

LINEAR ALGEBRA (MATH 115)

MOSHE KAMENSKY

Note: References in **bold** (such as **Example 4.1-2**) refer to [Nic06].

1. OVERVIEW

1.1. Linear equations. Linear algebra can be viewed as the study of systems of *linear equations*. Such a system has the form:

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

Here, the x_i are called the *unknown variables* or *indeterminates* of the system. The a_{ij} and b_i , on the other hand, are called parameters. Their value is unspecified, but is assumed to be known. A *solution* of the system is an assignment of values to the x_i for which all the equalities above hold. Thus, each solution of the system above is a sequence of n numbers. When the number of variables is small, we will usually use the letter x, y, z, \dots instead of x_1, x_2, x_3, \dots .

As an example, consider the following system:

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

The right hand side of each equation here is 0. Such systems are called *homogeneous*. A homogeneous system always has at least one solution, namely, the one where all variables are assigned the value 0. However, the system may have additional solutions. Indeed, the system above is also solved by the tuple $(1, 0, -1)$. In the following lecture, we will see how one can find solutions of such systems, and describe the set of all solutions (or prove that such solutions do not exist.) We may do this right away with the simplest kind of system: one equation in one variable.

Such a “system” has the form $ax = b$. Assume first that $a \neq 0$. In this case we may divide by a , and conclude that there is at most one solution, namely b/a . Substitution then shows that this is indeed a solution. At this point it proves useful to observe that we always seek solutions in a “number system” where we may divide by any non-zero number (can you think of a situation where this does not hold?) On the other hand, assume that $a = 0$. Then we again have two cases. If $b \neq 0$, there is no solution. If $b = 0$, any number will solve the equation. We thus classified the possible sets of solutions: it is either empty, or everything, or consists of one solution. In the last case, we also found the solution explicitly. A similar analysis, though more complicated, holds for general systems.

Date: November 29, 2007.

Systems of linear equations arise both in purely mathematical context, and in applications. We shall see examples of the later during the course. An important example of the former is 2 and 3 dimensional geometry which we shall start discussing now.

1.2. Vector geometry. One of the most important discoveries in Mathematics, made by René Descartes in the 17th century ([Wik07]), is that the set of points in the plane can be identified with the set of pairs of real numbers. To do this, we need to fix a pair of perpendicular lines, and a unit of length. The pair of lines is called the x and y axes, drawn horizontally and vertically, respectively. Once this is chosen, we may map a point to a pair of numbers, called its *coordinates*, as follows: given a point P , there is a unique line passing through P and parallel to the y -axis. This line intersects the x -axis at a unique point P_x . The x -coordinate of P is the length of the interval from the point 0 where the axes intersect, to P_x (using the fixed length unit.) The y -coordinate is obtain similarly, by swapping x and y . Reversing the process, we see that to every pair of numbers, we may associate a unique point in the plane.

Similarly, 3-space can be identified with triples of real numbers, by fixing three perpendicular lines intersecting at the same point, and a unit of length.

Given this identification, we may start asking about the relation between geometric properties of points, lines, etc. in the plane (or space), and the algebraic properties of the corresponding tuples (or sets of tuples) of numbers. For example, we saw above that the solution set of a system of linear equations in two variables is a set of tuples. Given such a system, we may thus consider its set of solutions as a set of points in the plane. What shape may this set have? Answers to this and similar questions provide intuition, and sometimes tools, that allow us to develop the theory of linear equations.

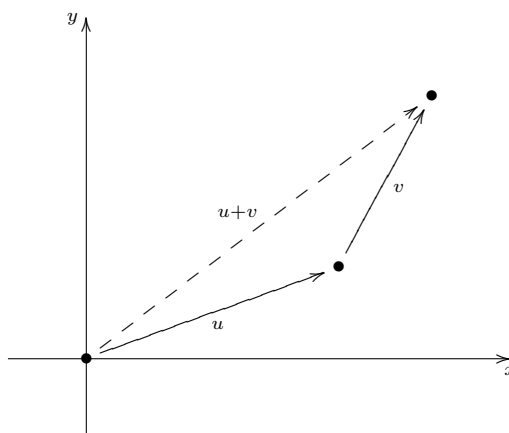
Conversely, given a line in the plane, can we describe it algebraically (say, using some finite tuple of numbers)? Can we deduce from this description, whether (say) given two lines are parallel? We will see that these questions amount to questions about linear equations, and that this correspondence allows us to solve geometric problems using algebra.

Following [Lan89], we say that a *located vector* in the plane (or in 3-space) is simply an order pair of points. The first is called the tail, and the second the tip (or head) of the located vector. We visualise it as an arrow going from the tail to the tip. Two such located vectors are *equivalent* if they have the same length and direction. In other words, if one can be translated to the other. A *vector* is a located vector up to this equivalence. Thus, a vector has length and direction, but not head or tip.

Once we choose a coordinate system, any located vector is equivalent to a unique one whose tail is the origin, and therefore, we may identify vectors with points in the plane (and with tuples of numbers.) However, we will see that it is sometimes helpful to choose a different representative.

1.3. vector spaces. There are natural operations that can be defined on tuples of numbers. If $\bar{x} = (x_1, \dots, x_n)$ and $\bar{y} = (y_1, \dots, y_n)$ are two such tuples, we may define their sum $\bar{x} + \bar{y}$ to be the tuple $(x_1 + y_1, \dots, x_n + y_n)$. Likewise, if \bar{x} is a tuple and c is a number, we may define $c\bar{x}$ as (cx_1, \dots, cx_n) .

Such operations can also be defined on vectors in the plane. If u and v are two vectors, we define their sum as follows: choose a representative of v whose tail is equal to the tip of u (when represented with tail at the origin.) The sum is then the vector whose tip is the tip of v and whose tail is the origin, as in the following picture:



Likewise, if c is a positive number, and u is a vector, cu is the vector with the same direction, and with length $c|u|$ (where $|u|$ is the length of u .) If $c < 0$, cu is the vector of length $-c|u|$, in the direction opposite to u .

We will see that the identification of vectors with pairs of numbers preserves these operations: if \bar{x} and \bar{y} correspond to u and v , then $\bar{x} + \bar{y}$ corresponds to $u + v$, and $c\bar{x}$ corresponds to cu . A structure with these operations (that satisfies appropriate axioms) is called a *vector space*. As we shall see, vector spaces play a central role in the study of linear equations.

2. LINEAR GEOMETRY

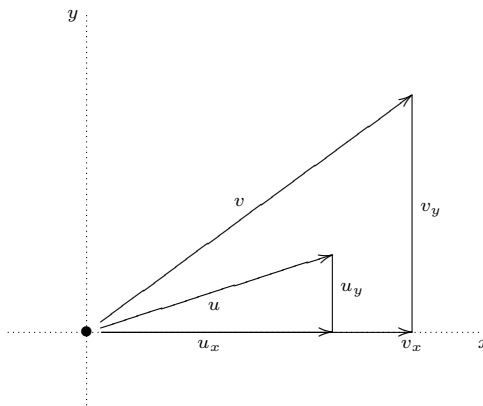
In this lecture we will study the algebraic properties of certain “linear” sets of points in the plane and in 3-space. We recall that by fixing axes, we may identify a (geometric) point with its coordinates, a tuple of (real) numbers. We also recall, that a *vector* is an “arrow” in the space, determined by its length and direction. By fixing a point 0 in space, we may associate with each vector a point in space, namely the head of the located vector representing the given vector, whose tail is at 0 . In particular, if we fix a coordinate system (i.e., axes and a length unit), we may take 0 to be the intersection point of the axes, and then we may identify a vector with a tuple, called its coordinates. We now express some geometric properties of the vector in terms of these coordinates.

