Linear Algebra for MAS 162 and MAS 163

Sem. 2, 2001/02, Lecturer: Claas Röver

Lecture 1

1 Linear equations

1.1 One unknown

What are the solutions of the following equation?

$$a \cdot x = c, \qquad a, c \in \mathbb{R}. \tag{1}$$

In words, the set of solutions of (1) are all those real numbers which give a true statement when substituted for x in (1).

In order to find the solutions of (1), we can simply solve for x, and obtain $x = \frac{c}{a}$ if $a \neq 0$, as we are not allowed to divide by zero. But what if a is in fact zero? Then it clearly does not matter which value we substitute for x, we will always get $0 = c$. So if c is zero, then all real numbers are solutions of the equation. And if c is not zero, then the equation has no solutions.

Point to note. The equation (1) has
- a *unique* solution if $a \neq 0$, or
- *all real numbers* as solutions if $a = 0$ and $c = 0$, or
- *no* solution, if $a = 0$ and $c \neq 0$.

The equation (1) is the simplest example of a *linear equation*.

1.2 Two unknowns

A little more interesting are equations with two unknowns which have the general form

$$ax + by = c, \qquad a, b, c \in \mathbb{R}. \tag{2}$$

Again, we would like to know the solutions, i.e. all pairs of real numbers (s, t) which, when substituting s for x and t for y in (2), give a true statement.

It is clear that in case $a = 0$ it does not matter which value we assign to x, and the equation becomes $by = c$. And similarly, if $b = 0$ then every value will do for y and the equation becomes $ax = c$. In both cases we are then back in the situation of (1), but keep in mind that there was a second unknown whose value can be chosen freely.

Hence we assume now that a and b are non-zero. First observe that, if we fix a value for x, s say, then we are left to solve

$$as + by = c$$

which is the same as

$$by = d, \quad \text{where} \quad d = c - as \in \mathbb{R}.$$

And this has the same form as (1). In particular, we get that the solutions of (2) are all pairs $(s, \frac{c-as}{b})$ with $s \in \mathbb{R}$.

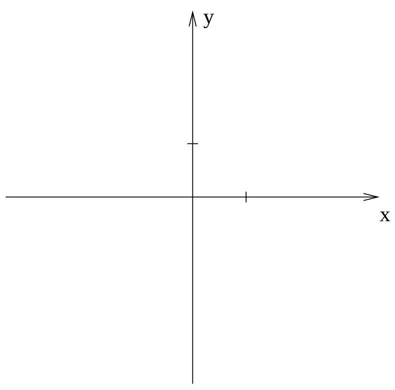
Point to note. The equation (2) has
- *no* solution, if $a = b = 0$ and $c \neq 0$, or
- *infinitely many* solutions, otherwise.

Why is there never a unique solution? Intuitively, the answer is that we have two unknowns but only one equation. So we can choose one unknown freely, and then the other one is determined uniquely, still assuming a and b are non-zero.

To express this dependence we could write $f(x)$ for y, and rewrite (2) as

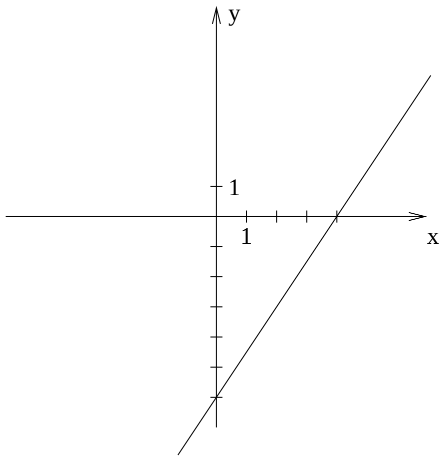
$$f(x) = \frac{c - ax}{b} = px - q, \quad \text{where} \quad p = -\frac{a}{b} \text{ and } q = \frac{c}{b}. \tag{3}$$

Now we know that this function describes a line in the usual coordinate system below, when we draw all the points $(x, f(x))$, $x \in \mathbb{R}$.



Conclusion. The solutions of (2) with a and b non-zero are all points on the line given by (3).

Example. The solutions of $3x - 2y = 12$ are all pairs $(x, 1.5x - 6)$, $x \in \mathbb{R}$. Or all points in the plane on the line given by $f(x) = 1.5x - 6$, which is drawn below.



Remark. Even if only one of a and b is non-zero, we can interpret the solutions of (2) as a line. This time it is

- a horizontal line if $a = 0$ and $b \neq 0$, or
- a vertical line if $b = 0$ and $a \neq 0$.

So whenever (2) has a solution, then it has infinitely many solutions which we can think of as the points on a certain line in the plane.

1.3 Two equations

The upshot of the previous section is that if we have two unknowns, then we need at least two equations to determine a unique solution. So suppose we are given the following *system of linear equations*

$$\begin{cases} a_1 x + b_1 y = c_1 \\ a_2 x + b_2 y = c_2 \end{cases} \tag{4}$$

where $a_1, a_2, b_1, b_2, c_1, c_2 \in \mathbb{R}$.

The solutions of (4) are now all pairs of real numbers which solve both equations simultaneously. Clearly, for a solution to exist, each of the two equations must have a solution, in which case the solutions of each equation are the points on a line in the plane. It follows now, that the solution of the system of equations (4) is the intersection of those two lines. But two lines in the plane could be parallel, in which case there is no solution, or the two lines could be identical, in which case there are infinitely many solutions.

Conclusion. The system of equations (4) has
- a unique solution, or
- infinitely many solutions, or
- no solution at all.

Examples.

$(a)\begin{cases} 2x-3y=7 \\ 3x+5y=1 \end{cases}$ has a unique solution, $x=2$, $y=-1$.

$(b)\begin{cases} 2x-3y=1 \\ 4x-6y=2 \end{cases}$ has infinitely many solutions $x=s$, $y=\frac{2}{3}x-\frac{1}{3}$.

$(c)\begin{cases} 2x-3y=1 \\ 4x-6y=3 \end{cases}$ has no solution: the equations are *inconsistent*.

1.4 The general case

In the course of this lecture we will learn how to solve *systems of linear equations* with m equations and n unknowns $x_1, x_2, \ldots, x_n$, m and n are integers ≥ 1 (which can be very large!). Such a system can be written as

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n} = b_2 \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in} = b_i \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn} = b_m \end{cases}$$

Here each a_{ij} is a real (or sometimes complex) constant, and so is each b_i. It is worth remembering the meaning of the indecees: a_{ij} is the *coefficient* of x_j in the ith equation.

Lecture 2

2 Gaussian elimination

Two systems of equations are called *equivalent* if they have the same solution set; in particular they must have the same number of unknowns.

The following three types of operations transform a system of linear equations into an equivalent one.

1. Multiplying one equation by a **non-zero** real (maybe complex) number.
2. Replacing the ith equation by the sum of the ith and the jth equation.
3. Reordering the equations.

Now consider the system of m linear equations in n unknowns given by

$$\left\{\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array}\right. \tag{5}$$

First reorder the equations such that a_{11} is non-zero; we always assume that the coefficient of x_j in the ith equation is a_{ij}. This is possible because the original system is assumed to be in n unknowns, so that in at least one of the equations x_1 has non-zero coefficient.

Then subtract $\frac{ai1}{a_{11}}$ times the first equation from the ith equation. After that the $i$the equation has zero coefficient for x_1. And we do this for all i with $2 \leq i \leq m$.

Now, the system consisting of all but the first equations, S say, has one less equation than the original system and is in one less unknown. So it is in some sense simpler and we restrict our attention to that system next. This makes sense, for if we have a solution $(s_2, \ldots, s_n)$of S, then we can find the s_1 which makes $(s_1, s_2, \ldots, s_n)$ a solution of the original system from the first equation, that is $s_1 = \frac{1}{a_{11}}(b - a_{12}s_2 - \cdots - a_{1n}s_n)$.

Applying the same procedure to S and so on, we eventually transform the system (5) into one with $a_{ij} = 0$ whenever $i < j$. And it comes as no surprise that there are three possibilities, $n < m$, $n = m$, and $n > m$.

If $n < m$, i.e. there are less unknowns than equations, then after $n-1$ reductions, we are left with a system in which the last $m-n+1$ equations only involve x_n. It is then often the case that the system will have no solutions at all because these equations contradict each other, eg $x_n = 2$ and $3x_n = 5$.

If, on the other hand, $n > m$, then the last equation in the transformed system may still have non-zero coefficients for more than one variable, and we find infinitely many solutions in that case.

Finally, if $n = m$, then it can happen that there is precisely one solution, although infinitely many or none is still possible because

Remark. Note that it can happen in this process that, when we eliminate the unknown x_i another unknown also completely disappears from the new smaller subsystem.

This procedure is called *Gaussian elimination*. It is probably best understood by looking at
Examples. (a) Consider the system of equations

$$\left\{\begin{array}{lllll} x & +y & -2z & +2u & =3 \\ 2x & +y & -z & +2u & =5 \\ 3x & +y & +z & -2u & =4 \\ 4x & +2y & -3z & +u & =6 \end{array}\right.$$

Subtracting twice the first eq. from the second, three times the first from the third, and four times the first from fourth yields

$$\left\{\begin{array}{lllll} x & +y & -2z & +2u & =3 \\ & -y & +3z & -2u & =-1 \\ & -2y & +7z & -8u & =-5 \\ & -2y & +5z & -7u & =-6 \end{array}\right.$$

Now subtract twice the second eq. from the third and fourth to get

$$\left\{\begin{array}{lllll} x & +y & -2z & +2u & =3 \\ & -y & +3z & -2u & =-1 \\ & & +1z & -4u & =-3 \\ & & -1z & -3u & =-4 \end{array}\right.$$

Finally, add twice the third eq. to the fourth, which results in

$$\left\{\begin{array}{lllll} x & +y & -2z & +2u & =3 \\ & -y & +3z & -2u & =-1 \\ & & z & -4u & =-3 \\ & & & -7u & =-7 \end{array}\right.$$

So $u = 1$, which when substituted into the third eq. implies $z = 1$. Now the second eq. tells us that $y = 2$, and then $x = 1$ follows from the first equation. Hence the solution is $(1, 2, 1, 1)$.

(b) Consider the system

$$\begin{cases} 2x & +3y & -z & = -5 \\ x & -2y & +4z & = 1 \\ 5x & 4y & +2z & = -9 \end{cases}$$

First interchange the first and second equation.

$$\begin{cases} x & -2y & +4z & = 1 \\ 2x & +3y & -z & = -5 \\ 5x & 4y & +2z & = -9 \end{cases}$$

Now subtract twice the first from the second and five times the first from the third to get

$$\begin{cases} x & -2y & +4z & = 1 \\ & 7y & -9z & = -7 \\ & 14y & -18z & = -14 \end{cases}$$

Finally, subtracting twice the second from the third gives.

$$\begin{cases} x & -2y & +4z & = 1 \\ & 7y & -9z & = -7 \\ & & 0 & = 0 \end{cases}$$

Now what does this mean, $0 = 0$? This is just to indicate that there was actually a third equation. But since it was the same as twice the second it has no influence on the solution set whatsoever.

