) $\lambda \cdot (\mathbf{a} + \mathbf{b}) = \lambda \cdot \mathbf{a} + \lambda \cdot \mathbf{b}$, because$\lambda \cdot ((x, y) + (u, v)) =$
$\lambda \cdot (x, y) + \lambda \cdot (u, v) =$.
(iii) $(\lambda + \mu) \cdot \mathbf{a} \neq \lambda \cdot \mathbf{a} + \mu \cdot \mathbf{a}$, because$(\lambda + \mu) \cdot (x, y) =$ however $\lambda \cdot (x, y) + \mu \cdot (x, y) =$
.
(iv) $(\lambda\mu) \cdot \mathbf{a} = \lambda(\mu \cdot \mathbf{a}) = \mu(\lambda \cdot \mathbf{a})$, i.e. the *associative law* holds (why?)
(v) $1 \cdot \mathbf{a} \neq \mathbf{a}$ (why?)

This demonstrate that the failing axioms could not be derived from the rest, i.e they are independent.

Consider how some consequences can be derived from the axioms:

Theorem 4.5. *Let V be any vector space and let $v \in V$ and $t \in \mathbb{R}$, then:*

(i) *The null vector $\mathbf{0}$ is unique in any vector space;*
(ii) *For any vector $\mathbf{a}$ there is the unique inverse vector $-\mathbf{a}$;*

(iii) *If* $\mathbf{a} + \mathbf{x} = \mathbf{a} + \mathbf{y}$ *then* $\mathbf{x} = \mathbf{y}$ *in* V.
(iv) $0 \cdot \mathbf{a} = \mathbf{0}$;
(v) $t \cdot \mathbf{0} = \mathbf{0}$;
(vi) $(-t) \cdot \mathbf{v} = -(t \cdot \mathbf{v})$;
(vii) *If* $t \cdot \mathbf{v} = \mathbf{0}$ *then either* $t = 0$ *or* $\mathbf{v} = \mathbf{0}$.

Proof. We demonstrate the 4.5(i), assume there are two null vectors $\mathbf{0}_1$ and $\mathbf{0}_2$ then:

Similarly we demonstrate the 4.5(ii), assume that for a vector $\mathbf{a}$ there are two inverses vectors $-\mathbf{a}_1$ and $-\mathbf{a}_2$ then: □

4.2. **Subspaces.** Consider again the vector space $\mathbb{R}^4$ of *all* 4-tuples (w, x, y, z) from the Example 4.1(vi). The vector space H of all solutions (w, x, y, z) to a system of homogeneous equations from Example 4.1(vi) is a smaller subset of $\mathbb{R}^4$ since not all such 4-tuple are solutions. So we have one vector space H inside another $\mathbb{R}^4$.

Definition 4.6. Let V be a real vector space. Let S be a non-empty subset of V. If S itself a vector space (using the *same* operations of addition and multiplication by scalars as in V) then S is a *subspace* of V.

Theorem 4.7. *Let* S *be a subset of the vector space* V. S *will be a* subspace *of* V *(and hence a vector space on its own right) if and only if*

(i) *S is not the empty set.*
(ii) *For every pair* $\mathbf{u}, \mathbf{v} \in S$ *we have* $\mathbf{u} + \mathbf{v} \in S$.
(iii) *For every* $\mathbf{u} \in S$ *and every* $\lambda \in \mathbb{R}$ *we have* $\lambda \cdot \mathbf{u} \in S$.

Example 4.8. Let $V = \mathbb{R}^3$ and $S = \{(x, y, z), x + 2y - 3z = 0\}$. Then S is a subspace of V. Demonstration:

(i) S is not empty (why not?)
(ii) Suppose $\mathbf{u} = (a, b, c)$ and $\mathbf{v} = (d, e, f)$ are in S. Then $\mathbf{u}+\mathbf{v} =$ and $\mathbf{u}+\mathbf{v}$ is in S if $(a+d)+2(b+e)-3(c+f) = 0$.We know $a+2b-3c = 0$ (why?) and $d + 2e - 3f = 0$ (why?). Then adding them we get
i.e. $\mathbf{u} + \mathbf{v}$ is in S.
(iii) Similarly we could check that$\lambda \cdot \mathbf{u} =$, because $\lambda a + 2\lambda b - 3\lambda c = 0$.

Example 4.9. Let $V = \mathbb{R}^3$ and $S = \{(x, y, z), x + 2y - 3z = 1\}$. Is S a subspace of V? Demonstration:

(i) S is not empty, so we continue our check...
(ii) Suppose $\mathbf{u} = (a, b, c)$ and $\mathbf{v} = (d, e, f)$ are in S. Then as before $\mathbf{u} + \mathbf{v} = (a + d, b + e, c + f)$ and $\mathbf{u} + \mathbf{v}$ is in S only if $(a + d) + 2(b + e) - 3(c + f) = 1$. We again know $a+2b-3c = 1$ (why?) and $d+2e-3f = 1$ (why?). But now we could not deduce $(a + d) + 2(b + e) - 3(c + f) = 1$ from that! To disprove the statement we have to give a **specific counterexample**.

$$\in S \qquad \in S \qquad \text{but} \qquad \notin S.$$

(iii) There is no point to verify the third condition since the previous already fails!

Remark 4.10. (i) The null vector **0** should belong to any subspace. Thus the n-tuple $(0, 0, \ldots, 0)$ should lie in every subspace of $\mathbb{R}^n$. *If* $(0, 0, \ldots, 0)$ *does not belong to* S *(as in the above example, then* S *is not a subspace*.

(ii) *Geometrically*: In $\mathbb{R}^2$ all subspaces correspond to straight lines coming through the origin $(0, 0)$, see Figure 5;
In $\mathbb{R}^3$ subspaces correspond to planes containing the origin $(0, 0, 0)$. Otherwise

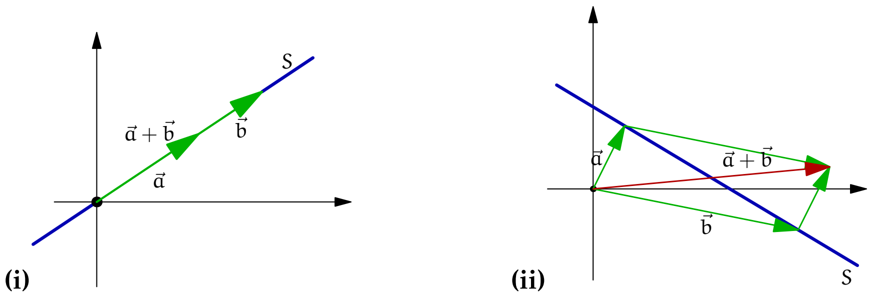


FIGURE 5. **(i)** A straight line coming through the origin is a subspace of the plane $\mathbb{R}^2$;
(ii) otherwise it is **not** a subspace;

they are not a subspaces of $\mathbb{R}^3$.

Example 4.11. Let $V = \mathbb{R}^4$ and $S = \{(x, y, z, t) : x \text{ is any integer}\}$. Is S a subspace of V?

(i) Is it non-empty? (give an example!);
(ii) Is any sum $\mathbf{u} + \mathbf{v}$ in S if both $\mathbf{u}$, $\mathbf{v}$ in S? (explain!);

(iii) Is any product $t \cdot \mathbf{v}$ in S if $\mathbf{v}$ in S and $t \in \mathbb{R}$? No, specific *counterexample:* $\mathbf{v} =$ and $t =$.

Conclusion: this is not a subspace.

Example 4.12. Let $V = \mathbb{R}^3$ and $S = \{(x, y, z) : xyz = 0\}$.
(i) Is it non-empty? (give an example!)
(ii) Is any sum $\mathbf{u} + \mathbf{v}$ in S if both $\mathbf{u}$, $\mathbf{v}$ in S? (explain!)

(iii) There is no need to verify the third condition!

Conclusion: this is not a subspace.

Example 4.13. Let V is the set of all 3×3 real matrices and S is the subset of symmetric matrices $\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$.
(i) Is it non-empty? (give an example!)
(ii) Is any sum $\mathbf{u} + \mathbf{v}$ in S if both $\mathbf{u}$, $\mathbf{v}$ in S? (explain!)

(iii) Is any product $t \cdot \mathbf{v}$ in S if $\mathbf{v}$ in S and $t \in \mathbb{R}$?