We denote the coordinates of a vector v by (v_x, v_y, v_z) (or (v_x, v_y) in the plane), or just by (x, y, z) if there is only one vector around.

2.1. The length of a vector. The length $\|v\|$ of v is given by $\sqrt{x^2 + y^2 + z^2}$. This follows from Pythagoras’ theorem. In particular, we see that $v = 0$ if and only if $\|v\| = 0$. We note that the length is computed in terms of the chosen length unit.

Example 1. Compute the lengths of the vectors represented by $(1, -2, 2)$ and $(4, -3)$.

2.2. The direction of a vector. We claim that non-zero v and u are *parallel* if and only if $v_x = u_x = 0$ or $v_y/v_x = u_y/u_x$ (in other words, $v_y u_x = u_y v_x$.) Indeed, assume that all coordinates are positive (the other cases are similar), and consider the triangles whose vertices are $0, (v_x, v_y), (v_x, 0)$ and $0, (u_x, u_y), (u_x, 0)$. They both have a right angle, so they are similar if and only if the vectors have the same direction. On the other hand, they are similar if and only if the ratios between the sides that meet on the right angle is the same. These ratios are v_y/v_x and u_y/u_x .



It follows that u and v have the same direction if and only if the above ratios are the same, the coordinates have the same sign. Likewise, in 3 dimensions, the vectors are parallel iff the ratios between all their components are the same, and they have the same direction iff in addition, all coordinates have the same sign.

In the plane, we may measure the direction of the vector v by looking at the angle θ between it and the x -axis, counter-clockwise (so the vector represented by $(0, 1)$ will have angle $\pi/2$.) Recall that the ratio v_y/v_x we considered are then denoted $\tan(\theta)$. The ratio $v_x/\|v\|$ is denoted by $\cos(\theta)$. We have the following formula that connects the lengths of sides of a triangle:

$$c^2 = a^2 + b^2 - 2ab \cos(\theta) \quad (1)$$

here, a , b and c are the lengths, and θ is the angle opposite the side of length c . This is called the *law of cosines*.

Example 2. Which of the following vectors have the same direction? Which are parallel?

- (1) $(1, 3, 2)$
- (2) $(-2, -6, -4)$
- (3) $(-1, 3, 2)$
- (4) $(3, 9, 6)$

Example 3. Is there a triangle with sides of length 1, 2, 4?

2.3. Multiplication by scalars. Recall that we defined multiplication of a tuple $\bar{x} = (x_1, \dots, x_n)$ by a number c via $c\bar{x} = (cx_1, \dots, cx_n)$. On the other hand we defined cv , when v is a geometric vector, to be, in case $c > 0$, the vector with the same direction as v , and with length $c\|v\|$ (if $c < 0$, it will be the opposite of $-cv$.) We will now show that these operations are compatible with the translation between vectors and tuples.

We need to show that (cv_x, cv_y) has the same length and direction as cv . By definition, the length of cv is $|c|\|v\|$. On the other hand, the length of (cv_x, cv_y) is

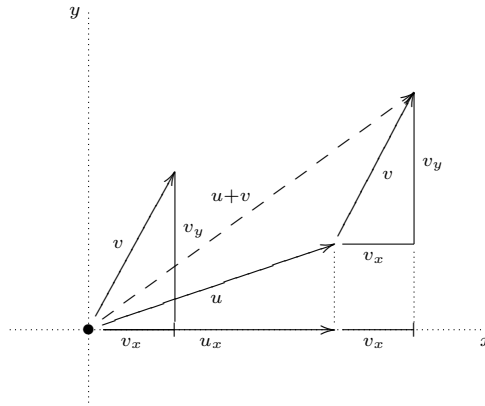
$$\sqrt{(cv_x)^2 + (cv_y)^2} = |c|\sqrt{v_x^2 + v_y^2} = |c|\|v\| \quad (2)$$

For the direction, assume that $c > 0$. Then (cv_x, cv_y) has the same direction as (v_x, v_y) , as we saw in 2.2, and this is the same as the direction of cv . The case $c < 0$ is similar.

Example 4. If $v \neq 0$, show that $v/\|v\|$ is the unique vector in the same direction as v with length 1. This is named a *unit vector* in that direction.

Start of lecture 3

2.4. Addition of vectors. Recall that we defined addition of vectors via the parallelogram law, while addition of tuples is defined term-wise. As with multiplication by scalars, these definitions are compatible. This follows from congruence of the right triangle determined by one of the vectors:

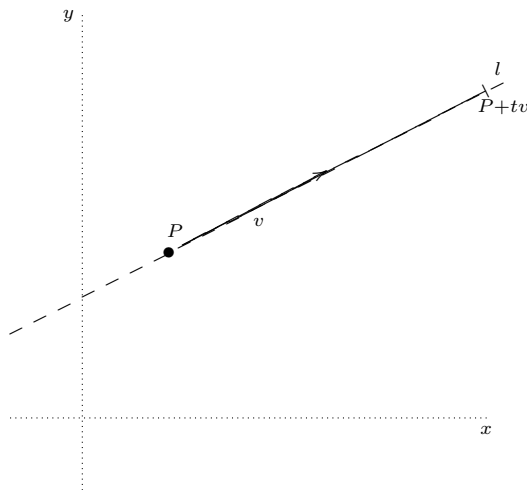


Example 5 (Example 4.1-2). Show that the diagonals of a parallelogram bisect each other.

Example 6 (Example 4.1-6). Show that the midpoints of the sides of any quadrangle form a parallelogram.

2.5. Lines in space. A line l in space is determined by a direction (more precisely a *parallelism class*), and a point through which it goes. Given a point P , the line through P in a given direction is the set of all points Q such that the vector defined by P, Q is parallel to the given direction. When the direction is given by a vector $v \neq 0$, we conclude from the previous results that a general point on the line through P in the direction of v has the form $P + tv$, for arbitrary t (we identified P with the vector it defines, and then identified the resulting vector with the point it defines.) If P has coordinates (P_x, P_y, P_z) , then the results above imply that a

general point on the line has coordinates $(P_x + tv_x, P_y + tv_y, P_z + tv_z)$.



Example 7 (Example 4.1-8). Find the equations for the line through $(2, 0, 1)$ and $(4, -1, 1)$.

Example 8 (Example 4.1-10). Determine the intersection point (if exists) of the lines determined by $P = (1, 2, 1), v = (-3, 5, 1)$ and $P = (-1, 3, 1), v = (1, -4, -1)$.