The second equation tells us that $y = \frac{9}{7}z - 1$. And then the first equation says that $x = 1 + 2(\frac{9}{7}z - 1) - 4z = -1 - \frac{10}{7}z$. So we get the following parametrised description of all the solutions: $(-1 - \frac{10}{7}s, \frac{9}{7}s - 1, s)$. That is to say that all those triples with $s \in \mathbb{R}$ are a solution of the given system of equations.

3 Homogeneous vs inhomogeneous

Again suppose we are given a system S of linear equations as in (5). Assume also that $(r_1, \ldots, r_n)$ and $(s_1, \ldots s_n)$ are solutions of S.

Is $(r_1 - s_1, \ldots, r_n - s_n)$ also a solution of S? Well, in general not. This is because we have

1. $a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n = b_1$, as $(s_1, \ldots s_n)$ is a solution,
2. $a_{11}r_1 + a_{12}r_2 + \cdots + a_{1n}r_n = b_1$, as $(r_1, \ldots r_n)$ is a solution, and
3. $a_{11}(s_1 - r_1) + a_{12}(s_2 - r_2) + \cdots + a_{1n}(s_n - r_n) = b_1$, if $(r_1 - s_1, \ldots, r_n - s_n)$ was a solution.

Since the left hand side of 3. equals

$$(a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n) - (a_{11}r_1 + a_{12}r_2 + \cdots + a_{1n}r_n) = b_1 - b_1,$$

by 1. and 2., $(r_1 - s_1, \ldots, r_n - s_n)$ can only be a solution if $b_1 = 0$, and similarly all the b_i have to be zero for $1 \leq i \leq m$.

Conclusion. If one of the b_i is non-zero, then $(r_1 - s_1, \ldots, r_n - s_n)$ is not a solution of S.

But the argument shows that $(r_1 - s_1, \ldots, r_n - s_n)$ is a solution of the system

$$\left\{ \begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{array} \right. \tag{6}$$

which is called the *homogeneous* system associated to (5). It follows from this that

All solutions of the inhomogeneous system (5) are of the form

$$(s_1 + q_1, \ldots, s_n + q_n),$$

where $(s_1, \ldots s_n)$ **is a fixed solution of (5) and** $(q_1, \ldots q_n)$ **is an arbitrary solution of the associated homogeneous system (6)**.

Why bother? Because (6) is a little easier to handle in the Gaussian elimination process (all right hand sides of the equations will stay zero).

Moreover, when $(r_1, \ldots, r_n)$ and $(s_1, \ldots s_n)$ are solutions of (6) then so are $(r_1 + s_1, \ldots, r_n + s_n)$ and $(\lambda r_1, \ldots, \lambda r_n)$ for all real (complex) numbers. So the solution set of the homogeneous system has nicer properties than the solution set of the inhomogeneous system.

4 Matrix and vector notation

Instead of carrying along all the x_i's while solving a system of equations, we could simply deal with the coefficients if we make sure we know in front of which unknown in which equation a given coefficient appears. But this is already achieved by our choice of indices in (5); the coefficient of x_i in the jth equation is a_{ji}.

Hence it is convenient to denote (6) by the scheme

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{7}$$

Such a scheme is called a *matrix*, more precisely, a matrix of type m by n, or an $m \times n$-matrix.

The inhomogeneous system (5) is often denoted by the scheme

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} & b_i \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right) \tag{8}$$

where the vertical line indicates that the entries to the right of it are not coefficients of unknowns but rather the right hand sides of the equations.

A *column vector* is now simply a $k \times 1$-matrix, and a *row vector* is a $1 \times k$-matrix, eg.

$$\begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_k \end{pmatrix}, \qquad \text{and} \qquad (r_1, r_2, \ldots, r_k)$$

are a column vector with n rows and a row vector with k columns, respectively.

The *product of a row vector by a column* is defined **if and only if** the they both have the same number of entries, say n, in which case it is defined by

$$(r_1, r_2, \ldots, r_n) \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = r_1 s_1 + r_2 s_2 + \cdots + r_n s_n. \tag{9}$$

Remark. Observe that this product is a real (or complex) number, or if you like a 1×1-matrix.

Extending this definition, by viewing a matrix as a collection of row vectors, we define the product of the matrix in (7) with the column vector $\underline{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}$ by

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_i \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n} \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n} \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in} \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn} \end{pmatrix}, \tag{10}$$

which is another column vector; its first entry is the product of the first row of the matrix with $\underline{s}$, its second entry is the product of the second row of the matrix with $\underline{s}$ and so on.

In particular, if we denote the matrix in (7) by A, and agree to write $\underline{y}$ for $\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ whatever letter y might be, then the system of equations (5) may be written as $A\underline{x} = \underline{b}$.

Lecture 3

Matrix and vector notation continued

We can now define the product of two matrices, by viewing the second matrix as collection of column vectors. In particular, the first matrix must have as many columns as the second matrix has rows for the product to be defined. The general formula is a bit complicated but in words it is as follows.

The entry in the i-the row and j-the column of the product AB is the product of the i-th row of A by the j-th column of B. In particular, the product of an $m \times n$-matrix by an $n \times k$-matrix is of type $m \times k$.

$$\boxed{(m \times n\text{-matrix})(n \times k\text{-matrix}) = m \times k\text{-matrix}} \tag{11}$$

Examples.

$$1) \qquad \begin{pmatrix} 2 & -1 & 3 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ -1 & 1 \\ -2 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 7 \\ -4 & 4 \end{pmatrix}$$

$$2) \qquad \begin{pmatrix} 2 & -1 & 3 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 3 \\ 0 & 2 & 1 \end{pmatrix} \quad \text{is not defined!!}$$

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

4.1 Properties of matrix multiplication

Matrix multiplication shares some but not all properties with ordinary multiplication of complex (real) numbers. The first difference to note is that the product of two matrices may not be defined. Recall the following

Facts. Multiplication of real (complex) numbers satisfies

- *associativity*; for all $a, b, c \in \mathbb{R}$, $a(bc) = (ab)c$.
- *commutativity*; for all $a, b \in \mathbb{R}$, $ab = ba$.
- *existence of identity*; there is $1 \in \mathbb{R}$ so that for all $a \in \mathbb{R}$, $a \cdot 1 = 1 \cdot a = a$.
- *existence of inverse*; for all $a \in \mathbb{R}$ there is $b \in \mathbb{R}$ such that $ab = ba = 1$.
- *existence of zero*; there is $0 \in \mathbb{R}$ so that for all $a \in \mathbb{R}$, $a \cdot 0 = 0 \cdot a = 0$.
- *no zero divisors*; for all non-zero $a, b \in \mathbb{R}$, $ab \neq 0$.

4.1.1 Associativity

***Theorem* 1** *Matrix multiplication is associative, i.e. for all matrices A, B, and C, $A(BC) = (AB)C$.*

Proof. The first thing to observe is that, by (11), both sides of the equality are either both defined or they are both not defined.

For a matrix A let A_{ij} denote its entry in the i-th row and j-th column. Then matrix multiplication can be expressed in the following way.

Let A be an $m \times n$-matrix, and let B be an $n \times k$ matrix. Then for every i and j with $1 \leq i \leq m$ and $1 \leq j \leq k$,

$$(AB)_{ij} = \sum_{p=1}^{n} A_{ip} B_{pj} = (i^{\text{th}} \text{ row of } A)(j^{\text{th}} \text{ column of } B),$$

Now let C be a $k \times l$-matrix. Then for all i and j with $1 \leq i \leq m$ and $1 \leq j \leq l$

$$\begin{aligned}
(A(BC))_{ij} &= \sum_{p=1}^{n} A_{ip}(BC)_{pj} && (12)\\
&= \sum_{p=1}^{n} A_{ip}\left(\sum_{q=1}^{k} B_{pq}C_{qj}\right) && (13)\\
&= \sum_{p=1}^{n}\sum_{q=1}^{k} A_{ip}(B_{pq}C_{qj}) && (14)\\
&= \sum_{p=1}^{n}\sum_{q=1}^{k} (A_{ip}B_{pq})C_{qj}, \quad \text{using associativity of } \mathbb{R}, && (15)\\
&= \sum_{q=1}^{k}\left(\sum_{p=1}^{n} A_{ip}B_{pq}\right)C_{qj}, \quad \text{changing order of summation,} && (16)\\
&= \sum_{q=1}^{k} (AB)_{iq}C_{qj} && (17)\\
&= ((AB)C)_{ij} && (18)
\end{aligned}$$

and the theorem is proved because i and j were arbitrary.

Recall the compact form for a system of linear equations, $A\underline{x} = \underline{b}$ that we obtained at the end of the last lecture. It clearly suggest that we study the behaviour of products of the type matrix times vector, $A\underline{v}$ say.

4.1.2 Matrix times vector, or linear mappings

Convention. I'm using $\underline{v}$ for an arbitrary vector and $\underline{x}$ for an unknown vector. I see that this might cause confusion at first. But on the other hand it exposes the different faces a vector has: it can be a variable or unknown in an equation, or it may just be a particular but not further described "point" in $\mathbb{R}^n$, the set of all n-tuples of real numbers. Let us also agree that writing $A\underline{v}$ implies that the product is defined. Just imagine your favourite two natural numbers m and n and assume that A is an $m \times n$-matrix and $\underline{v}$ is an *n-dimensional* column vector (we will pin down the notion of dimension mathematically precise in a future lecture).

Since $A\underline{v}$ is again a vector, say m-dimensional, we can interpret A as a mapping $\mathbb{R}^n \longrightarrow \mathbb{R}^m$. The first two properties of such mappings are

(Lin1) $A(\underline{v} + \underline{u}) = A\underline{v} + A\underline{u}$, and

(Lin2) $A(\lambda\underline{v}) = \lambda(A\underline{v})$,

where we use the following definitions with $\lambda \in \mathbb{R}$

$$\underline{v} + \underline{u} = \begin{pmatrix} v_1 + u_1 \\ \vdots \\ v_n + u_n \end{pmatrix}, \qquad \text{and} \qquad \lambda\underline{v} = \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix}.$$

In words the two properties (Lin1) and (Lin2) say that **matrix mappings respect addition and scalar multiplication of vectors**.[1]

We call $A\underline{v}$ the *image of $\underline{v}$ under A*. Since every vector $\underline{v}$ can be written as

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + v_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

it suffices to know the images of $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ under A and then use (Lin1) and (Lin2) to obtain the image of $\underline{v}$ under A.

[1]This is one incarnation of a general mathematical species called a *linear mapping* from one *vector space* to another vector space; hence the names for the rules. Most of the time we will work with $\mathbb{R}^n$, n-dimensional real or Euclidean space. Certain sets of functions can also form vector spaces, for instance all 2π-periodic functions from $\mathbb{R} \longrightarrow \mathbb{R}$.