Conclusion: this is a subspace.

4.3. Definitions of span and linear combination.

Definition 4.14. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ be vectors in V. A vector in the form

$$\mathbf{v} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k$$

is called *linear combination* of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$.

The collection S of **all** vectors in V which are linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ is called *linear span* of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ and denoted by $\mathrm{sp}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$.

We also say S *is spanned* by the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ of vectors.

Example 4.15. (i) Vector $\mathbf{z} = (-2.01, 0.95, -0.07)$ is a linear combination of $\mathbf{x} = (8.6, 9.1, -7.3)$ and $\mathbf{y} = (6.1, 5.8, -4.8)$ since $\mathbf{z} =$.

(ii) $(2, -53, 11) = \quad \cdot (4, -1, 7) - \quad \cdot (3, 1, 5)$.

Exercise 4.16. Does $(1, 27, 29)$ belongs to $\mathrm{sp}\{(2, -1, 3), (-1, 6, 4)\}$?

Yes, provided there exist α and β such that $(1, 27, 29) = \alpha \cdot (2, -1, 3) + \beta \cdot (-1, 6, 4)$. That means that:

$$\begin{aligned} 2\alpha - \beta &= \\ -\alpha + 6\beta &= \\ 3\alpha + 4\beta &= \end{aligned}$$

which has a solutions $\alpha = 3$ and $\beta = 5$.

The last example and its question *does a vector belong to a linear span* bring us to the important notion which we are going to study now.

4.4. Linear Dependence and Independence.

Definition 4.17. A set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of vectors in a vector space V is said to be *linearly dependent* (LD) if there are real numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ *not all zero* such that:

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n = \mathbf{0}.$$

A set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ of vectors in a vector space V is said to be *linearly independent* (LI) if the *only* real numbers $\lambda_1, \lambda_2, \dots, \lambda_n$ such that:

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n = \mathbf{0},$$

are all zero $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.

Example 4.18. (i) Set $\{(2, 1, -4), (2, 5, 4), (1, 1, 3), (1, 4, 5)\}$ is linearly dependent since

$$\cdot (2, 1, -4) - \quad \cdot (2, 5, 4) + \quad \cdot (1, 1, 3) + \quad \cdot (1, 4, 5) = (0, 0, 0).$$

Note: some coefficients *could be* equal to zero but *not all of them!*

(ii) Is set $\{(-1, 1, 1), (1, -1, 1), (1, 1, -1)\}$ linearly independent in $\mathbb{R}^3$?
Yes, provided $\alpha(-1, 1, 1) + \beta(1, -1, 1) + \gamma(1, 1 -1) = (0, 0, 0)$ implies $\alpha = \beta = \gamma = 0$. This is indeed true since

$$\alpha(-1, 1, 1) + \beta(1, -1, 1) + \gamma(1, 1 - 1) =$$

Then we got a linear system

$$\begin{aligned} -\alpha+\beta+\gamma &= && \text{(i)} \\ \alpha-\beta+\gamma &= && \text{(ii)} \\ \alpha+\beta-\gamma &= && \text{(iii)} \end{aligned}$$

Then (i)+(ii) gives $2\gamma =$. Similarly $2\beta = 2\alpha =$.

(iii) The set $\{(\ , \ , \ , \ ), (\ , \ , \ , \ ), (0,0,0,0)\}$ is always linearly dependent since

$$0 \cdot (\ , \ , \ , \ ) + 0 \cdot (\ , \ , \ , \ ) + 1 \cdot (0,0,0,0) =$$

Here there are some non-zero coefficients!

Corollary 4.19. *Any set containing a null vector is linearly dependent.*

(iv) Is the set $\{(4,7,-1),(3,4,1),(-1,2,-5)\}$ linearly independent?
Solution: Look at the equation

$$\alpha \cdot (4,7,-1) + \beta \cdot (3,4,1) + \gamma \cdot (-1,2,-5) = (0,0,0).$$

Equating the corresponding coordinates we get the system:

$$\begin{aligned} 4\alpha+3\beta-\gamma &= \\ 7\alpha+4\beta+2\gamma &= \\ -\alpha+\beta-5\gamma &= \end{aligned}$$

Solving by the Gauss Eliminations gives $\gamma =$, $\beta =$, $\alpha =$. For $c = 1$ we get

$$\cdot (4,7,-1) + \quad \cdot (3,4,1) + \quad \cdot (-1,2,-5) = (0,0,0),$$

thus vectors are linearly dependent.
We will describe an easier way to do that later.

Here is a more abstract example.

Example 4.20. Let $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ be a linearly independent set of vectors in some vector space V. Show that the set $\{\mathbf{u}+\mathbf{v},\mathbf{v}+\mathbf{w},\mathbf{w}+\mathbf{u}\}$ is linearly independent as well.

Proof. It is given that $\alpha \cdot \mathbf{u} + \beta \cdot \mathbf{v} + \gamma \cdot \mathbf{w} = \mathbf{0}$ implies $\alpha = \beta = \gamma =$.
Suppose $a(\mathbf{u}+\mathbf{v}) + b(\mathbf{v}+\mathbf{w}) + c(\mathbf{w}+\mathbf{u}) = \mathbf{0}$ then

$$(\qquad)\mathbf{u} + (\qquad)\mathbf{v} + (\qquad)\mathbf{w} = \mathbf{0}.$$

Then $a+c=$, $a+b=$, and $b+c=$. This implies that $a=b=c=$.
Conclusion: the set $\{\mathbf{u}+\mathbf{v},\mathbf{v}+\mathbf{w},\mathbf{w}+\mathbf{u}\}$ is linearly independent. □

A condition for linear dependence in term of the linear combination is given by

Theorem 4.21. *The set $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is linearly dependent if and only if one of the $\mathbf{v}_k$ is a linear combination of the of its predecessors.*

Proof. If the set $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$ is linearly dependent then there such λ_1, λ_2, ..., λ_n such that:

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n = \mathbf{0}.$$

Let λ_k be the non-zero number with the biggest subindex k, then

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_k\mathbf{v}_k = \mathbf{0} \quad \text{(why?)}$$

But then $\mathbf{v}_k = -\frac{\lambda_1}{\lambda_k}\mathbf{v}_1 - \frac{\lambda_2}{\lambda_k}\mathbf{v}_2 - \cdots - \frac{\lambda_{k-1}}{\lambda_k}\mathbf{v}_{k-1}$ (why?)

Conversely, if for some k the vector $\mathbf{v}_k$ is a linear combination $\mathbf{v}_k = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_{k-1}\mathbf{v}_{k-1}$ then $\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_{k-1}\mathbf{v}_{k-1} - 1\cdot\mathbf{v}_k + 0\cdot\mathbf{v}_{k+1} + 0\cdot\mathbf{v}_n = \mathbf{0}$ (why?), i.e. vectors are linearly dependent (why?). □

Example 4.22. Considering again Example 4.18(i) the set of vectors $\{\mathbf{u} = (2,1,-4), \mathbf{v} = (2,5,4)$,
$\mathbf{w} = (1,4,5)\}$ is linearly dependent.

But $\mathbf{v}$ is not a linear combination of $\mathbf{u}$, but $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$, indeed

$(1,4,5) = \quad (2,1-4) + \quad (2,5,4)$.

Theorem 4.23. *Let A and B be row equivalent $m \times n$ matrices—that is, each can be obtained from the other by the use of a sequence of elementary row operations. Then, regarding the Rows of A as vectors on $\mathbb{R}^n$*

(i) *The set of rows (vectors) of A is linearly* independent *if and only if the set of rows (vectors) of B is linearly* independent*;*
(ii) *The set of rows (vectors) of A is linearly* dependent *if and only if the set of rows (vectors) of B is linearly* dependent*;*
(iii) *The rows of A span exactly the same subspace of $\mathbb{R}^n$ as do the rows of B.*

Definition 4.24. In the 4.23(iii) above the subspace of $\mathbb{R}^n$ spanned by the rows of A (and B!) is called *row space* of A (and B!).

Remark 4.25. So Theorem 4.23(iii) says: if A and B are row equivalent $m \times n$ matrices then A and B have same row space.