2.6. Scalar products. We have defined the operations of addition between vectors, and multiplication of a vector by a scalar (number.) Both of these yield a vector as a result. We now define an operation between tuples that yields a *scalar*. For tuples $\bar{x} = (x_1, \dots, x_n)$ and $\bar{y} = (y_1, \dots, y_n)$, the *scalar product* (also called the *inner product* or the *dot product*) is defined to be $x_1y_1 + \dots + x_ny_n$. It is denoted by $\bar{x} \cdot \bar{y}$ or by $\langle \bar{x}, \bar{y} \rangle$. The scalar product has the following (easily verified) properties:

- (1) $\bar{x} \cdot \bar{y} = \bar{y} \cdot \bar{x}$
- (2) $\bar{x} \cdot \mathbf{0} = 0$
- (3) $\bar{x} \cdot \bar{x} = \|\bar{x}\|^2$
- (4) $(c\bar{x}) \cdot \bar{y} = c(\bar{x} \cdot \bar{y})$
- (5) $\bar{x} \cdot (\bar{y} + \bar{z}) = \bar{x} \cdot \bar{y} + \bar{x} \cdot \bar{z}$

Example 9 (Example 4.2-2). Compute $\|\bar{x} - 3\bar{y}\|$ if $\|\bar{x}\| = 2$, $\|\bar{y}\| = 1$ and $\bar{x} \cdot \bar{y} = 2$.

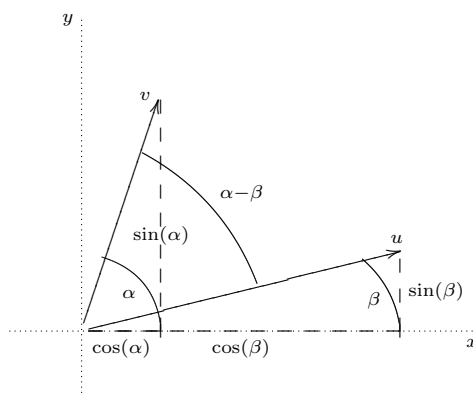
We now interpret the scalar product from the geometric point of view. The basic formula, which follows from the law of cosines, is

$$u \cdot v = \|u\| \|v\| \cos(\theta) \quad (3)$$

where θ is the angle between u and v . We note that we may restrict the attention to $0 \leq \theta \leq \pi$. In this range, $\cos(\theta) = 1$ iff $\theta = 0$, and $\cos(\theta) = 0$ iff $\theta = \pi/2$. We thus conclude that $u \cdot v = \|u\| \|v\|$ iff u and v are parallel, and u is perpendicular to v iff $u \cdot v = 0$. In general $u \cdot v$ carries information about the lengths of u and v , as well as the angle between them.

Example 10 (Example 4.2-3). Compute the angle between $(-1, 1, 2)$ and $(2, 1, -1)$.

Example 11. Show that for any α, β , $\cos(\alpha - \beta) = \cos(\alpha) \cos(\beta) + \sin(\alpha) \sin(\beta)$



Example 12 (Example 4.2-5). Diagonals of a parallelogram are perpendicular iff it is a rhombus.

Example 13. If u and v are orthogonal to w , then so are $u + v$ and cu for any number c .

Start of lecture 4

2.7. Projections. Let v be a non-zero vector, and let u be a vector. Using the scalar product, we may compute the *projection* of u on v . By definition, this is a vector w parallel to v , such that $u - w$ is orthogonal to v . Since w is parallel to v , we may write $w = cv$ for some number c . Using the scalar product description of orthogonality, we get the equation $(u - cv) \cdot v = 0$. Solving for c , we get that $c = u \cdot v / \|v\|^2$.

This construction has the following geometric descriptions: w is the unique vector parallel to v such that the triangle determined by u, w has a right angle; and $w - u$ is the shortest vector from the tip of u to a point on the line determined by v (so $\|w - u\|$ is the distance from the tip of u to this line.)

Example 14 (Example 4.2-8). Find the shortest distance from $(1, 3, -2)$ to the line through $(2, 0, -1)$ in the direction $(1, -1, 0)$.

2.8. Planes. Consider a plane in 3-space. Given a point P on this plane, there is a unique line through it that is perpendicular to the plane. If we fix P as a base point (so that vectors are represented with their tail at P), we see that any vector v representing the given line is orthogonal to every vector that lies in the plane. Conversely, every vector orthogonal to v is contained in the plane. Therefore, using the scalar product, a plane is given by the set of all points p solving the equation $v \cdot (p - p_0) = 0$, for a fixed vector v and point p_0 (the same plane can be represented by many such equations.) To write this in coordinates, let $v = (a, b, c)$, $p_0 = (x_0, y_0, z_0)$ and $p = (x, y, z)$. Then we get

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \quad (4)$$

Example 15. Show that any plane is given by an equation $ax + by + cz = d$, where $abc \neq 0$, and any such equation is the equation of a plane. What is the condition for the plane to contain 0?

Example 16 (Example 4.2-11). Find the distance from $(2, 1, -3)$ to the plane $3x - y + 4z = 1$.

3. LINEAR EQUATIONS

A *linear equation* is an equation of the form $a_1x_1 + \cdots + a_nx_n = b$. This equation has x_1, \dots, x_n as indeterminates, and a_1, \dots, a_n are their coefficients. b is called the constant term. A *system of linear equations* is a finite set of equations of this kind. A *solution* of such a system is a tuple (c_1, \dots, c_n) , such that all the equations are satisfied when \bar{c} is substituted for \bar{x} . The system is called *homogeneous* if all the constant terms are 0. A homogeneous system always has at least one solution: 0. We have seen that in dimension 3 (when $n = 3$), one linear equation generically defines a plane, and two linear equations (generically) define a line.

We would like to find a way to solve systems of linear equations. What do we mean by solving? We would like to describe *all* solutions of the system. Since there might be infinitely many, we can't simply list them. Instead, we shall see that we can describe the solutions in the following form:

$$\begin{aligned}x_1 &= d_{1,0} + d_{1,1}y_1 + \cdots + d_{1,k}y_k \\x_2 &= d_{2,0} + d_{2,1}y_1 + \cdots + d_{2,k}y_k \\&\dots\end{aligned}$$

where the y_i are *arbitrary*. We note that this last set of equations is again a system of linear equations, of a special form. It is equivalent to the original system, in the sense that they have the same sets of solutions. Our task is thus to transform an arbitrary system into an equivalent one, which has a special form.

3.1. Equivalent systems. Two systems a linear equations are *equivalent* if they have the same set of solutions. We will examine some simple ways to get from a system to an equivalent one.

Let $a_1x_1 + \cdots + a_nx_n = b$ be an equation, and consider the equation $ca_1x_1 + \cdots + ca_nx_n = cb$, obtained by multiplying both sides of the equation by a number c . If \bar{x} is a solution of the first equation, it will also solve the second. The converse need not be true: if $c = 0$, then any tuple solves the second equation. However, if $c \neq 0$, then any solution of the second equation also solves the first one. This is true since the first equation is obtained from the second by multiplication by $1/c$. The main point is that the operation of multiplying an equation by an *non-zero* number can be inverted.

Here are examples of operations of the same kind on systems of equations:

- (1) Multiplication of an equation by a non-zero scalar
- (2) Adding one equation to another

These are called *elementary operations* on the equations. These operations transform the system to an equivalent one. Later we will see that any system can be transformed to any equivalent one by a sequence elementary operations.