Lecture 4

Matrix and vector notation further continued

One important consequence of associativity (Theorem 1) is the following result which says that composition of mappings corresponds to matrix multiplication.

***Corollary* 2** *Let A and B be an $n \times k$-matrix and an $m \times n$-matrix respectively, and let $\underline{v}$ be a k-dimensional vector. Then the image of $\underline{v}$ under BA is the image of $A\underline{v}$ under B.*

To see this, view $\underline{v}$ as $k \times 1$-matrix and note that the statement is equivalent to $(BA)\underline{v} = B(A\underline{v})$ which follows from Theorem 1.

4.1.3 Identity, Zero, and zero-divisors

In analogy to the real number one, let us define an *identity matrix* I to be one that satisfies $AI = IA = A$ for all A for which both products are defined. Let us show that identity matrices exist. That every identity matrix must be square follows from (1) on page 12. And hence for AI and IA to be defined, A has to be a square matrix as well.

For each n there is an $n \times n$ (or n-square) identity matrix I_n. To see this define

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

or using the notation from the proof of Theorem 1,

$$(I_n)_{ij} = \begin{cases} 1 & \text{, if } i = j \\ 0 & \text{, otherwise} \end{cases} .$$

Now the multiplication formula gives

$$(AI_n)_{ij} = \sum_p A_{ip}(I_n)_{pj} = A_{ij} \qquad \text{and}$$

$$(I_nA)_{ij} = \sum_p (I_n)_{ip}A_{pj} = A_{ij}.$$

***Theorem* 3** *There is only one n-square identity matrix for each dimension n.*

Proof. Suppose I and I' are both n-square identity matrices. Then $I = II' = I'$, and so $I = I'$.

The $m \times n$ *zero matrix*, denoted $0_{m\times n}$ or 0_n when $n = m$, is defined to be the $m \times n$-matrix which has all entries equal to zero. It is easy to check (do it?!) that for every $n \times k$ matrix A and every $l \times m$-matrix B, $0_{m\times n}A = 0_{m\times k}$ and $B0_{m\times n} = 0_{l\times n}$.

***Theorem* 4** *For every $n \geq 2$, the set of $n \times n$-matrices has zero divisors, i.e. there are non-zero $n \times n$-matrices A and B with $AB = 0_n$.*

Proof. The case $n = 2$ is Exercise $2(a)$ on problem sheet 2, eg.

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 5 & 4 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

And if $n > 2$, the matrices which have their upper left 2×2 corner equal to the matrices above and zeros elsewhere will do the trick.

4.1.4 Inverses

By definition the matrix A has a *left inverse* if there is a matrix B with $BA = I_n$ for some n. Note that B and A have to be of type $n \times m$ and $m \times n$ respectively, if $BA = I_n$ holds. This, in turn, implies that AB exists as well.

Example of a left inverse which is not a right inverse for the same matrix.

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

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

Let us record the following consequence of the existence of zero divisors: not every matrix has a left inverse.

***Corollary* 5** *For every $n \geq 2$, there are n-square matrices without left inverse.*

Proof. Let $n \geq 2$. By the theorem, there are non-zero $n \times n$-matrices A and B with $AB = 0_n$. Assume that A has a left inverse, C say. Then

$$0_n = C0_n = C(AB) = (CA)B = I_nB = B$$

which is a contradiction, as $B \neq 0_n$. Hence A has no left inverse.
Remark. Note the use of Theorem 1 in the proof.

We say that A has an *inverse* B if $AB = BA = I_n$ for some n.

***Theorem* 6** *If a matrix A has an inverse B, then A is a square matrix and B is unique.*

Proof. It follows from (1) on page 12 that A has to be a square matrix. Now assume B and B' are inverses of A, then $B' = B'(AB) = (B'A)B = B$. This proves the theorem (again using Theorem 1).

A matrix A with an inverse, usually denoted A^{-1}, is called *invertible* or *non-singular*. Since an inverse is in particular a left inverse, there are *singular* matrices, that is matrices without inverse, by the corollary.

Problem. When does a square matrix have an inverse?

This will be addressed later in the course. Here is one reason. Suppose you can find the inverse A^{-1} of the matrix A. Then, because of $\underline{x} = (A^{-1}A)\underline{x} = A^{-1}\underline{b}$, you solved the system $A\underline{x} = \underline{b}$ of linear equations.

***Theorem* 7** *Let I be a matrix satisfying $AI = A$ for all matrices A such that AI is defined. Then $I = I_n$ for some n. In particular, if A is square then $IA = A$ holds too.*

Proof. First note that I has to be a square matrix, $n \times n$ say. Now define a matrix $E_{(kl)}$ with $1 \leq k, l \leq n$ by

$$(E_{(kl)})_{ij} = \begin{cases} 1, & \text{if } i = k \text{ and } j = l \\ 0, & \text{otherwise} \end{cases} . \tag{19}$$

We exploit that $E_{(kl)}I = E_{(kl)}$ for any choice of k and l. So choose i with $1 \leq i \leq n$ and check

$$1 = (E_{(ki)})_{ki} = (E_{(ki)}I)_{ki} = \sum_{q=1}^{n} (E_{(ki)})_{kq} I_{qi} = I_{ii}.$$

And if $i \neq j$, then

$$0 = (E_{(ki)})_{kj} = (E_{(ki)}I)_{kj} = \sum_{q=1}^{n} (E_{(ki)})_{kq} I_{qj} = I_{ij}.$$

It follows now that $I = I_n$ and the proof is complete.

4.1.5 Commutativity

***Theorem* 8** *Matrix multiplication is not commutative, not even for square matrices.*

Proof. Check

$$\begin{pmatrix} 5 & 4 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 5 \\ 0 & 0 \end{pmatrix}$$

and compare with the product in the proof of Theorem 4, q.e.d.

Remark. Non commutativity is a common feature of mappings, eg. if $f(x) = x^2 + 1$ and $g(x) = \frac{1}{x}$, then $(f \circ g)(x) = \frac{1}{x^2} + 1 \neq (g \circ f)(x) = \frac{1}{x^2+1}$.

4.1.6 Addition and distributivity

Let A and B be matrices of the same type, $n \times m$ say. Then we define $A+B$ as the matrix whose ij-entry is the sum of the ij-entries of A and B. In our compact notation this is $(A + B)_{ij} = A_{ij} + B_{ij}$. Let us also define λA for a real (complex) number λ, by $(\lambda A)_i j = \lambda a_{ij}$.

The following facts follow directly from (Lin1) and (Lin2) on page 14, by viewing the second factor as a collection of columns. Assume that all the products and sums below are defined.

1. $A(B + C) = AB + AC$
2. $(B + C)D = BD + CD$
3. $A(\lambda B) = \lambda(AB) = (\lambda A)B$.

The first two laws are called *left and right distributivity*, respectively.

Lecture 5

4.2 Matrices doing row operations

Recall the definition of the matrix $E_{(kl)}$ from page 17. Suppose $E_{(kl)}$ is an $m \times n$-matrix and let A be an $n \times p$-matrix. Then

$$(E_{(kl)}A)_{ij} = \sum_{q=1}^{n} (E_{(kl)})_{iq} A_{qj} = \begin{cases} A_{lj}, & \text{if } i = k \\ 0, & \text{otherwise} \end{cases} .$$

We can express this as follows.

The k-th row of $E_{(kl)}A$ is the l-th row of A and all other entries are zero.

Now we can use distributivity (cf. 4.1.5.) to see that $(I_n + \lambda E_{(kl)})A$ is the matrix obtained from A by adding λ times the l-th row to the k-th row. Notice that $E_{(kl)}$ has to be a square matrix for the sum and the product to be defined. But that is no problem.

Example. (row addition)

$$\begin{aligned}(I_3 + 2E_{(21)}) \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \\ &= \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 + 2a_1 & b_2 + 2b_1 & c_2 + 2c_1 \\ a_3 & b_3 & c_3 \end{pmatrix}\end{aligned}$$

Next consider $B = (I_n - E_{(ll)} - E_{(kk)} + E_{(kl)} + E_{(lk)})A$. I claim that this is the matrix obtained from A by interchanging its l-th and k-th row. Look at the l-th row of B: it is

$$l^{\text{th}} \text{ row of } A - l^{\text{th}} \text{ row of } A + k^{\text{th}} \text{ row of } A$$

and the k-th row of B is

$$k^{\text{th}} \text{ row of } A - k^{\text{th}} \text{ row of } A + l^{\text{th}} \text{ row of } A.$$

All other rows of B are equal to the same row of A. This verifies the claim.

Example. (INTERCHANGING ROWS)

$$(I_3 - E_{(11)} - E_{(22)} + E_{(21)} + E_{(12)}) \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = \begin{pmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

Finally consider $B = (I_n + (\lambda - 1)E_{(kk)})$. Then the k-th row of BA is

$$k^{\text{th}} \text{ row of } A + (\lambda - 1) \text{ times the } k^{\text{th}} \text{ row of } A,$$

which is just λ times the k-th row of A. All other rows are the same as for A.

Example. (SCALAR MULTIPLE OF ROW)

$$\begin{aligned} (I_3 + 4.5E_{33}) \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5.5 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \\ &= \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ 5.5a_3 & 5.5b_3 & 5.5c_3 \end{pmatrix} \end{aligned}$$

Thus we have expressed all row operations, as defined on page 5, in terms of matrix multiplication. That this is useful is the theme of the next section.

5 INVERTIBLE MATRICES

Suppose A is an n-square matrix. Let us investigate what can be said under the assumption that A has an inverse A^{-1}.

Consider the system of equations

$$A\underline{x} = \underline{b}, \tag{20}$$

where $\underline{b}$ is an arbitrary n-dimensional vector.

If (20) has a solution $\underline{v}$, i.e. $A\underline{v} = \underline{b}$, then

$$\underline{v} = I_n\underline{v} = A^{-1}A\underline{v} = A^{-1}\underline{b}$$

i.e., $\underline{v} = A^{-1}\underline{b}$. On the other hand $A^{-1}\underline{b}$ is a solution of (20), as $A(A^{-1}\underline{b}) = \underline{b}$. In particular, for every $\underline{b}$, the equation (20) has a unique solution.

Now assume that A is an n-square matrix such that $A\underline{x} = \underline{b}$ has a unique solution for every $\underline{b}$. Consider the Gaussian algorithm on A. It transforms A into a matrix of the form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & 0 & & a_{nn} \end{pmatrix}.$$

We must have $a_{nn} \neq 0$ because $\underline{b}$ was arbitrary. Hence we can add multiples of the last row to the others to eliminate all the a_{in} with $i < n$. Now we have a matrix of the form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & 0 \\ 0 & a_{22} & \cdots & a_{2n-1} & 0 \\ & & \ddots & \vdots & \vdots \\ & 0 & & a_{n-1n-1} & 0 \\ & & & 0 & a_{nn} \end{pmatrix}.$$

Again, it follows that $a_{n-1n-1} \neq 0$, and we can eliminate the a_{in-1} for $i < n-1$, by adding multiples of the $(n-1)$-st row to the others.