Example 4.26. Is the set of vectors $\{(1,3,4,-2,1), (2,5,6,-1,2), (0,-1,-2,3,0), (1,1,1,3,-1)\}$ linearly independent or linearly dependent?

Solution: Form the 4×5 matrix A using vectors as its rows: $A = \begin{pmatrix} 1 & 3 & 4 & -2 & 1 \\ 2 & 5 & 6 & -1 & \\ 0 & -1 & -2 & 3 & \\ 1 & 1 & 1 & 3 & \end{pmatrix}$,

then

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

B

The set of rows vectors of B is linearly dependent in $\mathbb{R}^4$ (why?)
Hence the given set of vectors is linearly dependent as well.

Similarly we could solve the following question.

Example 4.27. Is the set of vectors $\{(1,2,3,4), (5,6,7,8), (1,3,6,10)\}$ linearly independent or linearly dependent?

Solution: Form a 3×4 matrix A using the vectors as rows: $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 3 & 6 & 10 \end{pmatrix}$, then

$$A \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & 1 & 3 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 3 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{pmatrix} = B.$$

Rows of B form a linearly set in $\mathbb{R}^3$ (why?)
Form a linear combination of its rows with arbitrary coefficients:
$\cdot (1,2,3,4) + \cdot (0,1,2,3,) + \cdot (0,0,1,3) =$, then
$(\ , \ , 3\alpha + 2\beta + \gamma, 4\alpha + 3\beta + 3\gamma) =$. From the equality of first coordinates $\alpha =$, **then** from the equality of second coordinates $\beta =$, and **then** from the equality of third coordinates $\gamma =$.
Conclusion: rows of B are linearly independent and so are rows of A.

The above example illustrates a more general statement:

Theorem 4.28 (about dependence and zero rows). (i) *The rows of a matrix in the echelon form are linearly dependent if there is a row full of zeros.*
(ii) *The rows of a matrix in the echelon form are linearly independent if there is **not** a row full of zeros.*

Proof. The first statement just follows from the Corollary 4.19 and we could prove the second in a way similar to the solution of Example 4.27. □

We could also easily derive the following consequence of the above theorem.

Corollary 4.29. *In a matrix having echelon form any non-zero rows form a linearly independent set.*

Summing up: To study linear dependence of a set of vector:
(i) Make a matrix using vectors as rows;
(ii) Reduce it to the echelon form;
(iii) Made a conclusion based on the presence or absence of the null vector.

Exercise 4.30. Show that a set of m vectors in $\mathbb{R}^n$ is *always* linearly dependent if $m > n$.

4.5. **Basis and dimensions of a vector space.** Let us revise the Example 4.27. We demonstrate that the identity $\cdot (1,2,3,4) + \cdot (0,1,2,3,) + \cdot (0,0,1,3) =$ implies that $\alpha = \beta = \gamma =$. However the same reasons shows that an attempt of equality $\cdot (1,2,3,4) + \cdot (0,1,2,3,) + \cdot (0,0,1,3) =$ requires that $\alpha = \beta = \gamma =$, and thus is impossible (why?). We conclude that row vectors of the initial matrix A do not span the whole space $\mathbb{R}^4$. Yet if we add to the all non-zero rows of B the vector itself then they span together the whole $\mathbb{R}^4$ (why?). Then by the Theorem 4.23 the rows of A and the vector $(0,0,0,1)$ span $\mathbb{R}^4$ as well.

Similarly solving the homogeneous system $\begin{aligned} 3x + 2y - 2z + 2t &= 0 \\ 2x + 3y - z + t &= 0 \\ 5x - 4z + 4t &= 0 \end{aligned}$, we find a general solution of the form of the span of two vectors:

$$(x,y,z,t) = \left(-\frac{4c}{5} + \frac{4d}{5}, \frac{c}{5} - \frac{d}{5}, d, c\right) =$$

The similarity in the above examples is that to describe a vector space we may provide a small set of vectors (without linearly dependent), which span the whole space. It deserves to give the following

Definition 4.31. The set $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of vectors in the vector space V is *basis* for V if the both

(i) B is linearly independent;
(ii) B spans V.

Remark 4.32. (i) From the second part of the definition: any vector in V is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.
(ii) If we adjoin more vectors to a basis B then this bigger set still spans V, **but** now the set will be linearly dependent.
(iii) If we remove some vectors from a basis B then the smaller set still be linearly independent, **but** it will *not* span V any more.
(iv) A basis contain a smallest possible number of vectors needed to span V.

Remark 4.33. In this course we always assume that a basis B for a vector space V contain a finite number of vectors, then V is said to be *finite dimensional vector space*. However there are objects which represent a *infinite dimensional vector space*. Such spaces are studied for example in the course on Hilbert Spaces.

Example 4.34. All polynomials obviously form a vector space P. There is no possibility to find a finite number of vectors from P (i.e. polynomials) which span the entire P (why?). However any polynomial is a linear combinations of the monomials $1, x, x^2, \ldots, x^k, \ldots$. Moreover the monomials are linearly independent. Thus we could regard them as a basis of P.

Example 4.35. (i) The set $B = \{(1,0),(0,1)\}$ is a basis for $\mathbb{R}^2$ (called the natural basis of $\mathbb{R}^2$). Indeed
(a) B is linearly independent because ...
(b) B spans $\mathbb{R}^2$ because ...
Particularly $\mathbb{R}^2$ is finite dimensional and has infinitely many other bases. For example
$\{(\quad,\quad),(\quad,\quad)\}$, or $\{(\quad,\quad),(\quad,\quad)\}$.
(ii) Similarly the $B = \{(1,0,0,\ldots,0),(0,1,0,\ldots,0),(0,0,1,\ldots,0),\ldots,(0,0,0,\ldots,1)\}$ of n n-tuples is a basis for $\mathbb{R}^n$.
(a) B is linearly independent because ...

$$\alpha(1,0,0,\ldots,0)+\beta(0,1,0,\ldots,0)+\gamma(0,0,1,\ldots,0)+\cdots+\omega(0,0,0,\ldots,1)=$$

(b) B spans $\mathbb{R}^2$ because ...
(iii) The set $\{(1,1,0),(2,-1,3),(-1,2,-3)\}$ is **not** a basis for $\mathbb{R}^3$ because

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

i.e. the set is linearly dependent.

(iv) The set $\{(1,2,1),(2,4,5),(3,7,9),(1,0,4)\}$ is **not** a basis for $\mathbb{R}^3$ since it is linearly dependent:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 3 & 7 & 9 \\ 1 & 0 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & \\ 0 & -2 & \\ 0 & -3 & \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 3 \\ 0 & 0 & \\ & & \\ 0 & 0 & \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 3 \\ 0 & 0 & 3 \\ 0 & 0 & \end{pmatrix} \text{(why?)},$$

(v) $\{(1,1,2,3),(-1,2,1,-2),(4,1,5,6)\}$ is **not** a basis for $\mathbb{R}^4$, it is linearly independent:

$$\begin{pmatrix} 1 & 1 & 2 & 3 \\ -1 & 2 & 1 & -2 \\ 4 & 1 & 5 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 & 3 \\ 0 & 3 & 3 & \\ 0 & -3 & -3 & \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 2 & 3 \\ 0 & 3 & 3 & 1 \\ 0 & 0 & 0 & \end{pmatrix},$$

but does not span $\mathbb{R}^4$, because the vector is not in the row space (why?), because it is obviously

Remark 4.36. The last two examples suggests that

(i) **No** set of *four or more* vectors can be a basis for $\mathbb{R}^3$ since it must be a linearly independent set.

(ii) **No** set of *three or less* vectors can be a basis for $\mathbb{R}^4$ since they cannot span $\mathbb{R}^4$.

In fact we can prove the important

Theorem 4.37 (about dimension of bases)**.** *Let V be a finite dimensional vector space. If one basis B for V has k elements, then **all** basis for V have k elements.*

The proof of this theorem relies on the following another important results.

Theorem 4.38 (about spanning and independent sets)**.** *Let V be a vector space. If V can be spanned by a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ and if $I = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ is a linearly independent subset of V then $r \leqslant s$.*

Proof of the Theorem 4.37. (Easy) Let B be the given basis with k elements and let C be any other basis with, say, l elements.