3.2. Special forms of a system. We now describe more precisely what kind of form we would like the linear system to have. A system is said to be in *row echelon form* if the following conditions hold:

- (1) The first non-zero coefficient in each equation is 1 (if exists)
- (2) For any two equations the first non-zero entry (if exists) is different

If a system is in this form, we can order the equations according to their first non-zero coefficient (and all equations all of whose coefficients are zero at the end.) In

this ordering, they may look like this:

$$\begin{aligned} x_{k_1} + a_{1,k_1+1}x_{k_1+1} + a_{1,k_1+2}x_{k_1+2} + a_{1,k_1+3}x_{k_1+3} + \cdots + a_{1,n}x_n &= b_1 \\ x_{k_1+2} + a_{2,k_1+3}x_{k_1+3} + \cdots + a_{2,n}x_n &= b_2 \\ x_{k_1+3} + \cdots + a_{3,n}x_n &= b_3 \\ &\dots \\ 0 &= b_m \end{aligned}$$

It is in *reduced row echelon form* if, in addition, a variable that appears as the variable in an equation, does not appear in any other equation (so, in the above example, $a_{1,k_1+2} = a_{1,k_1+3} = a_{2,k_1+3} = 0$.) This form is what we called earlier a solution: first, a system will have solutions if and only if the constant term is 0 for any equations each of whose coefficients are 0. If that is the case, let us call all the variables that do not appear first in any equation *parameters*. Then for any assignment of arbitrary values to the parameters, we get a solution to the equations (and all solutions are of this form.) What is the condition to have a unique solution?

We now explain that any system can be transformed to a (reduced) row echelon form. This goes under the name *Gauss elimination*. It goes as follows: we first find an equation with a non-zero coefficient a of minimal index k (if there is no such equation, the all equations are 0, and we are done!). Since a is non-zero, we may divide by a , to obtain an equation with the first coefficient 1. We now subtract suitable multiples of this equation from any other equation, to obtain equations with zero coefficient at that index. At this point all other equations have only variables of index higher than k . We now remove the equation we used, and perform the same operations with the rest of the equations. Since the number of variables (and equation) is strictly less on each stage, we are bound to stop. This gives an equivalent system in echelon form. To obtain a reduced form, we now subtract an appropriate multiple of each equation from any equation that contains its leading variable.

We summarise what we showed:

Theorem 17. *Any system is equivalent to a system in reduced echelon form (and we know how to compute it.)*

Example 18 (Example 1.2-2). Solve the system

$$\begin{aligned} x + 10z &= 5 \\ 3x + y - 4z &= -1 \\ 4x + y + 6z &= 1 \end{aligned}$$

Example 19 (Example 1.2-4). For which values of (a, b, c) does the following system have a solution?

$$\begin{aligned} x + 3y + z &= a \\ -x - 2y + z &= b \\ 3x + 7y - z &= c \end{aligned}$$

3.3. Uniqueness and rank. Given a system, there are many ways to perform the Gauss elimination on it, and many equivalent row echelon forms. However, we have the following theorem:

Theorem 20. *A given system has exactly one equivalent reduced echelon form*

Proof. The existence was proved above. Let A and B be two equivalent systems in reduced row echelon form, and assume they are different. For an equation in A (or in B), let us call the index of the first non-zero coefficient the index of the equation. We may assume, by symmetry, that A contains an equation not in B , of highest non-zero index k . There are two cases:

B contains an equation of index k : Let E be the difference of the two equations. Since the two equations are distinct, the difference is not the trivial equation $0 = 0$, and therefore there is a tuple \bar{x} that does *not* solve E . By minimality of k , all variables appearing in E are parameter variables in both A and B . Therefore, any assignment to them can be extended to a solution of A (and B). Extending \bar{x} to such a solution, we find a solution of all equation of A and B , but not of the difference E of two of the equations. This is a contradiction.

x_k is a parameter in B : By minimality, all parameters in given equation are also parameters in B . An assignment to these parameters determines uniquely the value of x_k in A , but not in B . This is a contradiction.

□

Corollary 21. *If A and B are equivalent systems, it is possible to reach from A to B by elementary operations.*

Since the reduced row echelon form is unique for a given equivalence class of equations, we may use it to define quantities that depend only on the set of solutions, and not on the particular system of equations, or the way we implement Gaussian elimination. For example, the *rank* of a (consistent) system is defined to be the number of non-zero equations in the (reduced) row echelon form. The uniqueness tells us that the rank is well defined, and depends only on the solution set.

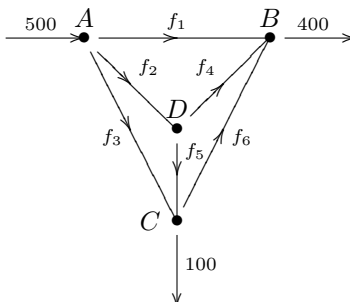
Example 22. If the number of variables is 3, a system of equations defines a subset of 3 space. Equivalent systems define the same subset. The system will have ranks 0, 1, 2 and 3 if the subset is, respectively, the whole space, a plane, a line and a point. Thus, the rank carries information about the “size” of the set of solutions.

Example 23. If a consistent system of equations has more variables than equations, then it has infinitely many solutions: since the number of equation in the row echelon form can only go down, there are more than zero parameters. Since the system is consistent, any assignment to these parameters gives a solution. In particular, a homogeneous system with more variables than equations has infinitely many solutions.

Example 24 (Example 1.3-2). Show that there is a conic through any five points not on the same line.

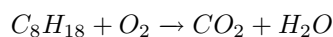
3.4. Examples.

Example 25 (Example 1.4-1). Compute the possible flows of cars in the following network of one-way roads:



Example 26 (Example 1.5-1).

Example 27 (Example 1.6-1). Balance the reaction



4. MATRICES

Given a system of linear equations,

$$a_{1,1}x_1 + \cdots + a_{1,n}x_n = b_1$$

$$a_{2,1}x_1 + \cdots + a_{2,n}x_n = b_2$$

...

$$a_{m,1}x_1 + \cdots + a_{m,n}x_n = b_m$$

we may organise the coefficients in a *matrix*, as follows

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ \cdots & \cdots & \cdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

We will say that this matrix has size (or dimensions) $m \times n$. Writing \bar{x} for the tuple of variables, and \bar{b} for the tuple of constant terms, we will see that the system of equations can be written as

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

This suggests that a system of linear equations can be thought of in analogy with a single linear equation $ax = b$, where the *number* a is replaced by a *matrix* A . To explain the notation, and the analogy, we study algebraic operations defined on matrices.

4.1. Vector space operations. A matrix can be regarded as a tuple, written in a funny way: if we write the rows of a matrix (of size $m \times n$) in a sequence, we get a tuple (of size mn .) Thus, a matrix can be thought of as a tuple, together with the *extra information* of the size n of each row. Viewed this way, we see that the familiar operations of addition of two tuples, and multiplication of a tuple by a scalar can be applied to matrices of a given size. All axioms that hold for tuples will hold automatically for matrices. Thus, if A, B and C are matrices of a given size, and x, y are numbers, the following hold:

(1) $A + B = B + A$

- (2) $A + (B + C) = (A + B) + C$
- (3) $A + 0 = A$
- (4) $-A + A = 0$
- (5) $x(A + B) = xA + xB$
- (6) $(x + y)A = xA + yA$
- (7) $(xy)A = x(yA)$
- (8) $1A = A$

(Thus, in the terminology of 5, the set of matrices of size $m \times n$ forms a *vector space*.)

We stress that addition is only defined between matrices of the *same size*.

4.2. Transpose. The transpose A^T of a matrix A is defined by the formula $(A^T)_{i,j} = A_{j,i}$. Thus it is the reflection of A along the main diagonal. In other words, the rows of A^T are the columns of A . We see from any of these definitions that the following holds:

- (1) $(A^T)^T = A$
- (2) $(xA)^T = xA^T$
- (3) $(A + B)^T = A^T + B^T$

A matrix is called *symmetric* if $A^T = A$. Such a matrix has to be *square* (of size $n \times n$ for some n .)