We will always find that the ii-entries are non-zero, by choosing the right $\underline{b}$, and continuing the process, we eventually transform A into a *diagonal matrix*; that is a matrix in which only the ii-entries for $1 \leq i \leq n$ are non-zero. And in our case we even know that the diagonal entries are all non-zero, because otherwise $A\underline{x} = \underline{0}$ would have more than one solution.

Since every row operation can be interpreted as multiplying A by the appropriate matrix on the left, we have shown that there is a matrix B, the product of all the row operations we used, so that

$$BA = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix},$$

where all the d_i are non-zero. And therefore

$$\begin{pmatrix} \frac{1}{d_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{d_n} \end{pmatrix} BA = I_n.$$

So A has a left inverse, call it C. Next we show that C is in fact an inverse of A. To do this we have to show that $AC = I_n$.

First observe that, if $AC\underline{v} = \underline{v}$ for every n-dimensional vector $\underline{v}$, then $(AC)X = X$ for every matrix X such that the product is defined (view X as a collection of columns). And then the obvious analogue of Theorem 7 (see Question 1 on sheet 4) tells us that $AC = I_n$. So we only have to show that $AC\underline{v} = \underline{v}$ holds for every $\underline{v}$.

To this end, let $\underline{v}$ be an arbitrary n-dimensional vector. By assumption $A\underline{x} = \underline{v}$ has a solution, $\underline{w}$ say. Hence

$$AC\underline{v} = ACA\underline{w} = AI_n\underline{w} = A\underline{w} = \underline{v},$$

and we are done. Let us summarise this.

***Theorem* 9** *An n-square matrix A has an inverse if and only if, for every n-dimensional vector $\underline{b}$ the equation $A\underline{x} = \underline{b}$ has precisely one solution.*

Now this might seem not very helpful because how does one check whether (20) has a unique solution for every $\underline{b}$. Well you don't have to do it because the arguments above show that one can transform A into the identity matrix using row operations if A is invertible.

This also gives us a **procedure to invert an invertible matrix** A:
Write A and I_n next to each other and do the Gaussian algorithm on A and simultaneously apply the same sequence of row operations to I_n. When you have transformed A into I_n, then I_n has been transformed into A^{-1}. Because in effect you have multiplied both A and I_n with a matrix, P say. But then P is nothing but the matrix into which I_n got transformed. Since $PA = I_n$, $P = A^{-1}$. If you cannot transform A into I_n, then it is not invertible.

Remark. It does not matter which sequence of row operations one chooses as long as they transform A into the identity matrix.

Examples.

1.

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

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

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

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

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

And it is alway worth checking for mistakes. But

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

2.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 0 \\ -1 & 0 & -14 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -6 \\ 0 & 2 & -11 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & -2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} -14 & 6 & -3 \\ 28 & -11 & 6 \\ 5 & -2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} -70 & 28 & -15 \\ 28 & -11 & 6 \\ 5 & -2 & 1 \end{pmatrix}$$

And again one should check that

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 0 \\ -1 & 0 & -14 \end{pmatrix} \begin{pmatrix} -14 & 6 & -3 \\ 28 & -11 & 6 \\ 5 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Lecture 6

6 Linear combinations and linear independence

6.1 Linear combinations and linear span

Let $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_m$ be n-dimensional vectors. By definition, a *linear combination* of the vectors v_i $(1 \leq i \leq m)$ is a sum of scalar multiples of the $\underline{v}_i$; that is an expression of the form

$$\alpha_1 \underline{v}_1 + \alpha_2 \underline{v}_2 + \cdots + \alpha_m \underline{v}_m, \text{ with } \alpha_i \in \mathbb{R}, (1 \leq i \leq m).$$

Examples.

1. $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ is a linear combination of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ because
$$\begin{pmatrix} 4 \\ 2 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

2. $\begin{pmatrix} 11 \\ 1 \\ -7 \end{pmatrix}$ is a linear combination of $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$, and $\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ because
$$\begin{pmatrix} 11 \\ 1 \\ -7 \end{pmatrix} = 11 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + (-7) \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}.$$

3. $\begin{pmatrix} 11 \\ 1 \\ -7 \end{pmatrix}$ is not a linear combination of $\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ because the first entry of
$$\alpha \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$
is zero for any choice of real numbers α and β.

The set of all linear combinations of the n-dimensional vectors $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_m$ is denoted $[\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_m]$ and called the *linear span of the* $\underline{v}_i$ $(1 \leq i \leq m)$.

Examples.

1. The linear span of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is the set of all 2-dimensional vectors. For $\begin{pmatrix} a \\ b \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

2. The linear span of $\underline{v}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\underline{v}_2 = \begin{pmatrix} -1 \\ -1.5 \end{pmatrix}$ does not contain $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$, because $\underline{v}_2 = -\frac{1}{2}\underline{v}_1$, and therefore every linear combination of $\underline{v}_1$ and $\underline{v}_2$ is in fact a linear combination of $\underline{v}_1$ alone. But this means that every vector in the linear span of $\underline{v}_1$ and $\underline{v}_2$ is a scalar multiple of $\underline{v}_1$, and $\begin{pmatrix} 2 \\ 2 \end{pmatrix} \neq \alpha\underline{v}_1$ for all $\alpha \in \mathbb{R}$.

These examples show that there are pairs of 2-dimensional vectors $\underline{v}_1$ and $\underline{v}_2$ so that $[\underline{v}_1, \underline{v}_2]$ is the set of all 2-dimensional vectors. But somehow $\underline{v}_1$ and $\underline{v}_2$ cannot be completely arbitrary. We shall make this more precise in the following section.

6.2 Linear independence

A sequence $\langle\, \underline{v}_1, \ldots, \underline{v}_m \,\rangle$ of n-dimensional vectors $\underline{v}_1, \ldots, \underline{v}_m$ is called *linearly independent* if there is a unique way to express the zero-vector $\underline{0}$ as a linear combination of the $\underline{v}_i$ $(1 \leq i \leq m)$.

Since $\underline{0} = 0\underline{v}_1 + 0\underline{v}_2 + \cdots + 0\underline{v}_m$, we have that $\langle\, \underline{v}_1, \ldots, \underline{v}_m \,\rangle$ is linearly independent if and only if $\alpha_1\underline{v}_1 + \alpha_2\underline{v}_2 + \cdots + \alpha_m\underline{v}_m = \underline{0}$ implies $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$.

Examples.

1. $\langle \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -8 \\ -12 \end{pmatrix} \rangle$ is not linearly independent because

$$4\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} -8 \\ -12 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

2. $\langle \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix} \rangle$ is linearly independent. To see this suppose

$$\underline{0} = \alpha\begin{pmatrix} 2 \\ -2 \end{pmatrix} + \beta\begin{pmatrix} 4 \\ 5 \end{pmatrix}.$$

But this is the same as

$$\begin{pmatrix} 2 & 4 \\ -2 & 5 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus we try to find the solutions of this system of equations. Using matrix notation we want to solve

$$\left(\begin{array}{cc|c} 2 & 4 & 0 \\ -2 & 5 & 0 \end{array} \right).$$

Adding the first row to the second gives

$$\left(\begin{array}{cc|c} 2 & 4 & 0 \\ 0 & 9 & 0 \end{array} \right).$$

And now the second row implies that $\beta = 0$ and then it follows from the first row that also $\alpha = 0$. Hence the zero-vector has a unique expression as a linear combination of $\begin{pmatrix} 2 \\ -2 \end{pmatrix}$ and $\begin{pmatrix} 4 \\ 5 \end{pmatrix}$, i.e. $\langle \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \end{pmatrix} \rangle$ is linearly independent.

6.2.1 Properties of linearly independent sequences

Throughout this subsection we assume that $\langle \underline{v}_1, \ldots, \underline{v}_m \rangle$ is a linearly independent sequence of n-dimensional vectors.

Claim 1. For every **non-zero** scalar α and every i with $1 \leq i \leq m$, the sequence $\langle \underline{v}_1, \ldots, \underline{v}_{i-1}, \alpha\underline{v}_i, \underline{v}_{i+1}, \ldots, \underline{v}_m \rangle$ is linearly independent.

To see this suppose

$$\underline{0} = \alpha_1\underline{v}_1 + \cdots + \alpha_{i-1}\underline{v}_{i-1} + \alpha_i(\alpha\underline{v}_i) + \alpha_{i+1}\underline{v}_{i+1} + \cdots + \alpha_{m\underline{v}_m}.$$

Since $\alpha_i(\alpha\underline{v}_i) = (\alpha_1\alpha)\underline{v}_i$ and $\langle \underline{v}_1, \ldots, \underline{v}_m \rangle$ is a linearly independent, we get that $\alpha_1 = \cdots = \alpha_{i-1} = \alpha_i\alpha = \alpha_{i+1} = \cdots = \alpha_m$. But α was not zero, so α_i must be zero, and we have established the claim.

Lecture 7

Properties of linearly independent sequences continued

We still assume that $\langle\, \underline{v}_1, \ldots, \underline{v}_m \,\rangle$ is a linearly independent sequence of n-dimensional vectors.

Claim 2. For every pair of distinct indices i and j with $1 \leq i, j \leq m$, the sequence obtained by replacing $\underline{v}_i$ by $\underline{v}_i \underline{v}_j$, i.e. $\langle\, \underline{v}_1, \ldots, \underline{v}_{i-1}, \underline{v}_i + \underline{v}_j, \underline{v}_{i+1}, \ldots, \underline{v}_m \,\rangle$, is also linearly independent. This follows from

$$\begin{aligned} \underline{0} &= \alpha_1 \underline{v}_1 + \cdots + \alpha_{i-1} \underline{v}_{i-1} + \alpha_i (\underline{v}_i + \underline{v}_j)) + \alpha_{i+1} \underline{v}_{i+1} + \cdots + \alpha_m \underline{v}_m \\ &= \alpha_1 \underline{v}_1 + \cdots + \alpha_{j-1} \underline{v}_{j-1} + (\alpha_i + \alpha_j) \underline{v}_j) + \alpha_{j+1} \underline{v}_{j+1} + \cdots + \alpha_m \underline{v}_m \end{aligned}$$

which implies $\alpha_1 = \cdots = \alpha_{j-1} = \alpha_j + \alpha_i = \alpha_{j+1} = \cdots = \alpha_m = 0$. So $\alpha_i = 0$ and $\alpha_j + \alpha_i = 0$, and hence $\alpha_j = 0$, which is what we needed to show.

Recall that we can view a real $m \times n$-matrix A as a mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. And one could ask which vectors in $\mathbb{R}^m$ are in the image of A, i.e., for which $\underline{v} \in \mathbb{R}^m$ does there exist a vector $\underline{w} \in \mathbb{R}^n$ with $A\underline{w} = \underline{v}$?