- Then $k \leqslant l$ (why?) by the previous Theorem, since C span V and B is a linearly independent set in V;
- Then $l \leqslant k$ (why?) by the previous Theorem, since B span V and C is a linearly independent set in V;

Consequently $k = l$. □

Idea of a proof of Theorem 4.38. Clearly we could assume (why?) that none of $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_s$ is the null vector. Then we see that

$$\operatorname{sp}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\} \subseteq \operatorname{sp}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\} = V.$$

Now the set $\{\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ is linearly dependent (why?), then by the Theorem on linear dependence one of v_i is a linear combinations of preceding vectors $\{\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}\}$, thus the linear span of the set $\{\mathbf{w}_1, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ **without** $\mathbf{v}_i$ is still the entire space V. We could continue this procedure for following $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$, ... Each time we could kick out one vector, which could **not** be any $\mathbf{w}_k$ (why?), i.e. should be some $\mathbf{v}_j$, then s could not be less than r. □

Definition 4.39. The common number k of elements in **every** basis of V is called the *dimension* of V.

Example 4.40. The Euclidean space $\mathbb{R}^n$ is n-dimensional (i.e. has the dimension n), since its natural basis $\{(1,0,0,\ldots,0),\ (0,1,0,\ldots,0),\ (0,0,1,\ldots,0),\ldots,(0,0,0,\ldots,1)\}$ has n elements in it.

There are some important consequences of the previous Theorems.

Theorem 4.41. *Let V be a vector space of dimension n. Then*

(i) ***Every** set of $n+1$ (or more) vectors in V is linearly dependent.*
(ii) ***No set** of $n-1$ (or fewer) vectors in V can span V.*
(iii) *Each linearly independent set of n vectors in V must also span V and hence forms a basis of V;*
(iv) *Each of n vectors which spans V must also be linearly independent and hence forms a basis of V;*

Proof of 4.41(iv) only. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ spans V. Then S must be a linearly independent set (why?) because if **not** then, by the Theorem about condition for linear dependence at least one of the $\mathbf{v}_i$ is a linear combination of its predecessors and could be omitted without change of the span. Hence V could be spanned by less than n vectors—contradiction to the Theorem about dimension of basis. ☐

(i) Vectors $\{(\quad,\quad,\quad),(\quad,\quad,\quad),(\quad,\quad,\quad),(\quad,\quad,\quad)\}$ is linearly dependent in $\mathbb{R}^3$;
(ii) Vectors $\{(\quad,\quad,\quad,\quad,\quad),(\quad,\quad,\quad,\quad,\quad),(\quad,\quad,\quad,\quad,\quad),(\quad,\quad,\quad,\quad,\quad)\}$ cannot span $\mathbb{R}^5$;
(iii) If you somehow know that $S = \{(1,1,2,3),\ (-1,2,1,-2),\ (0,0,1,0),\ (4,1,5,6)\}$ is linearly independent subset of $\mathbb{R}^4$ (if you are not sure look at Example 4.35(v)) then you may conclude that it also *necessarily*, span $\mathbb{R}^4$—and so is a basis for $\mathbb{R}^4$.

Example 4.42. Let

$$A = \begin{pmatrix} 1 & 2 & 3 & 3 & 13 \\ 2 & 0 & 6 & -2 & 6 \\ 0 & 1 & 0 & 2 & 5 \\ 2 & -1 & 6 & -4 & 1 \end{pmatrix}$$

(i) Find a basis for the rows space of A, then find a second basis.
(ii) State the dimension of the row space of A;
(iii) Find a basis for the solution space of $A\mathbf{x} = \mathbf{0}$ and find its dimension.

Solution:

(i) $$A = \begin{pmatrix} 1 & 2 & 3 & 3 & 13 \\ 2 & 0 & 6 & -2 & 6 \\ 0 & 1 & 0 & 2 & 5 \\ 2 & -1 & 6 & -4 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 3 & 13 \\ 0 & 1 & 0 & 2 & 5 \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \end{pmatrix} = B.$$

By the Theorem about dependence and zero rows the set $\{(1,2,3,3,13),(0,1,0,2,5)\}$ forms a linearly independent set and by Theorem 4.23(iii) it spans the same subspace of $\mathbb{R}^5$ as do the four given vectors. Another basis is $\{(1,3,3,5,18),(0,1,0,2,5)\}$ (how do we obtain it?), and another is ....

(ii) The dimension of the row space of A is, therefore, 2 (why?)

(iii) Let $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \\ w \\ t \end{pmatrix}$ and $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Then solving $A\mathbf{x} = \mathbf{0}$ is exactly the same as solving $B\mathbf{x} = \mathbf{0}$. Now the general solution of the homogeneous system:

$$\begin{aligned} x + 2y + 3x + 3w + 13t &= 0 \\ y + 2w + 5t &= 0 \end{aligned}$$

is given by putting, for example, $t = \alpha$, $w = \beta$, so $y =$, then $z = \gamma$ (so that $x = -13\alpha - 3\beta - 3\gamma - 2(-5\alpha - 2\beta) =$. So the general solution is

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \\ w \\ t \end{pmatrix} = \begin{pmatrix} -3\alpha + \beta - 3\gamma \\ -5\alpha - 2\beta \\ \gamma \\ \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} -3 \\ -5 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Hence the set $S = \left\{ \begin{pmatrix} -3 \\ -5 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix} \right\}$ spans the solutions space of $A\mathbf{x} = \mathbf{0}$. It is also not difficult to see that S is also linearly independent (why?) Hence the solution space of $A\mathbf{x} = \mathbf{0}$ has the dimension .

Example 4.43. Find 3 solutions of $A\mathbf{x} = \mathbf{0}$ for which $y = 1$.

Solution: This could be achieved as follows: $\alpha = -\frac{1}{5}$, $\beta = \gamma = 0$ (actually γ could be any number); $\alpha = 0$, $\beta = -\frac{1}{2}$, $\gamma = 0$ (actually γ could be any number); $\alpha = 1$, $\beta = -3$, $\gamma = 0$ (actually γ could be any number);

Remark 4.44. In the above case $2 + 3 = 5$, i.e. the dimension of row space of A plus the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$ is equal to the dimension of the whole space $\mathbb{R}^5$. This is true in general as well.

4.6. Row Rank of a Matrix.

From the Theorem 4.37 about dimension of bases we can deduce:

In reducing a matrix A to echelon form we always get the same number of non-zero rows (why?), because that number is the **dimension** of the row space of A.

Definition 4.45. The dimension of the row space of a matrix A is called *row rank* of the matrix A.

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

Remark 4.46. If A is $m \times n$ matrix then

(i) the raw space of A is a subspace of $\mathbb{R}^n$.

(ii) the column space of A is a subspace of .

Thus it is *remarkable* that

Proposition 4.47. *The column rank of a matrix is equal to its row rank. This common number is called* rank *of* A.

Example 4.48. An illustration:

$$\begin{pmatrix}1 & 3 & 3 & 1\\ 3 & 1 & 5 & 11\\ 1 & 1 & 2 & 3\end{pmatrix} \to \begin{pmatrix}1 & 3 & 3 & 1\\ 0 & -8 & -4 & \\ 0 & -2 & -1 & \end{pmatrix} \to \begin{pmatrix}1 & 3 & 3 & 1\\ 0 & -8 & -4 & \\ 0 & & & \end{pmatrix}.$$

Thus row rank is .

$$\begin{pmatrix}1 & 3 & 3 & 1\\ 3 & 1 & 5 & 11\\ 1 & 1 & 2 & 3\end{pmatrix} \to \begin{pmatrix}1 & 0 & 0 & 0\\ 3 & -8 & -4 & 8\\ 1 & -2 & -1 & 2\end{pmatrix} \to \begin{pmatrix}1 & 0 & 0 & 0\\ 3 & -8 & & \\ 1 & -2 & & \end{pmatrix}$$

Thus column rank is .

We skip the proof of the last theorem and will state the another important result.

Theorem 4.49. *The* n *rows of an* $n \times n$ *matrix* A *form a linearly independent set if and only if and only if* $\det A \neq 0$.

Equivalently, the n *rows of an* $n \times n$ *matrix* A *form a linearly dependent set if and only if and only if* $\det A = 0$.