Example 28 (Example 2.1-11). The set of symmetric matrices is closed under the vector space operations.

Example 29 (Example 2.1-12). If $A = 2A^T$ then $A = 0$.

4.3. Matrix multiplication. Let A be the matrix associated to a system L of linear equations. Any equation in L has the form $a_{i,1}x_1 + \cdots + a_{i,n}x_n = b_i$, for some row i of the matrix A . If we write \bar{a}_i for the tuple of which this row consists, we see that the LHS of the above equation can be written as the scalar product $\bar{a}_i \cdot \bar{x}$ of \bar{a}_i and \bar{x} . Thus the system L can be written as

$$\begin{aligned} \bar{a}_1 \cdot \bar{x} &= b_1 \\ &\dots \\ \bar{a}_m \cdot \bar{x} &= b_m \end{aligned}$$

This motivates the following definition:

definition 30. The *product* AB of the matrices A of size $m \times n$ and B of size $n \times k$ is a matrix of size $m \times k$, with $(AB)_{i,j} = A_{i,-} \cdot B_{-,j}$, where $A_{i,-}$ is the i -th row of A , and $B_{-,j}$ is the j -th row of B .

Thus, if \bar{x} and \bar{b} above are both considered as $m \times 1$ matrices (*column vectors*), the system can be written as $A\bar{x} = \bar{b}$.

Example 31 (Example 2.2-2). Compute the product of

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

and

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

Example 32 (Example 2.2-4). Let $A = \begin{bmatrix} 6 & 9 \\ -4 & -6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$. Compute A^2 , AB and BA .

The *identity matrix* of size n is the square $n \times n$ matrix I that has 1 on the main diagonal and 0 elsewhere. The following theorem lists the properties of the product:

Theorem 33. *The following identities between matrices hold (assuming they make sense)*

- (1) $IA = A, BI = B$
- (2) $A(BC) = (AB)C$
- (3) $A(B + C) = AB + AC$
- (4) $(B + C)A = BA + CA$
- (5) $k(AB) = A(kB)$
- (6) $(AB)^T = B^T A^T$

We note that the i -th column of AB is, by definition, $AB_{-,i}$ (where $B_{-,i}$ is the i -th column of B .) Similarly, the i -th row of AB is $A_{i,-}B$. Using this observation (or directly from the definition), we see that it is enough to prove, e.g., part 3 for the case that A is a row vector, and B is a column vector. The other parts also follow easily from the definition, except for 2, which will be proved later.

We note that in general AB is *not* equal to BA .

Example 34 (Example 2.2-7). $AB = BA$ iff $(A - B)(A + B) = A^2 - B^2$

To properly understand the meaning of the product of matrices, and to prove more easily the above properties, we need to re-interpret matrices as linear transformations.

5. VECTOR SPACES AND LINEAR TRANSFORMATIONS

As we saw several times, the operations of addition of tuples, and multiplication by a scalar play a central role in the theory. The existence and properties of these operations can be summarised by saying that the set of n -tuples is a vector space.

definition 35. A *vector space* is a set V , together with an operation $+$ between elements of V , an operation of multiplication by numbers on elements of V , and an element 0 . For any elements u, v and w of V , and for any numbers x and y , the following properties should hold:

- (1) $u + v = v + u$
- (2) $u + (v + w) = (u + v) + w$
- (3) $u + 0 = u$
- (4) $x(u + v) = xu + xv$
- (5) $(x + y)u = xu + yu$
- (6) $(xy)u = x(yu)$
- (7) $1u = u$

Example 36. The set \mathbb{k}^n of n -tuples of numbers is a vector space

Example 37. The set of geometric vectors in the plane (or in 3-space) is a vector space. In honor of this example, an element of an arbitrary vector space is called a vector.

Example 38. The set of matrices of size $m \times n$ is a vector space

Example 39. The set of solutions of a homogeneous system of linear equations is a vector space

We have seen that by choosing a coordinate system, we may identify a vector v in the plane with a pair of numbers $T(v)$. We noted that this identification satisfies the following properties: $T(u+v) = T(u) + T(v)$, and $T(xu) = xT(u)$ (where x is a number.) This can be summarised by saying that T is a linear transformation from the vector space of vectors in the plane to the vector space of pairs of numbers.

definition 40. Let U and V be two vector spaces. A *linear map* (or *linear transformation*) T from U to V is a map associating to any element u of U an element $T(u)$ of V , such that for any $u, w \in U$ and number x , $T(u+w) = Tu + Tw$ and $T(xu) = xT(u)$.

Example 41. Let U be the space of triples, and V the space of pairs. The map that sends a triple (x, y, z) to (x, y) is a linear transformation from U to V . The map $(x, y) \mapsto (y, x)$ is a linear transformation from V to itself, and so is the map $(x, y) \mapsto (x + 2y, 0)$.

Example 42. If T and S are two linear maps from U to V , then so are $T+S$ (defined as $(T+S)(u) = Tu + Su$) and xT , where x is a number (where $(xT)(u) = x(T(u))$.) In other words, the set of linear maps from U to V is itself a vector space!

An important example arises from linear equations. Let U be the set of solutions of an homogeneous system of linear equations. We observed above that this is a vector space. When the system is brought into reduced row echelon form, each tuple substituted for the parameters gives a solution to the system. If the number of parameter variables is m , let T be the map from \mathbb{k}^m to U that assigns to each tuple the solution obtained in this way. It is immediate that T is a linear map. Since any solution is obtained in this way uniquely, we see that T identifies the set of solutions with \mathbb{k}^m .

What is m ? Recall that the rank r of the system was defined to be the number of equations in the reduced form. Since any equation contains exactly one non-parameter variable (and these are all different), this is also the number of non-parameter variables. We thus get that m is $n - r$, where n is the total number of variables.

5.1. Bases. Matrices provide an example of of linear transformations: if A is an $m \times n$ matrix, and \bar{x} is an n -tuple, viewed as an $n \times 1$ matrix, then $A\bar{x}$ is an m -tuple. Thus A defines a map from \mathbb{k}^n to \mathbb{k}^m . This map is a linear transformation by 3 and 5 of theorem 33.

To study the relation between linear transformations and matrices it is convenient to introduce bases. If V is a vector space, and $v_1, \dots, v_n \in V$ are vectors, an expression of the form $a_1v_1 + \dots + a_nv_n$ is called a *linear combination* of v_1, \dots, v_n . A *basis* for V is a sequence v_1, \dots, v_n such that any vector in V can be presented uniquely as a linear combination of v_1, \dots, v_n .

Example 43. Let V be the space \mathbb{k}^n , and consider the vectors e_i that have 1 on the i -th place, and 0 elsewhere. Any n -tuple \bar{a} can be written (uniquely!) as $\bar{a} = a_1e_1 + \dots + a_ne_n$. Thus the e_i form a basis for \mathbb{k}^n , called the *standard basis*.

It is a fundamental fact that any vector space has a basis, and that the size of any two bases for the same vector space is the same. However, we do not currently need these results. Instead we will use just the standard basis.

- Theorem 44.** (1) For any matrix A , Ae_i is the i -th column of A .
 (2) Let U be a vector space, let $(u_i)_i$ be a basis for U and let T and S be two linear transformations from U to another vector space. If $T(u_i) = S(u_i)$ for all i , then $T = S$.
 (3) Any linear transformation from \mathbb{k}^n to \mathbb{k}^m is associated to a unique $m \times n$ matrix.