***Theorem* 10** *Let A be a real $m \times n$-matrix. Then a vector $\underline{v} \in \mathbb{R}^m$ is in the image of A if and only if $\underline{v}$ is a linear combination of the columns of A.*

Proof. This follows from

$$\begin{aligned} \underline{v} = A\underline{w} &= A \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = A \left[\begin{pmatrix} w_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ w_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ w_n \end{pmatrix} \right] \\ &= A \begin{pmatrix} w_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + A \begin{pmatrix} 0 \\ w_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + A \begin{pmatrix} 0 \\ \vdots \\ 0 \\ w_n \end{pmatrix} \\ &= w_1 A_{*1} + w_2 A_{*2} + \cdots + w_n A_{*n} \end{aligned}$$

where A_{*j} denotes the j-th column of A.

Another important result is the following.

***Theorem* 11** *A square matrix is invertible if and only if the sequence of its columns is linearly independent.*

6.2.2 Dimension of a vector space

A *real vector space* is a generalisation of the vector space $\mathbb{R}^n$ which we have been working with. It is a mathematical structure satisfying certain rules, the vector space axioms, which capture the essence of the things we know about $\mathbb{R}^n$.

Whatever the precise definition of a vector space V is, there are notions of linear combinations and linear independence. And one defines the *dimension* of V to be the maximal length of a linearly independent sequence of vectors in V or infinity if there are such sequences of arbitrary length.[2]

It turns out that $\mathbb{R}^n$ has precisely dimension n, as one would hope.

Remark. The set of all polynomials with real coefficients forms a real vector space of infinite dimension.

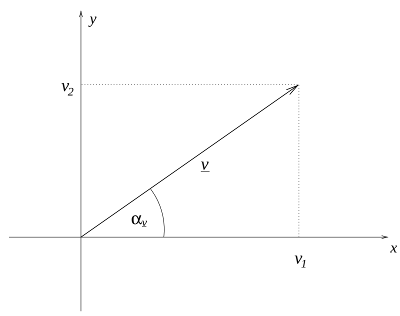
7 Some geometry and determinants

Very useful features of $\mathbb{R}^n$, alias Euclidean n-space, are distances and angles. One way of thinking of a vector in $\mathbb{R}^n$ is to view it as a 'direction together with a length'. This is why we often depict a vector as an arrow instead of just marking it as a point.

The direction of a vector is certainly relative to some preferred direction which in the following figures will be the x-axis.

Remark. Observe the link to polar coordinates here.

Fix an arbitrary vector $\underline{v}$ in $\mathbb{R}^2$ and imagine it as shown in the following figure.

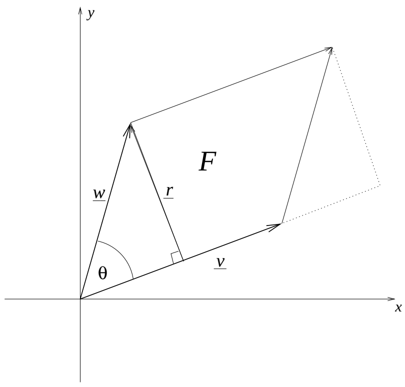


[2]This is not in the syllabus, so if you want to know more about this consult a proper Linear Algebra book, eg. R.B.J.T. Allenby *Linear Algebra*. There are over a hundred others in the library.

Let $|\underline{v}|$ denote the length of $\underline{v}$. From the figure we see, using Pythagoras' Theorem and the basic sin and cos rules, that

$$|\underline{v}| = \sqrt{v_1^2 + v_2^2}, \qquad \frac{v_1}{|\underline{v}|} = \cos\alpha_{\underline{v}}, \qquad \frac{v_2}{|\underline{v}|} = \sin\alpha_{\underline{v}}.$$

Now let us try to determine the area F of the parallelogram with sides $\underline{v}$ and $\underline{w}$ as in the following figure.



We find that

$$F = |\underline{v}||\underline{r}| = |\underline{v}||\underline{w}||\sin\theta|.$$

The absolute value of $\sin\theta$ is needed, as areas cannot be negative. Now $\theta = \alpha_{\underline{w}} - \alpha_{\underline{v}}$, so

$$\sin\theta = \sin\alpha_{\underline{w}}\cos -\alpha_{\underline{v}} + \cos\alpha_{\underline{w}}\sin -\alpha_{\underline{v}} = \sin\alpha_{\underline{w}}\cos\alpha_{\underline{v}} - \cos\alpha_{\underline{w}}\sin\alpha_{\underline{v}},$$

by the symmetries of sin and cos and the addition theorems (see problem sheet 2). And hence

$$\begin{aligned} F &= |\underline{v}||\underline{w}|\left(\sin\alpha_{\underline{w}}\cos\alpha_{\underline{v}} - \cos\alpha_{\underline{w}}\sin\alpha_{\underline{v}}\right) \\ &= |\underline{v}||\underline{w}|\left(\tfrac{w_2}{|\underline{w}|}\tfrac{v_1}{|\underline{v}|} - \tfrac{w_1}{|\underline{w}|}\tfrac{v_2}{|\underline{v}|}\right) \\ &= v_1w_2 - v_2w_1. \end{aligned}$$

This observation inspires the following definition.

The *determinant* of the 2×2-matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, which we denote as $\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right|$, is defined by

$$\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = ad - bc.$$

Example. $\left|\begin{pmatrix} 2 & 1 \\ 2 & 3 \end{pmatrix}\right| = 6 - 2 = 4$

We can summarise what we know about invertible 2×2-matrices.

***Theorem* 12** *For a 2×2-matrix A the following are equivalent.*

1. *A is invertible*
2. *$A\underline{x} = \underline{v}$ has a unique solution for every two dimensional vector $\underline{v}$.*
3. *The sequence consisting of the tow columns of A is linearly independent.*
4. *The parallelogram whose sides are the columns of A has non-zero area.*
5. *$|A| \neq 0$*

Proof. The implications 1. $\Longrightarrow$ 2. $\Longrightarrow$ 3. $\Longrightarrow$ 4. $\Longrightarrow$ 5. are Theorem 9, specification of $\underline{v}$ to $\underline{0}$, the fact that linearly independent vectors are not lying on the same line, and the discussion above, respectively. To see 6. $\Longrightarrow$ 1., check that

$$\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = I_2.$$

The determinant of the 3×3-matrix $\begin{pmatrix} a & b & c \\ d & d & f \\ g & h & k \end{pmatrix}$ is defined by

$$\left|\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix}\right| = a\left|\begin{pmatrix} e & f \\ h & k \end{pmatrix}\right| - b\left|\begin{pmatrix} d & f \\ g & k \end{pmatrix}\right| + c\left|\begin{pmatrix} d & e \\ g & h \end{pmatrix}\right|$$

Example.

$$\left|\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & -1 \\ 3 & 0 & 2 \end{pmatrix}\right| = 1\left|\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}\right| - 2\left|\begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix}\right| + 0\left|\begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}\right| = 2 - 14 = -12$$

Lecture 8

Determinants continued

For $n > 3$, the determinant of an n-square matrix A is defined as follows. Let $A^{(ij)}$ denote the $(n-1)$-square matrix obtained from A by deleting the i-th row and the j-th column. Now $|A|$ is defined by

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

Example.

$$\begin{aligned}\left|\begin{pmatrix} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{pmatrix}\right| &= 11\left|\begin{pmatrix} 22 & 23 & 24 \\ 32 & 33 & 34 \\ 42 & 43 & 44 \end{pmatrix}\right| - 12\left|\begin{pmatrix} 21 & 23 & 24 \\ 31 & 33 & 34 \\ 41 & 43 & 44 \end{pmatrix}\right| \\ &+ 13\left|\begin{pmatrix} 21 & 22 & 24 \\ 31 & 32 & 34 \\ 41 & 42 & 44 \end{pmatrix}\right| - 14\left|\begin{pmatrix} 21 & 22 & 23 \\ 31 & 32 & 33 \\ 41 & 42 & 43 \end{pmatrix}\right|\end{aligned}$$

The general theory of determinants is not very easy, so we just collect the most important properties in the following three theorems. Proofs of these result can be found in the literature.

***Theorem* 13** *Let A be an n-square matrix.*

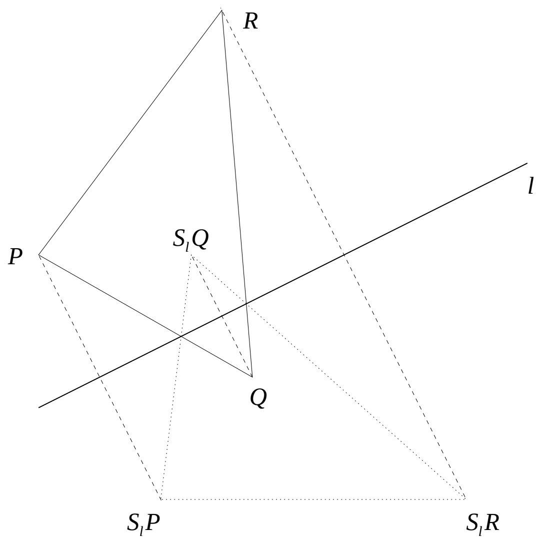
1. *When B is the matrix obtained from A by multiplying the i-th row with the real (or complex) number λ, then $|B| = \lambda|A|$.*
2. *When C is the matrix obtained from A by replacing the i-th row by the sum of the i-th and the j-th row with $i \neq j$, then $|B| = |A|$.*
3. *When D is obtained from A by interchanging the i-th and the j-th row where $i \neq j$, then $|D| = -|A|$.*
4. *When A is upper (or lower) triangular, then $|A|$ is equal to the product of the diagonal entries, i.e. $|A| = \prod_{i=1}^{n} A_{ii}$.*

***Theorem* 14** *Let A and B be n-square matrices, $n \geq 1$. Then $|AB| = |A||B|$; that is, the determinant respects multiplication when looked at as a map from n-square matrices into the real (or complex) numbers.*

***Theorem* 15** *A square matrix is invertible if and only if its determinant is non-zero.*

8 Eigenvectors and eigenvalues

How would you describe the reflection S_l in the line l depicted below?



The first point to note is that we need to **choose the coordinates**. This involves

1. choosing an origin O, and
2. choosing two other points E_1 and E_2 such that O, E_1, and E_2 do not lie on a line.