Proof. Let B be the echelon form of the matrix A obtained by the elementary row operations. Then by Theorem 4.23 both A and B have either *linearly dependent or linearly independent sets* of rows at the same time. The Rules 3.4 of evaluation of determinants also guarantee that both $\det A$ and $\det B$ are *zero or non-zero* at the same time. For the matrix B we clearly have the following alternative:

- All rows of B are non-zero, then all diagonal elements of B are equal to 1. In this case rows of B are linearly independent **and** $\det B = 1$.
- At least the last row of B is zero. In this case rows of B are linearly dependent **and** $\det B = 0$.

As pointed above the linear dependence of rows A and B and non-zero value of their determinants are simultaneous. Thus we also have these alternatives for the matrix A.

- Rows of A are linearly independent **and** $\det A \neq 0$.
- Rows of A are linearly dependent **and** $\det A = 0$.

This finishes the proof. □

5. Eigenvalues and eigenvectors

The following objects are very important in both theoretical and applied aspects. Let us consider the following example.

5.1. Use for Solving Differential Equations.

Example 5.1. Consider a differential equation like $\dot{X}(t) = kX(t)$, where $\dot{X}(t)$ denote the derivative of $X(t)$ with respect to t and k is a constant. The variable t is oftenly associated with a time parameter. The differential equation is easy to solve, indeed:

$$\frac{dX(t)}{dt} = kX(t) \equiv \frac{dX(t)}{X(t)} = k dt \equiv \int \frac{dX(t)}{X(t)} = \int k\, dt,$$

hence $\log X(t) = kt + c$ and $X(t) = e^{kt+c} = e^c e^{kt} = Ce^{kt}$, where $C = e^c$ is some constant.

A more interesting example is provided by a pair of differential equations:

Example 5.2. Consider a system

$$\left\{ \begin{array}{rcl} \dot{x}_1 & = & 7x_1 + 6x_2 \\ \dot{x}_2 & = & -9x_1 - 8x_2 \end{array} \right. \tag{5.1}$$

If a change of variables $(x_1, x_2) \to (X_1, X_2)$ could convert this system to a system like

$$\left\{ \begin{array}{rcl} \dot{X}_1 & = & qX_1 + 0 \cdot X_2 \\ \dot{X}_2 & = & 0 \cdot X_1 + rX_2, \end{array} \right. \tag{5.2}$$

then we could solve it in a ways similar to the previous Example and thus provide a solution for the original system. Let us consider such a change of coordinates. Let $\mathbf{u} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, then let $\dot{\mathbf{u}} = \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix}$ and $\dot{\mathbf{v}} = \begin{pmatrix} \dot{X}_1 \\ \dot{X}_2 \end{pmatrix}$. We also introduce matrices $A = \begin{pmatrix} 7 & 6 \\ -9 & -8 \end{pmatrix}$ and $D = \begin{pmatrix} q & 0 \\ 0 & r \end{pmatrix}$—being a diagonal matrix.

Then systems (5.1) and (5.2) became correspondingly

$$\dot{\mathbf{u}} = A\mathbf{u} \qquad \text{and} \qquad \dot{\mathbf{v}} = D\mathbf{v}. \tag{5.3}$$

If we assume that $\left\{ \begin{array}{rcl} x_1 & = & \alpha X_1 + \beta X_2 \\ x_2 & = & \gamma X_1 + \delta X_2 \end{array} \right.$, i.e. $\mathbf{u} = P\mathbf{v}$ where $P = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ then we may write (provided P has a multiplicative inverse!) $\mathbf{v} = P^{-1}\mathbf{u}$ and then

$$\dot{\mathbf{v}} = P^{-1}\dot{\mathbf{u}} = P^{-1}AP\mathbf{v}. \tag{5.4}$$

Comparing the systems (5.4) and (5.3) we see that we can get $\dot{\mathbf{v}} = D\mathbf{v}$ provided $D = P^{-1}AP$ for some suitable invertible matrix P.

Definition 5.3. If we can find such a P such that $D = P^{-1}AP$ for a diagonal matrix D we say that we can *diagonalise matrix* A.

If A and B such a matrices that $P^{-1}AP = B$ then $A = PBP^{-1}$ and A and B are called *similar matrices*.

5.2. Characteristic polynomial for eigenvalues.

Problem: Given a matrix A how could we find such a P which diagonalises A? **Solution:** from $D = P^{-1}AP$ we get $PD = AP$. In our example

$$\begin{pmatrix} 7 & 6 \\ -9 & -8 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} q & 0 \\ 0 & r \end{pmatrix} = \begin{pmatrix} \alpha q & \beta r \\ \gamma q & \delta r \end{pmatrix}.$$

We could split the last identity into two vector equations:

$$\begin{pmatrix} 7 & 6 \\ -9 & -8 \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha q \\ \gamma q \end{pmatrix} = \begin{pmatrix} q & 0 \\ 0 & q \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} \text{ and } \begin{pmatrix} 7 & 6 \\ -9 & -8 \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} = \begin{pmatrix} \beta r \\ \delta r \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix}.$$

In turn we could write them as

$$\begin{pmatrix} 7-q & 6 \\ -9 & -8-q \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 7-r & 6 \\ -9 & -8-r \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The last two equation are actually identical (apart of letters used in them) and represent a pair of simultaneous homogeneous linear equations. They have a non zero-solution provided $\begin{vmatrix} 7-q & 6 \\ -9 & -8-q \end{vmatrix} = 0$.

Thus to find P we should fist to find possible values for q (and r), which motivate the following definition.

Definition 5.4. Let A be an $n \times n$ matrix and let $\mathbf{x} \neq \mathbf{0}$ be a non-zero column vector such that $A\mathbf{x} = \lambda\mathbf{x}$ (i.e. $A\mathbf{x} = \lambda I\mathbf{x}$, i.e $(A - \lambda I)\mathbf{x} = \mathbf{0}$) for some real number λ. Then λ is called an *eigenvalue* for A and $\mathbf{x}$ is called an *eigenvector* for A corresponding to the eigenvalue λ.

Remark 5.5. (i) Note that the identity $A\mathbf{x} = \lambda\mathbf{x}$ does **not** implies that $A = \lambda$, since the later identity means $A\mathbf{y} = \lambda\mathbf{y}$ for any vector $\mathbf{y}$.

(ii) If $\mathbf{x}$ is an eigenvector of A with an eigenvalue λ then for any real number $t \neq 0$ the vector $t\mathbf{x}$ is also eigenvector of A with the eigenvalue λ. Indeed:

$$A(t\mathbf{x}) = t(A\mathbf{x}) = t(\lambda\mathbf{x}) = \lambda(t\mathbf{x}).$$

Therefore to determine the eigenvalues for A we just solve the equation

$$\det(A - \lambda I) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{12} - \lambda & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{pmatrix} = 0$$

Definition 5.6. The expression $\det(A - \lambda I)$ a polynomial of degree n in indeterminate λ, it is called the *characteristic polynomial* of A.

Remark 5.7. Both terms *eigen*value and *eigen*vector come from the following German word:

eigen—(Germ.) own, peculiar, peculiarly, to own.

Example 5.8. Consider matrix $A = \begin{pmatrix} \frac{5}{3} & -\frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. It has two eigenvectors $(1, 3)$ and $(2, 1)$ with corresponding eigenvalues $\frac{2}{3}$ and 1.5. See Figure 6 for illustration of this.