Proof. (1) Directly from the definition
 (2) Let u be any vector. Write u as $a_1u_1 + \dots + a_nu_n$, and apply T and S .
 (3) Let T be such a transformation, and let A be the matrix with columns Te_i . It follows from the previous items that the transformation associated to A is equal to T , and that A is unique. □

Example 45 (Example 2.5-7). Find the linear transformation T satisfying $T(1, 1) = (2, -3)$ and $T(1, -2) = (5, 1)$.

Theorem 44 can help when verifying matrix identities. As an example, we prove the associativity of matrix multiplication (part 2 of theorem 33): we first prove it in the case $C = e_i$. In this case, according to the first part of theorem 44, we should prove that the i -th column of AB is $AB_{-,i}$. We already observed that this is true. We next assume that C is a column vector. In this case, it can be written as a linear combination of the e_i , and the statement follows from the other parts of the theorem. Finally, if C is arbitrary, it is enough, according to theorem 44, to prove $(A(BC))e_i = ((AB)C)e_i$ for all i . This reduces to the special cases already proved.

The associativity and the last theorem also allow us to re-interpret the product of matrices. Let A and B be two matrices such that AB is defined, and let T_A and T_B be the corresponding transformations. Recall that the composition of the two maps $T_A \circ T_B$ is defined by $(T_A \circ T_B)(v) = T_A(T_B(v))$. We claim that $T_A \circ T_B = T_{AB}$, the map associated with AB . This holds since, as we proved, $A(Bv) = (AB)v$. Therefore, matrix multiplication is nothing but composition of functions.

We may also re-interpret systems of linear equations. Recall that such a system can be written as $A\bar{x} = \bar{b}$ for some matrix A and tuple \bar{b} . Viewing A as a map, we see that the set of solutions is the set of vectors mapped to b .

5.2. Examples.

Example 46. Let $T : \mathbb{k}^3 \rightarrow \mathbb{k}^2$ be a linear transformation with $T(1, 2, 1) = (1, -2)$, $T(1, -1, 0) = (0, 1)$ and $T(0, 2, -1) = (-1, -2)$. Compute $T(5, 10, 15)$, and find the matrix of this transformation.

We have seen that the set of solutions of a homogeneous system $A\bar{x} = 0$ forms a vector space, and that the reduced row echelon form gives rise to a basis for this space. If \bar{x}_0 is a solution of a (general) system $A\bar{x} = \bar{b}$, and \bar{x}_1 is another solution of the same system, then $\bar{x}_1 - \bar{x}_0$ solves the *associated homogeneous system* $Ax = 0$. Therefore, any solution of the original system can be written as $x_0 + x$, where x is a solution of the homogeneous one. If y_1, \dots, y_n is a basis for the space of solutions of the homogeneous system, then the general form of a solution for the original equation is $x_0 + a_1y_1 + \dots + a_ny_n$.

Example 47. Write the general form of the solution for the following equations:

$$\begin{aligned} 3x + 4y - z + 3w &= 2 \\ x + z - 3w &= 2 \\ x + 4y - 3z + 9w &= -2 \end{aligned}$$

In the plane, linear transformations can be described geometrically:

Example 48. Show that the following are linear transformations, and find their matrices:

- (1) Reflection in any of the axes
- (2) Scaling in the direction of any axis
- (3) Projection on any axis
- (4) X -shear
- (5) Rotation by any angle θ
- (6) Reflection, scaling and projection on any line through the origin

6. INVERTIBLE MATRICES AND TRANSFORMATIONS

Let $T : X \rightarrow Y$ be a function. A function $S : Y \rightarrow X$ is called an *inverse* of T if $S \circ T = id_X$ and $T \circ S = id_Y$. T is called *invertible* if it has an inverse. The function T is *injective* (or *one-to-one*) if $T(x) \neq T(y)$ for $x \neq y$. It is *surjective* (or *onto*) if for any $y \in Y$, $T(x) = y$ for some $x \in X$.

Theorem 49. *Let $T : X \rightarrow Y$ be a function*

- (1) *The inverse of T , if exists, is unique*
- (2) *T is invertible if and only if it is one to one and onto.*
- (3) *If X and Y are vector spaces, and T is a linear transformation and is invertible, then the inverse of T is also a linear transformation*

A corollary of the last two items is that a linear transformation has an inverse linear transformation if and only if it is one to one and onto. If T is invertible, the unique (linear) inverse is denoted T^{-1} .

Proof. (1) Assume that S_1 and S_2 are both inverses of T . We get

$$S_1 = S_1 \circ I = S_1 \circ (T \circ S_2) = (S_1 \circ T) \circ S_2 = I \circ S_2 = S_2$$

(Note that we showed more: if T has both a left and a right inverse then they are equal, and T is invertible)

- (2) If T is injective and surjective, let S be defined as follows: $S(y) = x$ if $T(x) = y$. x exists since T is surjective, and is unique since T is injective. Clearly S is the inverse of T . The other direction is easy.
- (3) Let $y_1, y_2 \in Y$. By the previous part, $y_i = T(x_i)$ for some (unique) $x_i \in X$. Hence

$$\begin{aligned} S(y_1 + y_2) &= S(T(x_1) + T(x_2)) = S(T(x_1 + x_2)) = \\ &= x_1 + x_2 = S(T(x_1)) + S(T(x_2)) \end{aligned}$$

Multiplication by scalar is similar. □

The following additional properties of the inverse are easy to verify: The identity function is invertible, if T and S are invertible then so is $T \circ S$ (assuming it makes sense), and $(T \circ S)^{-1} = S^{-1} \circ T^{-1}$, if T is invertible, so is T^{-1} , and $(T^{-1})^{-1} = T$.

An invertible linear map is also called a *linear isomorphism*. If there is a linear isomorphism between two vector spaces, they are said to be *isomorphic*.

Example 50. Rotations, reflections and scaling by non-zero scalar are all invertible. Projections are not invertible. The transformation given by the matrix $\begin{bmatrix} 0 & 0 \\ 1 & 3 \end{bmatrix}$ is not invertible (example 2.3-2).

Example 51. The linear map that assigns to each geometric vector in the plane its pair of coordinates is an isomorphism. Hence, the plane is isomorphic to \mathbb{k}^2 .

Example 52. The map that assigns to each $m \times n$ matrix A the linear transformation T_A is a linear map from the space of $m \times n$ matrices to the space of linear transformations from \mathbb{k}^n to \mathbb{k}^m . It is an isomorphism since any linear transformation comes from a unique matrix.

It is easy to see that an isomorphism maps a basis to a basis: if $T : U \rightarrow V$ is a linear isomorphism, and (u_i) is a basis for U , then $(T(u_i))$ is a basis for V . This is used in the following example:

Example 53. Let L be a homogeneous system of linear equations, with solution space V . We saw that the reduced row echelon form gives rise to a linear map from \mathbb{k}^m to V . This map is an isomorphism: the projection that maps each solution to the values of the parameter variables is an inverse. In particular (as we saw), a basis for the set of solutions is obtained from the standard basis of \mathbb{k}^m .