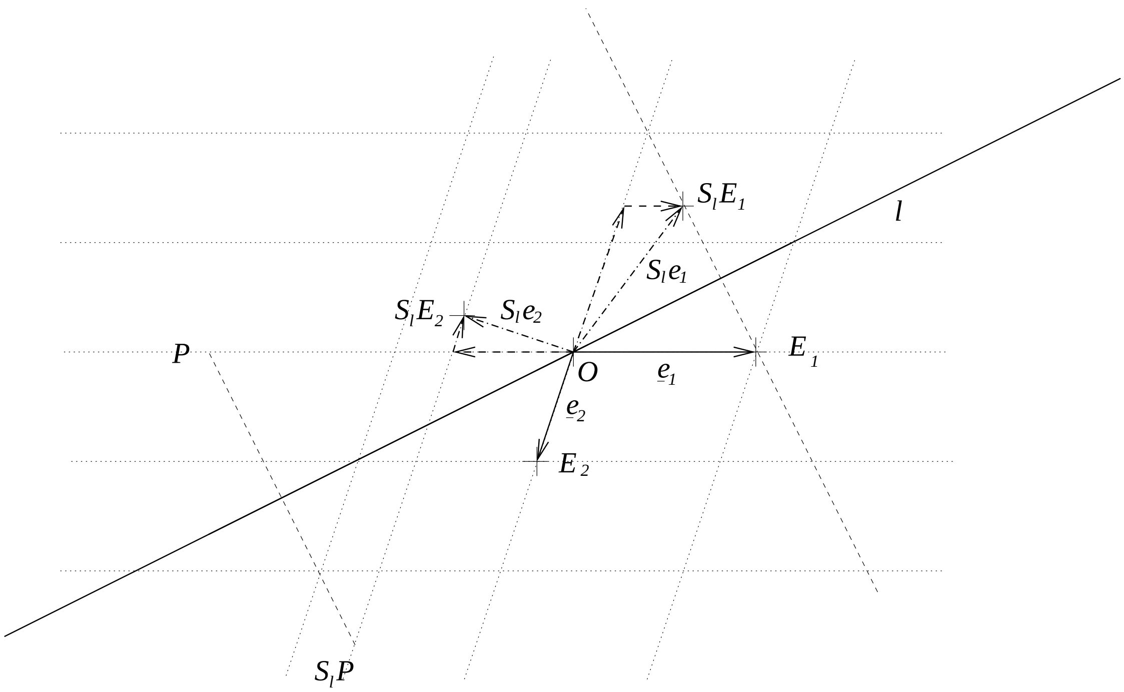
Let $\underline{e}_i$ be the vector from O to E_i, $i = 1, 2$. Then every point in the plane can be expressed as a linear combination of $\underline{e}_1$ and $\underline{e}_2$. Now there is some hope that we can describe S_l. Hopefully we can do this with a matrix.

Since every matrix mapping leaves the origin fixed, and l is fixed by S_l, we need to choose the origin on the reflecting line l, if we want to have any chance of finding a matrix describing S_l.

Now we look at the images $S_l\underline{e}_1$ and $S_l\underline{e}_2$ of $\underline{e}_1$ respectively $\underline{e}_2$, and express them in terms of our coordinates, say

$$S_l\underline{e}_1 = a_{11}\underline{e}_1 + a_{21}\underline{e}_2, \qquad \text{and} \qquad S_l\underline{e}_2 = a_{12}\underline{e}_1 + a_{22}\underline{e}_2.$$

It is no coincidence that this looks familiar. Most often, our favourite choice for E_1 and E_2 is such that $\underline{e}_1$ and $\underline{e}_2$ are orthogonal to each other and of length one. However, one could choose any coordinate system. But if you think of $\underline{e}_1$ and $\underline{e}_2$ as $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$, then $\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$ describes the reflection S_l with respect to the coordinates given by $\underline{e}_1$ and $\underline{e}_2$. So assume we choose E_1 and E_2 as in the following picture.



Then we find that $S_l\underline{e}_1 = \frac{1}{3}\underline{e}_1 - \frac{4}{3}\underline{e}_2$ and $S_l\underline{e}_2 = -\frac{2}{3}\underline{e}_1 - \frac{1}{3}\underline{e}_2$. So with respect to this choice of coordinates, S_l is given by $\begin{pmatrix}1/3 & -2/3\\ -4/3 & -1/3\end{pmatrix}$.

Having this freedom in choosing coordinates, we should take $\underline{e}_1$ and $\underline{e}_2$, or equivalently E_1 and E_2, so that their images are easy to express as linear combinations. Since l is fixed, it seems clever to take E_1 on the line l, for then $S_l\underline{e}_1 = \underline{e}_1$. If next we choose E_2 on a line through O and orthogonal to l, then $S_l\underline{e}_2 = -\underline{e}_2$. So with this choice of coordinates, S_l is given by $\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$.

This last description of S_l has a very simple form; it is just a diagonal matrix. In this case the images of the coordinate vectors are just multiples of these. This motivates the following definition.

A vector $\underline{v}$ is called an *eigenvector* for the matrix A if $\underline{v} \neq \underline{0}$ and $A\underline{v} = \alpha\underline{v}$ for some real (or complex) number α. In case $\underline{v}$ is an eigenvector, α is its corresponding eigenvalue. An arbitrary scalar α is an *eigenvalue* of A if A has an eigenvector with eigenvalue α.

Example. Take the matrix $S = \begin{pmatrix} 1/3 & -2/3 \\ -4/3 & -1/3 \end{pmatrix}$ from above. Now it should be true that a point on the line l is fixed because we believe that S describes the reflection in l. From the picture above we see that $\underline{v} = \underline{e}_1 - \underline{e}_2$ is a point on the line l. So what is the image of $\underline{v}$ under S? The answer is

$$S\underline{v} = S\underline{e}_1 - S\underline{e}_2 = \frac{1}{3}\underline{e}_1 - \frac{4}{3}\underline{e}_2 - (-\frac{2}{3}\underline{e}_1 - \frac{1}{3}\underline{e}_2) = \underline{e}_1 - \underline{e}_2 = \underline{v},$$

or directly in matrix notation $\begin{pmatrix} 1/3 & -2/3 \\ -4/3 & -1/3 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Thus $\underline{v} = \underline{e}_1 - \underline{e}_2$ is an eigenvector of S with corresponding eigenvalue 1.

Suppose that α is an eigenvalue of the matrix A. Let $\underline{v}$ be an eigenvector of A with eigenvalue α. Then $A\underline{v} = \alpha\underline{v}$ implies $A\underline{v} - \alpha\underline{v} = (A - \alpha I)\underline{v} = \underline{0}$. Since $\underline{v} \neq \underline{0}$ this means that $A - \alpha I$ is not invertible, or equivalently, that $|A - \alpha I| = 0$.

Let A be a square matrix, and let x be a variable. The *characteristic polynomial* of A is defined as $|A - xI|$ and it will be denoted by $\chi_A(x)$. The previous paragraph says that an eigenvalue of the matrix A must be a root (or zero) of $\chi_A(x)$.

Example. Again take $S = \begin{pmatrix} 1/3 & -2/3 \\ -4/3 & -1/3 \end{pmatrix}$. Then

$$\begin{aligned} \chi_S(x) &= \left|\begin{pmatrix} 1/3 - x & -2/3 \\ -4/3 & -1/3 - x \end{pmatrix}\right| \\ &= (1/3 - x)(-1/3 - x) - 8/3 \\ &= x^2 - 1 \\ &= (x-1)(x+1). \end{aligned}$$

So the characteristic polynomial of S is $x^2 - 1$ and the possible eigenvalues are 1 and -1.

***Theorem* 16** *The eigenvalues of the square matrix A are the zeros of the characteristic polynomial of A.*

Proof. We saw above that an eigenvalue must be a root of the characteristic polynomial. For the converse suppose that α is a root of the characteristic polynomial, then $|A - \alpha I| = 0$, and hence $A - \alpha I$ is not invertible, whence $(A - \alpha I)\underline{x} = \underline{0}$ has more than one solution. And at least one of these solutions, $\underline{w}$ say, is different from $\underline{0}$. It is easy to see that $\underline{w}$ is an eigenvector of A with eigenvalue α. The theorem is proved.

Lecture 9
Eigenvectors and eigenvalues revisited

8.1 A typical problem [3]

In order to get more familiar with the important notions introduced in Lecture 8, let us work out an example solution of the following

Problem. Let K be a real number and let $A = \begin{pmatrix} K & -K & 0 \\ -K & 2K & -K \\ 0 & -K & K \end{pmatrix}$. Find the eigenvalues and eigenvectors of A.
Solution. First we compute the characteristic polynomial of A:

$$\begin{aligned} \chi_A(x) &= |A - xI_3| = \left| \begin{pmatrix} K-x & -K & 0 \\ -K & 2K-x & -K \\ 0 & -K & K-x \end{pmatrix} \right| \\ &= (K-x)((2K-x)(K-x) - K^2) + K(-K(K-x)) \\ &= (K-x)(x^2 - 3Kx) \\ &= x(K-x)(x-3K). \end{aligned}$$

Now the eigenvalues are the roots of $\chi_A(x)$, i.e., 0, K, and $3K$. In order to find the eigenvector with eigenvalue 0 we have to find a nontrivial solution of $A\underline{x} = \underline{0}$. It follows easily that $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is an eigenvector with eigenvalue 0.

A nontrivial solution of $\underline{0} = (A - KI_3)\underline{x} = \begin{pmatrix} 0 & -K & 0 \\ -K & K & -K \\ 0 & -K & 0 \end{pmatrix} \underline{x}$ is an eigenvector with eigenvalue K, eg $\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$.

Finally, $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ is a solution of $\begin{pmatrix} -2K & -K & 0 \\ -K & -K & -K \\ 0 & -K & -2K \end{pmatrix} \underline{x} = \underline{0}$, and hence an eigenvector with eigenvalue $3K$.

[3] If this is the last time you encounter this type of question in your physics degree, this isn't worth the paper it's written on.

8.2 A theorem about characteristic polynomials

***Theorem* 17** *Let A be an n-square matrix and let T be an invertible n-square matrix. Then A and TAT^{-1} have the same characteristic polynomial, and hence also the same eigenvalues.*

Proof. This follows from

$$|TAT^{-1} - xI_n| = |TAT^{-1} - xTI_nT^{-1}| = |T(A - xI_n)T^{-1}| = |A - xI_n|,$$

where the last equality follows from Theorem 14.
Note that we haven't proved Theorem 14. However, Theorem 17 is easy to understand conceptually by thinking about the effect of passing from A to TAT^{-1}. And that is precisely what we do next.

8.3 Choice and change of coordinates

In our investigation of the reflection in Lecture 8 we saw that we had to choose coordinates in order to describe the reflection by a matrix.

In mathematical terms, a choice of coordinates for a vector space V is a basis of V. By definition, a *basis* of V is a maximal sequence of linearly independent vectors $\mathcal{B} =\subset V$. Recall that the number of elements in $\mathcal{B}$ is, by definition, the *dimension* of V.

Fact 1. When $\mathcal{B}$ is a basis of V, then every vector $\underline{v} \in V$ has a unique expression as a linear combination of the basis vectors $\underline{b}_i \in \mathcal{B}$, i.e, there are unique $v_i \in \mathbb{R}$ (or complex v_i) with $\underline{v} = \sum_{i=1}^{n} v_i \underline{b}_i$, where n is a positive integer less than the dimension of V.[4]

Fact 2. Assume $\underline{b}_1, \ldots, \underline{b}_n$ and $\underline{c}_1, \ldots, \underline{c}_n$ are both bases of the vector space V, then there is a linear mapping [5] $\phi : V \longrightarrow V$ with $\phi(\underline{b}_i) = \underline{c}_i$.