Example 5.9. $A = \begin{pmatrix} 7 & 6 \\ -9 & -8 \end{pmatrix}$ has eigenvalues given by solving $\begin{vmatrix} 7-\lambda & 6 \\ -9 & -8-\lambda \end{vmatrix} = \quad$, i.e. $(7-\lambda)(-8-\lambda) - (-9)6 = \quad$, i.e. $\lambda^2 + \lambda - 2 = \quad$, i.e. $\lambda = \qquad$. To find an eigenvector corresponding to $\lambda = 1$ we must solve

$$\begin{pmatrix} 7-1 & 6 \\ -9 & -8-1 \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} \quad \end{pmatrix}, \text{ i.e. } \begin{cases} 6\alpha + 6\gamma & = & 0 \\ -9\alpha - 9\gamma & = & 0 \end{cases} \text{ with solution } \alpha =$$

Thus corresponding to the eigenvalue $\lambda = 1$ the eigenvectors are $\begin{pmatrix} \alpha \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha \\ -\alpha \end{pmatrix} = \alpha \begin{pmatrix} \quad \end{pmatrix}$ for any non-zero real number α. For $\lambda = -2$:

$$\begin{pmatrix} 7-(-2) & 6 \\ -9 & -8-(-2) \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} = \begin{pmatrix} \quad \end{pmatrix}, \text{ i.e. } \begin{matrix} 9\beta + 6\delta & = & 0 \\ -9\beta - 6\delta & = & 0 \end{matrix} \text{ with solution } 2\delta =$$

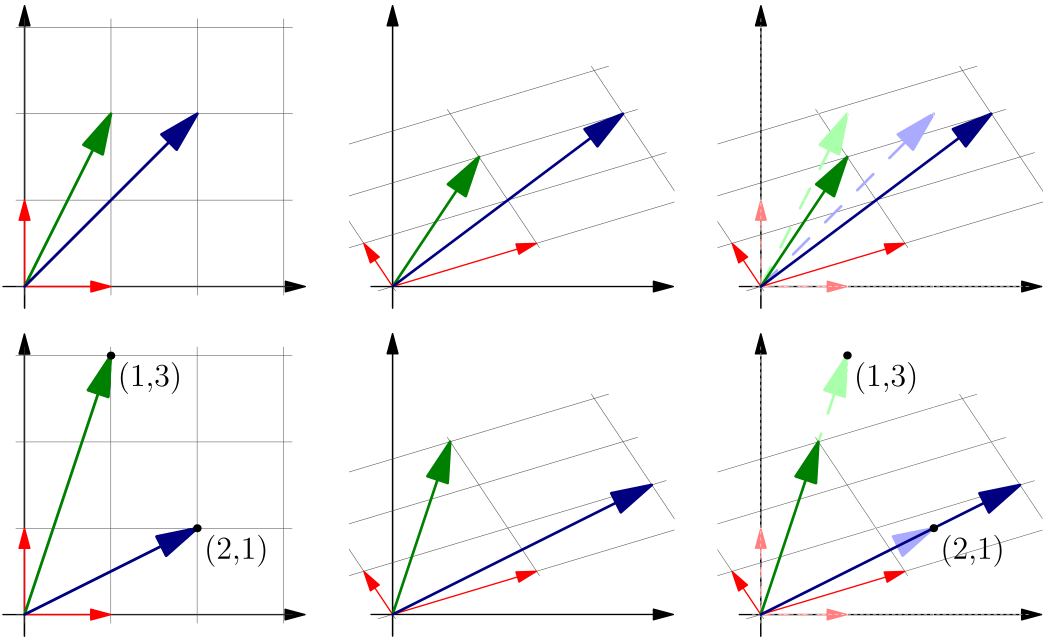


FIGURE 6. Eigenvalues and Eigenvectors

Thus corresponding to the eigenvalue$\lambda = -2$ the eigenvectors are $2\begin{pmatrix}\beta\\ \delta\end{pmatrix} = \begin{pmatrix}2\beta\\ -3\beta\end{pmatrix} = \beta\begin{pmatrix} \ \\ \ \end{pmatrix}$, with any real number $\beta \neq 0$.

Example 5.10. We can complete Example 5.2. Given $A = \begin{pmatrix}7 & 6\\ -9 & -8\end{pmatrix}$ we found eigenvalues $\lambda = 1, -2$. Then the transformation matrix $P = \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} = \begin{pmatrix}1 & \\ -1 & \end{pmatrix}$ is made out of the eigenvectors $\begin{pmatrix}1\\ -1\end{pmatrix}$ and $\begin{pmatrix}2\\ -3\end{pmatrix}$. We found its inverse $P^{-1} = \begin{pmatrix}3 & 2\\ & \end{pmatrix}$ and the matrix $D = \begin{pmatrix} & 0\\ 0 & \end{pmatrix}$ has eigenvalues on the diagonal. They related by $D = P^{-1}AP$.

Then from equation $\dot{\mathbf{v}} = D\mathbf{v}$ (5.3) we get $\left\{\begin{array}{lcl}\dot{X}_1 & = & 1\cdot X_1 + \ \cdot X_2\\ \dot{X}_2 & = & \ \cdot X_1 - 2\cdot X_2\end{array}\right.$ implying $X_1 = c_1e^t$ and $X_2 = c_2e^{-2t}$. Then from the identity $\mathbf{u} = P\mathbf{v}$ (5.4) we get the solution $\left\{\begin{array}{lcl}x_1 & = & c_1e \ + \ c\\ x_2 & = & c_1e\end{array}\right.$

The above methods works in general. Indeed let A be $n \times n$ matrix with $\lambda_1, \lambda_2, \ldots, \lambda_n$ its eigenvalues. Assume there exists corresponding eigenvectors $e_1, e_2, \ldots, e_n$ forming a linearly independent set in $\mathbb{R}^n$, which has to be a basis of $\mathbb{R}^n$.

Then on putting $P = (e_1, e_2, \dots, e_n)$ and $D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$ we find that $P^{-1}AP = D$.

If no such set of eigenvectors corresponding to $\lambda_1, \lambda_2, \dots, \lambda_n$ exists then A **cannot** be diagonalised.

Remark 5.11. Let A be $n \times n$ matrix. If $\mathbf{u}$, $\mathbf{v}$ are eigenvectors of A corresponding to the **same** eigenvalue λ then $\mathbf{u} + \mathbf{v}$ and $t\mathbf{u}$ (for any $t \neq 0$) are eigenvectors of A corresponding to λ (check!). Hence the set of all eigenvectors corresponding to λ—together with the zero vector—is a subspace of $\mathbb{R}^n$ called the *eigenspace* corresponding to eigenvalue λ.

Example 5.12. Find the eigenvalues and to each eigenvalues the full set of eigenvectors of the matrix $A = \begin{pmatrix} 3 & 11 & -11 \\ 1 & 3 & -2 \\ 1 & 5 & -4 \end{pmatrix}$.

Solution: first we solve $\begin{vmatrix} 3 - & 11 & -11 \\ 1 & 3 - & -2 \\ 1 & 5 & -4 - \end{vmatrix} =$. Evaluating determinant: $\begin{vmatrix} 3-\lambda & 11 \\ 1 & 3-\lambda \\ 1 & 5 \end{vmatrix}$

$= \begin{vmatrix} 3-\lambda & 11 & \\ 1 & 3-\lambda & \\ 1 & 5 & \end{vmatrix} = (1-\lambda) \begin{vmatrix} 3-\lambda & 11 & \\ 1 & 3-\lambda & \\ 1 & 5 & \end{vmatrix} = (1-\lambda) \begin{vmatrix} 3-\lambda & 11 & 0 \\ & -2-\lambda & \\ 1 & 5 & 1 \end{vmatrix} = (1-\lambda)\left((3-\lambda)(-2-\lambda)\right)$.

Thus eigenvalues are $\lambda = 1$, or -2, or 3.

For $\lambda = 1$ we solve $\begin{aligned} 2x + 11y - 11z &= 0 \\ x + 2y - 2z &= 0 \\ x + 5y - 5z &= 0 \end{aligned}$ and find $x = 0, y = z = \alpha$, i.e. eigenvectors are $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \alpha \begin{pmatrix} \end{pmatrix}$ where $\alpha \neq 0$.

For $\lambda = -2$ from $\begin{aligned} 5x + 11y - 11z &= 0 \\ x + 5y - 2z &= 0 \\ x + 5y - 2z &= 0 \end{aligned}$ eigenvectors are $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \beta \begin{pmatrix} \end{pmatrix}$, where $\beta \neq 0$.

For $\lambda = 3$ from $\begin{aligned} 0x + 11y - 11z &= 0 \\ x + 0y - 2z &= 0 \\ x + 5y - 7z &= 0 \end{aligned}$ eigenvectors are $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \gamma \begin{pmatrix} \end{pmatrix}$, where $\gamma \neq 0$.

Now the set $\left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 33 \\ -1 \\ 14 \end{pmatrix}, \begin{pmatrix} \\ \\ \end{pmatrix} \right\}$ is linearly independent in $\mathbb{R}^3$ (how do we know? see the next subsection!). So if we put $P = \begin{pmatrix} 0 & 33 & 2 \\ 1 & -1 & 1 \\ & & \end{pmatrix}$ and $D = \begin{pmatrix} & 0 & 0 \\ 0 & & 0 \\ 0 & 0 & \end{pmatrix}$ we have $P^{-1}AP = D$.