6.1. Inverse matrices. Let A be a square matrix, and consider the system of equations $A\bar{x} = \bar{b}$. We have seen that the set of solutions of the system can be interpreted as follows: it is the set of all vectors mapped by T_A to \bar{b} . The statement that T_A is surjective means that the system has a solution for any tuple \bar{b} . The statement that T_A is injective means that the solution (if exists) is unique. Thus, T_A is invertible if and only if any system $A\bar{x} = \bar{b}$ has a unique solution. In this case, applying $(T_A)^{-1}$ to both sides, we get that the solution is $\bar{x} = (T_A)^{-1}\bar{b}$. Since $(T_A)^{-1}$ is again a linear transformation, there is some matrix A^{-1} representing it, i.e., $(T_A)^{-1} = T_{A^{-1}}$. The above discussion shows that if we can calculate A^{-1} , we can find the unique solution to any equation $Ax = b$.

In general, assume that A^{-1} exists. As with any linear map, we can compute its columns by applying it to the standard basis. Suppose that $A^{-1}e_i = x_i$. Then $Ax_i = e_i$. Hence, to find the columns of A^{-1} , we need to solve all the systems $Ax = e_i$. We may do this using elementary row operations. The assumption that A is invertible means that the reduced row echelon form has the form $Ix = x_i$, where I is the identity matrix, and x_i is the solution. Since the row operation depend only on A , and not on the right hand side, we may perform them simultaneously for all e_i . Thus we will perform the row operations on the system $AX = I$, and will obtain the system $X = A^{-1}$. If the row operations do not yield the identity matrix, the inverse does not exist.

Example 54 (Example 2.3-6). Find the inverse of

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

Recall that T is injective if $T(x) = b$ has at most one solution for all b . In particular, the only solution of $T(x) = 0$ is 0. In fact, the converse is true: if the

only solution of $T(x) = 0$ is 0, then T is injective. Indeed, if x_1 and x_2 are two different solutions of $T(x) = b$, then $T(x_1 - x_2) = 0$, so $x_1 - x_2$ is a non-zero solution of $T(x) = 0$. Thus, to check that T is injective, it is enough to check that $T(x) = 0$ has only the 0 solution. If $T : \mathbb{k}^n \rightarrow \mathbb{k}^n$, more is true: if T is injective, then T is an isomorphism. This follows again from considering the reduced row echelon form of the corresponding matrix. If T is injective, this is given by the identity matrix, and so T is invertible. Thus, $T(x) = 0$ has only the 0 solution if and only if T is invertible. Similarly, to verify that a transformation from \mathbb{k}^n to itself (or a square matrix) has an inverse, it is enough to check that it has a one-sided inverse.

6.2. Elementary matrices. Each of the elementary row operations is a linear transformation, and hence has a corresponding matrix. Since each of the operations is invertible, so is the corresponding matrix. Thus, if A is an $m \times n$ matrix, performing an elementary operation on A amounts to multiplying A on the left by some invertible $m \times m$ matrix E . How to find E ? We have $EA = (EI)A$, so $E = EI$ is the result of performing the corresponding operation on the identity matrix.

Therefore, the process of going from a matrix to an equivalent one via row operation can be described as performing a sequence of multiplication by elementary matrices: $A \mapsto E_k \dots E_2 E_1 A$. Since we already know that we can get from any matrix to any equivalent one via row operations, we see that A and B are equivalent if and only if there is a sequence E_i of elementary matrices, such that $A = UB$, where $U = E_k \dots E_2 E_1$. U can be computed by performing row operations on the identity matrix, as before. In particular, A is invertible if and only if it is equivalent to the identity matrix. Thus we get: A square matrix is invertible if and only if it is a product of elementary ones.

7. DETERMINANTS

We continue looking for methods to for discovering whether a matrix A is invertible, and to compute the inverse.

Consider the case of 2×2 matrices. If A is a non-zero 2×2 matrix, it will either map the plane bijectively to itself, or to a line. The map is an isomorphism if and only if the image is a plane. This happens if and only if the vectors $A(1, 0)$ and $A(0, 1)$ are not on the same line. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $Ae_1 = (a, c)$ and $Ae_2 = (b, d)$. The condition that these two vectors are on the same line is $ad = bc$ or $ad - bc = 0$. The expression $ad - bc$ is called the *determinant*, $\det(A)$ of A . How to find the inverse of A ? The *adjugate* of a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is defined to be $\text{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. A direct calculation shows that $A \text{adj}(A) = \text{adj}(A)A = \det(A)I$. Hence, in the case that A is invertible, A^{-1} is given by $\frac{1}{\det(A)} \text{adj}(A)$.

Example 55 (Example 2.3-5). Solve the system

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

In general, we may imitate the above idea as follows. An $n \times n$ matrix A carries the standard basis e_i to the n -tuple of column vectors of A . A is invertible if and only if these vectors do not lie in a proper sub space of \mathbb{k}^n . We will try to pin down this condition by computing the (oriented) volume of the polytope determined these vectors. Thus we will consider the function $A \mapsto \det(A)$, sending a square

matrix to the volume of the polytope determined by the columns. This volume is 0 precisely if the matrix is not invertible. Being a volume function, it has the following properties:

- Axiom 56** (Properties of determinants). (1) $\det(I) = 1$
 (2) Let B be the matrix obtained from A by multiplying some column by a number u . Then $\det(B) = u \det(A)$.
 (3) Let A_1 and A_2 be two matrices with identical columns, except column i , and let B be the matrix with the same columns as A_1 and A_2 , except column i which is the sum of the corresponding columns. Then $\det(A_1) + \det(A_2) = \det(B)$ (see Figure 1 for an illustration in the two dimensional case.)
 (4) If A has two identical columns, then $\det(A)$ is equal to 0.

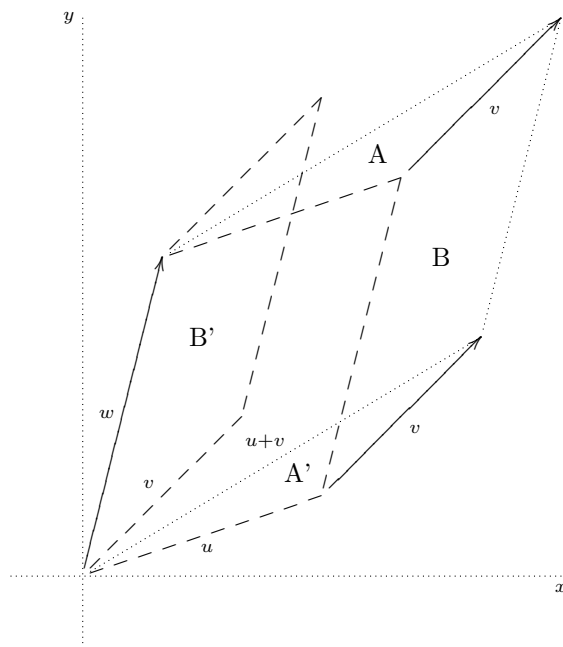


FIGURE 1. The area $\det(u + v, w)$ of the parallelogram with the dotted line is equal to the area of the shape with the triangle A translated to A' . This shape consists of the parallelogram B , and the one given by u and w . B , it turn, is a translate of B' , which is the parallelogram determined by v and w . Thus $\det(u + v, w) = \det(u, w) + \det(v, w)$.

It follows from these properties that $\det(A) = 0$ if and only if the columns of A all lie in a proper subspace. It also turns out that there is only one function \det that satisfies the above properties. Therefore, to find it, it is enough to find *some* function on the set of matrices that has these properties.