[4] In an infinite dimensional vector space every vector is a linear combination of finitely many basis vectors.

[5] See footnote on page 13.

Lecture 10

Coordinates continued

Fact 3. Every linear mapping $\phi : V \longrightarrow V$ can be described by a matrix once a basis of V has been chosen.

This goes as follows. Let $\phi : V \longrightarrow V$ be a linear mapping and let $\mathcal{B} = \{\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n\}$ be a basis for V. Let a_{ij} for $1 \leq i, j \leq n$ be real (or complex) numbers satisfying

$$\phi(\underline{b}_j) = a_{1j}\underline{b}_1 + a_{2j}\underline{b}_2 + \cdots + a_{nj}\underline{b}_n = \sum_{i=1}^{n} a_{ij}\underline{b}_i,$$

that is a_{ij} is the coefficient of $\underline{b}_i$ in the (unique) expression of $\phi(\underline{b}_j)$ as linear combination of the $\underline{b}_k$. Now let A be the n-square matrix with $(A)_{ij} = a_{ij}$. Then A describes ϕ with respect to the basis $\mathcal{B}$.

This becomes clearer when one views $\underline{b}_j$ as $\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, where the one is in the j-th row. For now $A\underline{b}_j$ is the j-th column of the matrix A which is nothing but $\phi(\underline{b}_j)$ expressed as linear combination of the $\underline{b}_k$.

Fact 4. When A is an invertible n-square matrix and $\mathcal{B} = \{\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_n\}$ is a basis of V, then $A\mathcal{B} = \{A\underline{b}_1, A\underline{b}_2, \ldots, A\underline{b}_n\}$ is also a basis of V.

Now suppose we have a matrix A describing a linear mapping $\phi : V \longrightarrow V$ with respect to the basis $\mathcal{B} = \{\underline{b}_1, \ldots, \underline{b}_n\}$ of V. Let $\mathcal{C} = \{\underline{c}_1, \ldots, \underline{c}_n\}$ be another basis. This poses

Problem. Which matrix describes ϕ with respect to the basis $\underline{c}_1, \ldots, \underline{c}_n$ of V?

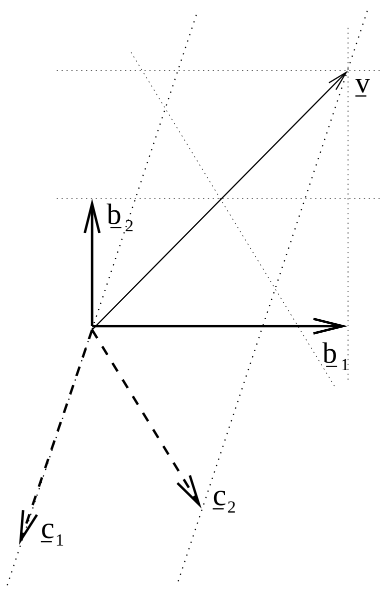
Let C be the matrix which describes the linear mapping $\underline{b}_i \mapsto \underline{c}_i$, $1 \leq i \leq n$ with respect to the basis $\mathcal{B}$. Then $C^{-1}AC$ is exactly the matrix that describes ϕ with respect to $\underline{c}_1, \ldots, \underline{c}_n$. Summing up we have.

***Theorem* 18** *Let $\mathcal{B} = \{\underline{b}_1, \ldots, \underline{b}_n\}$ and $\mathcal{C} = \{\underline{c}_1, \ldots, \underline{c}_n\}$ both be bases of the vector space V. Let C be the matrix which describes the linear mapping $\underline{b}_i \mapsto \underline{c}_i$, $1 \leq i \leq n$ with respect to the basis $\mathcal{B}$, and let A be any n-square matrix. Then A and $C^{-1}AC$ describe the same linear mapping but with respect to the bases $\mathcal{B}$ and $\mathcal{C}$ respectively.*

We can now understand Theorem 17 as follows. A linear mapping ϕ has eigenvalues and these are independent of the basis we choose to describe the linear mapping by a matrix; the eigenvalues of any matrix describing ϕ are the eigenvalues of the linear mapping ϕ. Since TAT^{-1} and A describe the same linear mapping with respect to different bases (read $T = C^{-1}$), it must be true that they have the same eigenvalues.

The point is that $\begin{pmatrix} 2 \\ -3 \end{pmatrix}$ has no meaning without choice of coordinates, but once a basis $\{\underline{b}_1, \underline{b}_2\}$ is chosen, we interpret $\begin{pmatrix} 2 \\ -3 \end{pmatrix}$ as $2\underline{b}_1 + 3\underline{b}_2$.

Example. Consider the figure below



Here $\mathcal{B} = \{\underline{b}_1, \underline{b}_2\}$ and $\mathcal{C} = \{\underline{c}_1, \underline{c}_2\}$ are both basis of $\mathbb{R}^2$. The vector $\underline{v}$ is given by $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ with respect to the basis $\mathcal{B}$, i.e. $\underline{v} = \underline{b}_1 + 2\underline{b}_2$. But with respect to the basis $\mathcal{C}$, the same vector $\underline{v}$ is given by $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$, because $\underline{v} = -2\underline{c}_1 + \underline{c}_2$. So with respect to $\mathcal{C}$, $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ describes the vector $\underline{c}_1 + 2\underline{c}_2$ which is quite different from $\underline{v}$.

With a ruler you can easily verify that

$$\underline{c}_1 = -\frac{2}{7}\underline{b}_1 - \frac{12}{7}\underline{b}_2 \qquad \text{and} \qquad \underline{c}_2 = \frac{3}{7}\underline{b}_1 - \frac{10}{7}\underline{b}_2.$$

In other words $\underline{c}_1 = \begin{pmatrix} -2/7 \\ -12/7 \end{pmatrix}$ with respect to $\mathcal{B}$ and $\underline{c}_2 = \begin{pmatrix} 3/7 \\ -10/7 \end{pmatrix}$ with respect to $\mathcal{B}$.

Thus, with respect to the basis $\mathcal{B}$, the linear mapping which sends $\underline{b}_1$ to $\underline{c}_1$ and $\underline{b}_2$ to $\underline{c}_2$ is described by the matrix $\begin{pmatrix} -2/7 & 3/7 \\ -12/7 & -10/7 \end{pmatrix}$; that is the matrix whose columns represent $\underline{c}_1$ and $\underline{c}_2$ with respect to $\mathcal{B}$.

Up to now we have mostly interpreted $A\underline{v}$, where A is an n-square matrix and $\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, as the image of $\underline{v}$ under the matrix mapping defined by the matrix A.

When A is invertible there is another way of interpreting $A\underline{v}$, namely as the vector represented by $\underline{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ with respect to the basis consisting of the columns of A which, in turn, are expressed with respect to some fixed basis, $\mathcal{B}$ say. This is because

$$A \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1\underline{A}_1 + v_2\underline{A}_2 + \cdots + v_n\underline{A}_n,$$

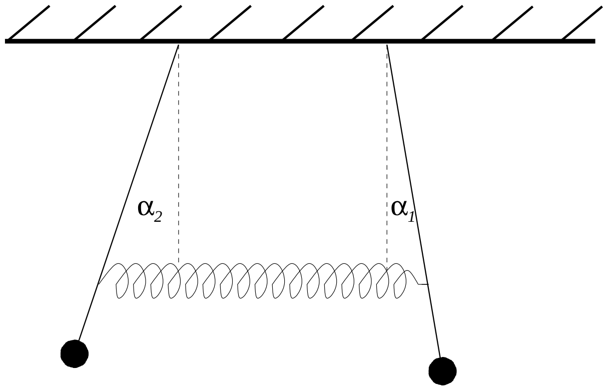
where $\underline{A}_i$ denotes the i-th column of A.

Now let C and A be n-square matrices with C invertible. Clearly $C^{-1}AC\underline{v} = \underline{w}$ is equivalent to $AC\underline{v} = C\underline{w}$, which we write as $A(C\underline{v}) = (C\underline{w})$. Clearly $C\underline{w}$ is the image of $C\underline{v}$ under A, and by the previous paragraph $C\underline{w}$ and $C\underline{v}$ are the vectors represented by $\underline{w}$ and $\underline{v}$ with respect to the basis $\mathcal{C}$ consisting of the columns of C, respectively. This means that $C^{-1}AC$ is the matrix which describes the same mapping as A but with respect to the basis $\mathcal{C}$.

Lecture 11

9 Coupled pendulums: an application

Consider two pendulums of mass m and length l coupled via a spring with spring constant D. We assume that they both move in the same plane. The situation is thus as in the following figure.



We would like to describe the behaviour of this system in time; that is we want functions $\alpha_1(t)$ and $\alpha_2(t)$ telling us how the pendulums move. Below we write $\dot{\alpha}$ for $\frac{d\alpha}{dt}$, the derivative of α with respect to time.

We proceed as follows:

(1) We find the Lagrange function $L(\alpha_i, \dot{\alpha}_i, t)$ for this situation.

(2) From L we obtain two differential equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\alpha}_i} - \frac{\partial L}{\partial \alpha_i} = 0, \quad i = 1, 2. \tag{21}$$

(3) We solve the differential equations.

In general, the Lagrange function has the form

$$L(\underline{x}_i, \dot{\underline{x}}_i, t) = T(\underline{x}_i, \dot{\underline{x}}_i, t) - U(\underline{x}_i, t),$$

where the first term T is the kinetic energy and U the potential energy of the system, and $\underline{x}_i$ is the coordinate vector of the i-th body (or particle) under consideration.

For small oscillations of the pendulums, that is small α_i, $\sin \alpha_i$ is roughly α_i, and we use this approximation in order to get a <u>linear</u> problem.