In particular the general solution of the of linear equations $\begin{array}{rcl} \dot{x}(t) & = & 3x + 11y - 11z \\ \dot{y}(t) & = & x + 3y - 2z \\ \dot{z}(t) & = & x + 5y - 4z \end{array}$

is given by $\begin{array}{rcl} x(t) & = & 0c_1e^t + 33c_2e^{-2t} + 2c_3e^{3t} \\ y(t) & = & 1c_1e^t - 1c_2e^{-2t} + 1c_3e^{3t} \\ z(t) & = & \end{array}$.

Summing Up:

To solve linear system of n equations we do the following steps:

(i) Form a matrix A from its coefficients.
(ii) Evaluate characteristic polynomial of A, i.e. $\det(A - \lambda I)$.
(iii) Find all eigenvalues $\lambda_1, \ldots, \lambda_n$ of A which are roots of the characteristic polynomial.
(iv) For all eigenvalues s find corresponding eigenvectors $e_1, \ldots, e_n$.
(v) Produce solutions $c_1e^{\lambda_1 t}, \ldots, c_ne^{\lambda_n t}$ of the diagonalised system of differential equations.
(vi) Make a matrix P using eigenvectors $e_1, \ldots, e_n$ as its columns.
(vii) Find the general solution of the initial system of differential equations in the form of product $P \begin{pmatrix} c_1e^{\lambda_1 t} \\ \vdots \\ c_ne^{\lambda_n t} \end{pmatrix}$.

5.3. Linear independence of eigenvectors for different eigenvalues.

Theorem 5.13. *If A is a $n \times n$ matrix and if A has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ then any corresponding set $\{e_1, e_2, \ldots, e_n\}$ of eigenvectors will be a linearly independent set in $\mathbb{R}^n$.*

Proof. By mathematical induction. The **base** $n = 1$: any non-zero vector e_1 form a linearly independent set.

The **step**: let any k-tuple of eigenvectors $e_1, e_2, \ldots, e_k$ with *different* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ are linearly independent, but there is e_{k+1} such that:
$e_{k+1} = a_1e_1 + a_2e_2 + \cdots + a_ke_k$ with some non-zero a_i. Then by linearity
$Ae_{k+1} = A(a_1e_1 + a_2e_2 + \cdots + a_ke_k) = \lambda_1a_1e_1 + \lambda_2a_2e_2 + \cdots + \lambda_ka_ke_k$.
Since $Ae_{k+1} = \lambda_{k+1}e_{k+1}$ we have two different expressions for e_{k+1}
$a_1e_1 + a_2e_2 + \cdots + a_ke_k = e_{k+1} = \frac{\lambda_1}{\lambda_{k+1}}a_1e_1 + \frac{\lambda_2}{\lambda_{k+1}}a_2e_2 + \cdots + \frac{\lambda_n}{\lambda_{k+1}}a_ke_k$, thus
$\left(1 - \frac{\lambda_1}{\lambda_{k+1}}\right)a_1e_1 + \left(1 - \frac{\lambda_2}{\lambda_{k+1}}\right)a_2e_2 + \cdots + \left(1 - \frac{\lambda_n}{\lambda_{k+1}}\right)a_ke_k = 0$. However this contradicts to the linear independence of $e_1, e_2, \ldots, e_k$. □

Example 5.14. Find eigenvalues and corresponding eigenvectors for the matrix $A = \begin{pmatrix} 3 & 1 \\ 2 & 4 \\ 1 & 1 \end{pmatrix}$
Is there an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

Solution: To find the eigenvalues of A we solve $= \begin{vmatrix} 3- & 1 & 1 \\ 2 & 4- & 2 \\ 1 & 1 & 3- \end{vmatrix} = \begin{vmatrix} 3-\lambda & & \\ 2 & 2-\lambda & \\ 1 & & \end{vmatrix}$

$$= (2-\lambda)\begin{vmatrix} 3-\lambda & 0 & 1 \\ 2 & & 2 \\ 1 & & 3-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 3-\lambda & 0 & 1 \\ 2 & 1 & 2 \\ & & \end{vmatrix} = (2-\lambda)\left((3-\lambda)(5-\lambda)-3\right)$$

$= (2-\lambda)(\lambda-2)(\lambda-6)$. Hence eigenvalues are 2, 2, 6, i.e. eigenvalue 2 has the *multiplicity*.

For $\lambda = 6$ we must solve $\begin{array}{rcl} -3x+y+z & = & \\ 2x-2y+2z & = & \\ x+y-3z & = & \end{array}$, and get $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \alpha \begin{pmatrix} \\ \\ \end{pmatrix}$, $\alpha \neq 0$.

For $\lambda = 2$ we must solve $\begin{array}{rcl} x+y+z & = & \\ 2x+2y+2z & = & \\ x+y+z & = & \end{array}$, which leads to $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\beta-\gamma \\ \gamma \\ \beta \end{pmatrix} =$
$\beta \begin{pmatrix} \\ \\ \end{pmatrix} + \gamma \begin{pmatrix} \\ \\ \end{pmatrix}$, β and γ are not both zero.

We see that all three eigenvectors $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$ form a linearly independent set in $\mathbb{R}^3$ and thus is a basis. Consequently setting $P = \begin{pmatrix} 1 & -1 & -1 \\ 2 & 0 & 1 \\ & & \end{pmatrix}$ and $D = \begin{pmatrix} & 0 & 0 \\ 0 & & 0 \\ 0 & 0 & \end{pmatrix}$ we get $P^{-1}AP = D$.

However if an eigenvalue has a multiplicity, the situation could be different.

Example 5.15. Can $A = \begin{pmatrix} 3 & 6 & 2 \\ 0 & -3 & -8 \\ 1 & 0 & -4 \end{pmatrix}$ be diagonalised? (i.e. does there exists an invertible matrix P and a diagonal D such that $P^{-1}AP = D$).

Solution: A has the characteristic equation $-\lambda^3 - 4\lambda^2 + 11\lambda - 6 = 0$ with eigenvalues $\lambda =$, , and .

Corresponding to $\lambda = -6$ we get eigenvectors $\alpha \begin{pmatrix} \\ \\ \end{pmatrix}$, where $\alpha \neq 0$.

Corresponding to $\lambda = 1$ we must solve $\begin{array}{rcl} 2x + 6y + 2z & = & 0 \\ -4y - 8z & = & 0 \\ x - 5z & = & 0 \end{array}$ giving $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \beta \begin{pmatrix} \quad \\ \quad \\ \quad \end{pmatrix}$, where $\beta \neq 0$. So only a single eigenvector spans the eigenspace corresponding to the double eigenvalue $\lambda = 1, 1$. Hence there do not exists matrices P, D as desired.

5.4. **Use in computing powers of matrix.** In many problems it is required to calculate a higher power of a matrix A. Eigenvectors provide an efficient way to do that.

If we can write $A = PDP^{-1}$ and then $A^k = PD^kP^{-1}$ since

$$A^k = PDP^{-1} \cdot PDP^{-1} \cdot \dots \cdot PDP^{-1} = PDIDI \cdots IDP^{-1} = PD^kP^{-1}.$$

Coming back to the matrix A from Exercise 5.14 we found:

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}, \; P = \begin{pmatrix} 1 & -1 & -1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \; D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \text{ and } P^{-1} = \tfrac{1}{4} \begin{pmatrix} -1 & -1 & 3 \\ -2 & 2 & -2 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then powers of A calculated by the formula $A^k = PD^kP^{-1}$ are:

$$\begin{aligned} A^k &= \frac{1}{4} \begin{pmatrix} 1 & -1 & -1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \quad & 0 & 0 \\ 0 & \quad & 0 \\ 0 & 0 & \quad \end{pmatrix} \begin{pmatrix} -1 & -1 & 3 \\ -2 & 2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \frac{1}{4} \begin{pmatrix} 3 \cdot 2^k + 6^k & -2^k + 6^k & -2^k + 6^k \\ -2 \cdot 2^k + 2 \cdot 6^k & 2 \cdot 2^k + 2 \cdot 6^k & -2 \cdot 2^k + 2 \cdot 6^k \\ \quad & \quad & \quad \end{pmatrix} \end{aligned}$$

Example 5.16. Find the element in the $(1, 2)$-th place of B^n where $B = \begin{pmatrix} 2 & 1 \\ 5 & 6 \end{pmatrix}$.