It turns out that for the coupled pendulums

$$L(\alpha_i, \dot{\alpha}_i, t) = \frac{1}{2}ml^2(\dot{\alpha}_1^2 + \dot{\alpha}_2^2) - \frac{1}{2}mgl(\alpha_1^2 + \alpha_2^2) - \frac{1}{2}Dl^2(\alpha_1 - \alpha_2)^2,$$

where T is the first summand and the other two make up U. Using (21) we obtain the following two differential equations.

$$\frac{d}{dt}ml^2\dot{\alpha}_1 + mgl\alpha_1 + Dl^2(\alpha_1 - \alpha_2) = 0, \text{ and}$$

$$\frac{d}{dt}ml^2\dot{\alpha}_2 + mgl\alpha_2 Dl^2(\alpha_2 - \alpha_1) = 0.$$

Carrying out the final derivation and dividing through by ml^2 we get

$$\ddot{\alpha}_1 + g/l\alpha_1 + D/m(\alpha_1 - \alpha_2) = 0, \text{ and } \ddot{\alpha}_2 + g/l\alpha_2 + D/m(\alpha_2 - \alpha_1) = 0.$$

In matrix notation this becomes

$$\underline{0} = \frac{d}{dt}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} + \begin{pmatrix} g/l + D/m & -D/m \\ -D/m & g/l + D/m \end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}.$$

Putting $\begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}v_1\\ v_2\end{pmatrix} e^{i\omega t}$, where $i^2 = -1$, we obtain

$$\underline{0} = \begin{pmatrix} g/l + D/m - \omega^2 & -D/m \\ -D/m & g/l + D/m - \omega^2 \end{pmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix}, \tag{22}$$

where we have already divided through by $e^{i\omega t}$ which is never zero. This equation has a non-trivial solution if and only if the determinant of the matrix is zero. We have

$$\begin{aligned}
&\left|\begin{pmatrix} g/l + D/m - \omega^2 & -D/m \\ -D/m & g/l + D/m - \omega^2 \end{pmatrix}\right| \\
&= (g/l + D/m - \omega^2)^2 - D^2/m^2 \\
&= \omega^4 - 2\omega^2(g/l + D/m) + (g/l + D/m)^2 - D^2/m^2,
\end{aligned}$$

and the only values for ω^2 which make this zero are

$$\omega^2 = (g/l + D/m) \pm \sqrt{D^2/m^2} = \begin{cases} g/l = \omega_1^2 \\ g/l + 2D/m = \omega_2^2 \end{cases}$$

For these two values of ω^2 we can now solve the equation (22):

$$\omega_1^2 : \qquad \begin{pmatrix} D/m & -D/m \\ -D/m & D/m \end{pmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix} = 0 \text{ is solved by } \underline{v}_1 = \begin{pmatrix}1\\ 1\end{pmatrix}$$

$$\omega_2^2: \qquad \begin{pmatrix} -D/m & -D/m \\ -D/m & -D/m \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0 \text{ is solved by } \underline{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Thus we get the following four **special solutions**, because each of the two values for ω^2 gives rise to two solutions, one with ω and the other with $-\omega$.

$$\underline{s}_1^{\pm} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{\pm i\omega_1 t} \quad \text{and} \quad \underline{s}_2^{\pm} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{\pm i\omega_2 t}.$$

Every **general solution** is a linear combination, or superposition in physics jargon, of the four special solutions, and thus is of the form

$$\begin{aligned} \underline{s} &= a_1\underline{s}_1^+ + b_1\underline{s}_1^- + a_2\underline{s}_2^+ + b_2\underline{s}_2^- \\ &= \begin{pmatrix} a_1e^{i\omega_1 t} + b_1e^{-i\omega_1 t} + a_2e^{i\omega_2 t} + b_2e^{-i\omega_2 t} \\ a_1e^{i\omega_1 t} + b_1e^{-i\omega_1 t} - a_2e^{i\omega_2 t} - b_2e^{-i\omega_2 t} \end{pmatrix} . \end{aligned}$$

Interpretation. Clearly there is not much to say about a general solution, as it is simply too difficult. But the special solutions correspond to very natural movements of the pendulums, namely synchronous and asynchronous oscillation. By this I mean that both either always move in the same direction or they always move in opposite direction. The first case corresponds to $\underline{s}_1^{\pm}$ and then the spring has no effect, as $\omega_1 = \pm\sqrt{g/l}$ which is just the frequency of a single pendulum of length l. The second case corresponds to $\underline{s}_2^{\pm}$ and here the frequency depends on D and also on m, as $\omega_2 = \pm\sqrt{g/l + D/m}$. These special solutions are called eigenstates of the system and the corresponding frequencies are the eigenfrequencies. This terminology should be clear from (22) which has the form $(A - \omega^2 I_2)\underline{v} = \underline{0}$.

Lecture 12

10 Scalar product and orthogonal matrices

The aim of this lecture is to explain how the geometry in $\mathbb{R}^n$ can be defined entirely in terms of linear algebra.

For $\underline{v}, \underline{w} \in \mathbb{R}^n$, their *dot product* or *scalar product* is defined by

$$\underline{v}.\underline{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i.$$

Properties. For all $\underline{v}, \underline{w}, \underline{v}_1, \underline{v}_2, \underline{w}_1, \underline{w}_2 \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ the following hold.

1. $(\underline{v}_1 + \underline{v}_2).\underline{w} = \underline{v}_1.\underline{w} + \underline{v}_2.\underline{w}$
2. $(\alpha \underline{v}).\underline{w} = \alpha(\underline{v}.\underline{w}$
3. $\underline{v}.(\underline{w}_1 + \underline{w}_2) = \underline{v}.\underline{w}_1 + \underline{v}.\underline{w}_2$
4. $\underline{v}.(\alpha \underline{w}) = \alpha(\underline{v}.\underline{w}$
5. $\underline{v}.\underline{w} = \underline{w}.\underline{v}$
6. $\underline{v}.\underline{v} \geq 0$ and $\underline{v}.\underline{v} = 0$ if and only of $\underline{v} = \underline{0}$

We now define the *length* or *norm* of $\underline{v} \in \mathbb{R}^n$ by

$$\|\underline{v}\| = \sqrt{\underline{v}.\underline{v}}.$$

It follows from 6. above that $\|\underline{v}\| = 0$ if and only if $\underline{v} = \underline{0}$.

And this allows us to define the *distance* between two vectors $\underline{v}$ and $\underline{w}$ by

$$d(\underline{v}, \underline{w}) = \|\underline{v} - \underline{w}\|.$$

The following is an often used result.

***Theorem* 19 (Cauchy-Schwarz inequality)** *For all $\underline{v}, \underline{w} \in \mathbb{R}^n$,*

$$|\underline{v}.\underline{w}| \leq \|\underline{v}\| \, \|\underline{w}\|$$

holds true.

Proof. First, if $\underline{w} = \underline{0}$, then both sides are zero, and we are done. So we assume $\underline{w} \neq \underline{0}$ and put $\lambda = \underline{w}.\underline{w}$ and $\mu = -\underline{v}.\underline{w}$. Then

$$\begin{aligned} 0 &\leq (\lambda\underline{v} + \mu\underline{w}).(\lambda\underline{v} + \mu\underline{w}) && \text{(by property 6.)} \\ &= \lambda^2\underline{v}.\underline{v} + 2\lambda\mu\underline{v}.\underline{w} + \mu^2\underline{w}.\underline{w} && \text{(using properties 1.-5.)} \\ &= \lambda\left((\underline{w}.\underline{w})(\underline{v}.\underline{v}) - 2(\underline{v}.\underline{w})^2 + (\underline{v}.\underline{w})^2\right) && \text{(by definition of } \lambda \text{ and } \mu) \\ &= \lambda(\|\underline{w}\|^2\|\underline{v}\|^2 - (\underline{v}.\underline{w})^2) && \text{(by definition of norm)} \end{aligned}$$

Since $\lambda > 0$, it follows that $(\underline{v}.\underline{w})^2 \leq \|\underline{w}\|^2\|\underline{v}\|^2$ which implies the result, as the square root is monotone increasing, i.e. $\sqrt{y} > \sqrt{z}$ whenever $y > z$, $0 \leq y, z \in \mathbb{R}$.

The Cauchy-Schwarz inequality implies the
triangle inequality: $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$.

This follows from monotony of the square root again and

$$\begin{aligned} \|\underline{v} + \underline{w}\|^2 &= (\underline{v} + \underline{w}).(\underline{v} + \underline{w}) \\ &= \underline{v}.\underline{v} + 2\underline{v}.\underline{w} + \underline{w}.\underline{w} && \text{(by properties 1.-5.)} \\ &\leq \|\underline{v}\|^2 + 2\|\underline{v}\|\,\|\underline{w}\| + \|\underline{w}\|^2 && \text{(by Cauchy-Schwarz ineq.)} \\ &= (\|\underline{v}\| + \|\underline{w}\|)^2. \end{aligned}$$

Finally we can define angles which besides distance are the other main feature of geometry. Again let $\underline{v}, \underline{w} \in \mathbb{R}^n$ and assume $\underline{v} \neq \underline{0} \neq \underline{w}$. By the Cauchy-Schwarz inequality, we have

$$-1 \leq \frac{\underline{v}.\underline{w}}{\|\underline{v}\|\,\|\underline{w}\|} \leq 1.$$

So there is a unique $\alpha \in [0, \pi]$ with

$$\cos\alpha = \frac{\underline{v}.\underline{w}}{\|\underline{v}\|\,\|\underline{w}\|},$$

and we define the *angle between* $\underline{v}$ *and* $\underline{w}$ to be this α.

Observation. If, for $\underline{v} \in \mathbb{R}^n$, $\underline{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ say, we define $\underline{v}^t$ to be the row vector $(v_1, v_2, \ldots, v_n)$, then the dot product $\underline{v}.\underline{w}$ becomes the matrix product $\underline{v}^t\underline{w}$.

In $\mathbb{R}^n$ we have now defined the geometric concepts of length, distance, and angles entirely in terms of the dot product, or matrix multiplication if you like. And it becomes an interesting question which n-square matrices preserve this geometry when we view them as mappings from $\mathbb{R}^n$ to itself. More precisely, which matrices A satisfy $(A\underline{v}).(A\underline{w}) = \underline{v}.\underline{w}$ for all $\underline{v}, \underline{w} \in \mathbb{R}^n$?

In terms of matrix multiplication this question is the following. Which matrices A satisfy $(A\underline{v})^t A\underline{w} = \underline{v}^t\underline{w}$ for all$\underline{v}, \underline{w} \in \mathbb{R}^n$? By Question 3 on problem sheet 6, $(AB)^t = B^t A^t$ for all matrices for which AB is defined. Here A^t is, by definition, the matrix whose i-th row is the i-th column of A and it is called the *transpose* of A. Hence the question boils down to the problem of finding those matrices A for which $\underline{v}^t A^t A\underline{w} = \underline{v}^t\underline{w}$ holds for all $\underline{v}, \underline{w} \in \mathbb{R}^n$.

It turns out that these are precisely those matrices satisfying $A^t = A^{-1}$ which are called *orthogonal matrices*. One reason for this name is the fact that columns of an orthogonal matrix constitute a so called *orthonormal basis*; that is each basis vector has length one, and distinct basis vectors are orthogonal to each other. Notice that non-zero vectors $\underline{v}$ and $\underline{w}$are orthogonal if and only if $\underline{v}.\underline{w} = 0$. Orthogonal matrices come up in different areas of physics, basically because they describe those transformations of Euclidean n-space which preserve the geometry.

WARNING! All the material of Section 10 is only valid in $\mathbb{R}^n$. In complex vector spaces things get a little more difficult, one reason being that property 6. (which we used a couple of times) does not hold any more, as can be seen from $(i, i)\begin{pmatrix} i \\ i \end{pmatrix} = -2$ for instance.