Solution: Equation $\begin{vmatrix} 2 - \quad & 1 \\ 5 & 6 - \quad \end{vmatrix} = 0$ yelds $\lambda^2 - 8\lambda + 12 - 5 = (\lambda - 1)(\lambda - 7) = 0$, so $\lambda = \quad , \quad .$

For $\lambda = 1$ we solve $\begin{array}{rcl} x + y & = & \quad \\ 5x + 5y & = & \quad \end{array}$ giving eigenvectors $c \begin{pmatrix} \quad \\ \quad \end{pmatrix}$, where $c \neq 0$.

For $\lambda = 7$ we solve $\begin{array}{rcl} -5x + y & = & \quad \\ 5x - y & = & \quad \end{array}$ giving eigenvectors $d \begin{pmatrix} \quad \\ \quad \end{pmatrix}$, where $d \neq 0$.

So $B^n = PD^nP^{-1} = \begin{pmatrix} 1 & 1 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} \quad & 0 \\ 0 & \quad \end{pmatrix} \frac{1}{6} \begin{pmatrix} 5 & -1 \\ 1 & 1 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} ? & \quad \\ ? & ? \end{pmatrix}$. Thus the answer is $\qquad$.

Appendix A. Additional Information

Remark A.1. The material of this Appendix is **not** in the syllabus of this course and is **not** examinable. But it worth to be known anyway!

A.1. **Why *Algebra*? (A historic Note).** It came from the Middle East together with *Arabic numbers* 1, 2, . . . (known, however from *India*). The name is formed from *al-jabr*, Arabic for "restoration," itself a transliteration of a Latin term, and just one of many contributions of Arab mathematicians.

Al-Khwarizmi (c.780-c.850), the chief librarian of the observatory, research center and library called the House of Wisdom in Baghdad produced the fundamental treatise, *"Hisab al-jabr w'al-muqabala"* (*"Calculation by Restoration and Reduction"*: widely used up to the 17th century), which covers linear and quadratic equations, was to solve trade imbalances, inheritance questions and problems arising from land surveyance and allocation—all *practical needs* raised by the civilisation!

Al-Karaji of Baghdad (953-c.1029), founder of a highly influential school of algebraic thought, defined higher powers and their reciprocals in his "al-Fakhri" and showed how to find their products. He also looked at polynomials and gave the rule for expanding a binomial, anticipating *Pascal's triangle* by more than *six centuries*.

Arab syntheses of Babylonian, Indian and Greek concepts also led to important developments in arithmetic, trigonometry and spherical geometry. The word *algorithm*, for instance, is derived from the name of al-Khwarizmi.

Another Arabic algebraist *Omar Khayyam* is also widely known for his poetry:

> Those who pursue the scientific way
> In a different language display
> Their ignorance and the way they pray.
> They too one day shall be dust and clay.

Exercise A.2. Why then bother to study anything?
Give your reasons!

A.2. **Linearisation of Natural Laws.** The almost any physical law studied at school is a linearised simplification of an (infinitely!) more complicated real process. Just few examples:

- **Hook's law:** Restoring force F of a spring *is proportional* to displacement x, i.e. $F = -kx$. (Untrue beyond elastisity)
- **Newton's Second Law:** Force $\vec{F}$ *is proportional* to caused acceleration $\vec{a}$ times (constant) mass m, i.e $\vec{F} = m\vec{a}$. (Untrue for high speeds in relativity)
- **Ohm's law:** Voltage V *is proportional* to current I times resistance R, i.e. $V = I \times R$. (Untrue for huge voltages—cause fire!)
- **Economics:** Profit from a mass production *is proportional* to the profit from a unit. (Untrue at a big scale, sparkle "globalisation") All these examples are manifestation of the same fundamental principle of mathematical analysis.
- **Analysis:** Increment $y - y_0$ of a function $y = f(x)$ *is proportional* to increment of $x - x_0$ times the derivative, i.e. $y = y_0 + f'(x_0)(x - x_0)$. (Untrue for not-so-small $x - x_0$—require higher derivatives!)

A.3. **Matrix Mechanics.** In the beginning of XX$^{\text{th}}$ century physicists tried to understand line spectrum of hydrogen atom. The revolutionary idea was that the electron may occupy fixed orbits only and it emits a photon during a spontaneous transition from one orbit to another.

	S_1	S_2	S_3	...
S_1	f_{11}	f_{21}	f_{31}	...
S_2	f_{12}	f_{22}	f_{32}	...
S_3	f_{13}	f_{23}	f_{33}	...
...	...	...	...	...

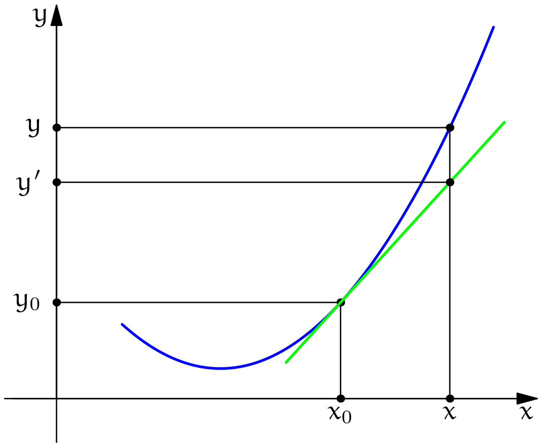


FIGURE 7. Linear part of a function increment gives a good approximation only for small $x - x_0$.

Heisenberg started out simply by tabulating the electron states of the hydrogen atom and the frequencies of photons that would be emitted by transitions between them, as shown above.

He went from this table of frequencies to develop corresponding tables of amplitudes, positions, and momenta, and began to painfully work out ways of performing calculations with them. He published a paper on it in July 1925, and also of course showed his work to his boss Born before leaving for Copenhagen to rejoin Bohr and his group. Born quickly saw the merit of Heisenberg's ideas and worked with one of his students, Pascual Jordan (1902:1980), to establish them on a more formal basis. Having two such transition matrices $[f_{ij}]$ and $[g_{ij}]$ we can calculate their composition according to the following rules, see Fig. 8:

(i) A resulting transition $S_i \to S_j$ can occur through any intermediate state S_k, e.g. $S_i \to S_1 \to S_j$, $S_i \to S_2 \to S_j$, etc.
(ii) The probability of a transition $S_i \to S_k \to S_j$ is $f_{ik}g_{kj}$.
(iii) Probability should be summed over all possible paths: $(fg)_{ik} = f_{i1}g_{1j} + f_{i2}g_{2j} + \cdots + f_{ik}g_{kj} + \ldots$

Born realized that Heisenberg's **sets of numbers could be represented as a square or rectangular grid of numbers known as a "matrix"**. Matrix math was already an established branch of mathematics, though it was not well known at the time, and in fact Born was one of the few physicists who understood it. Sets of matrix operations, including addition, subtraction, and multiplication, had been defined; Born and Jordan found that the rules of matrix mathematics could be directly applied to Heisenberg's ideas. **Heisenberg was not familiar with matrix math, but it is not a particularly complicated tool to learn how to use, and Heisenberg picked it up**. Read more...

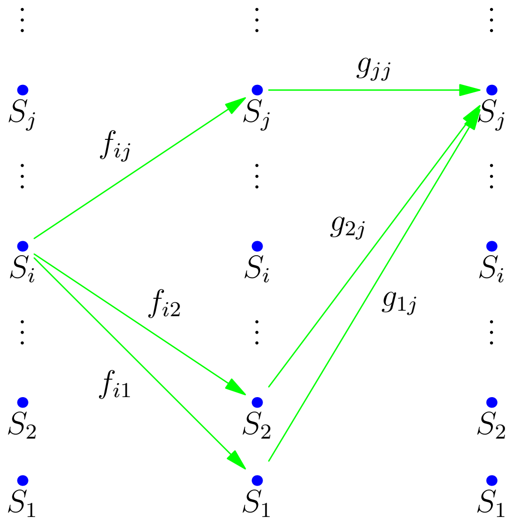


FIGURE 8. Transition amplitutdes and matrix multiplication

Index