Such a function is defined recursively as follows: if $n = 1$, \det is the identity function. Assuming \det is defined for matrices of size $n \times n$, we set, for A of size

$n + 1 \times n + 1$: $\det(A) = \sum_{i=1}^{n+1} a_{i,1} c_{i,1}(A)$, where the *cofactor* $c_{i,j}(A)$ is defined as $(-1)^{i+j} \det(A_{i,j})$, with $A_{i,j}$ the (i,j) -th minor of A : it is the matrix obtained from A by erasing row i and column j . This description is deduced directly from the axioms above: The second and the third axioms show that the function is determined by its values on the basis elements. The fourth axiom implies we may assume all columns are different basis elements, and determines what happens when switching the order. Thus we are reduced to the value on the identity matrix, which is determined by the first axiom.

The axioms also show that we may expand along any column: for any i ,

$$\det(A) = \sum_{j=1}^n a_{j,i} c_{j,i}(A)$$

Example 57 (Example 3.1-3-5). Compute the determinants of

$$\begin{bmatrix} 3 & 4 & 5 \\ 1 & 7 & 2 \\ 9 & 8 & -6 \end{bmatrix}$$

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

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

Example 58. A matrix is called *upper triangular* if all entries below the main diagonal are zero. The determinant of a triangular matrix is equal to the product of the entries on the main diagonal.

7.1. Properties of determinants. The following properties follow directly from the definition of the determinant:

Claim 59. *Let A be a matrix.*

- (1) *If B is obtained from A by adding a multiple of one column to another, then $\det(B) = \det(A)$.*
- (2) *If B is obtained from A by exchanging two columns, then $\det(B) = -\det(A)$.*
- (3) *If A has a zero column, then $\det(A) = 0$.*

Proof. Let C_1, \dots, C_n be the columns of A . We assume for definiteness that the two columns in question are C_1 and C_3 .

(1)

$$\begin{aligned} \det(C_1 + aC_3, C_2, C_3, \dots, C_n) &= && \text{by 56.3} \\ \det(C_1, C_2, C_3, \dots, C_n) + \det(aC_3, C_2, C_3, \dots, C_n) &= && \text{by 56.2} \\ \det(A) + a \det(C_3, C_2, C_3, \dots, C_n) &= && \text{by 56.4} \\ &= && \det(A) \end{aligned}$$

- (2)
- $$\begin{aligned}
 0 &= \det(C_1 + C_3, C_2, C_1 + C_3, \dots, C_n) && \text{by 56.4} \\
 &= \det(C_1, C_2, C_1, \dots, C_n) + \det(C_1, C_2, C_3, \dots) + \\
 &\quad \det(C_3, C_2, C_1, \dots) + \det(C_3, C_2, C_3, \dots) && \text{by 56.3} \\
 &= \det(C_1, C_2, C_3, \dots) + \det(C_3, C_2, C_1, \dots) && \text{by 56.4}
 \end{aligned}$$
- (3) A zero column is equal to itself multiplied by 0. Hence $\det(A) = 0 \det(A) = 0$.

□

It follows that performing column operations on A (i.e., row operations on A^T), results in multiplying the determinant by a non-zero number. In particular, if B^T is the reduced row echelon form of A^T , then $\det(B) = c \det(A)$, where $c \neq 0$. On the other hand $\det(B) = 1$ if B is the identity matrix, and B has a zero column otherwise, in which case $\det(B) = 0$. But B is the identity if and only if B^T is the identity, if and only if A^T is invertible, if and only if A is invertible (since $(A^T)^{-1} = (A^{-1})^T$.) We proved:

Theorem 60. $\det(A) \neq 0$ if and only if A is invertible.

Using this we may prove:

Theorem 61. For any matrices A and B , $\det(AB) = \det(A) \det(B)$

Proof. If B is not invertible, then it is not injective, hence so is AB . Therefore, $\det(AB) = \det(B) = 0$ and the statement is true. We thus may assume that B is invertible.

In this case, B can be written as product of elementary matrices. Hence, by induction, it is enough to prove the theorem when B is an elementary matrix. Multiplying by an elementary matrix on the right corresponds to performing an elementary operation on the column. The effect of each such operation on the determinant is described explicitly in claim 59. □

Remark 62. The above proof uses a calculation with elementary matrices which slightly obscures the meaning. A more conceptual (though, at the moment, less precise) argument is as follows: We defined the determinant of T to be the volume of the parallelogram determined by $T(e_1), \dots, T(e_n)$. However, there is nothing special about e_1, \dots, e_n . Given any n vectors v_1, \dots, v_n , denote by $v_1 \wedge \dots \wedge v_n$ the volume of the parallelogram determined by these vectors. Then, it follows from the linearity of the volume that for any vectors v_1, \dots, v_n , $T(v_1) \wedge \dots \wedge T(v_n) = \det(T)v_1 \wedge \dots \wedge v_n$. Once this is understood, the product formula in the theorem above follows immediately.

The following claim together with the product formula allow us to compute determinants in terms of other determinants:

Claim 63.

- (1) If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$
- (2) For any matrix A , $\det(A^T) = \det(A)$

Proof. (1) $\det(A^{-1}) \det(A) = \det(A^{-1}A) = \det(I) = 1$

- (2) We recall that A is invertible if and only if A^T is. Hence $\det(A) = 0$ if and only if $\det(A^T) = 0$. If both are invertible, using the product formula this reduces to elementary matrices. \square

The second part implies that we may compute the determinant by expanding along a row, rather than a column.

Example 64. If $\det(A) = 3$ and $\det(B) = -2$, compute $\det(B^2(A^{-1T}B^{-1}A)^{-1}A^T)$

Example 65 (Example 3.2-3). Show that if a product of matrices is invertible, then so is each matrix in the product.

7.2. Adjugates. Like in the two dimensional case, for any square matrix A we define the *adjugate* $\text{adj}(A)$ of A to be the transpose of the matrix $(c_{i,j}(A))$ of cofactors of A . From the formula for the determinant, we get that the (i, j) -th entry of $A \text{adj}(A)$ is equal to the determinant of the matrix obtained from A by replacing the j -th row by the i -th one. It follows that $A \text{adj}(A) = \det(A)I$, hence, if A is invertible, then $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$.

Example 66. Compute the adjugate of the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Example 67. Let A be a matrix with integer entries, such that $\det(A) = 1$. Show that A^{-1} also has integer entries.

Using the adjugate, we get a formula for the (unique) solution of the equation $Ax = b$, where A is invertible. From the formula for the inverse, we get $x = \frac{1}{\det(A)} \text{adj}(A)b$. Expanding this expression we get *Cramer's Rule*: $x_i = \frac{\det(A_i)}{\det(A)}$, where A_i is obtained from A by replacing column i with b . This rule allows us to calculate one x_i without calculating the others.

7.3. Application to polynomials. Using the tools we accumulated, we may prove a division with remainder theorem for polynomials. Recall that a *polynomial* is a function of the form $p(x) = a_n x^n + \cdots + a_1 x + a_0$. The highest index n for which a_n is non-zero is called the *degree* of the polynomial, denoted $\deg(p)$. A polynomial of degree n is called *monic* if $a_n = 1$. A *root* of p is a number a such that $p(a) = 0$.

Theorem 68. *Let $p(x)$ be a monic polynomial of degree n .*

- (1) *If $q(x)$ is another monic polynomial of degree $m \leq n$, then there are polynomials $r(x)$ of degree smaller than m and $t(x)$ monic of degree $n - m$, such that $p(x) = q(x)t(x) + r(x)$ (so $t(x)$ is $p(x)$ divided by $q(x)$, with remainder $r(x)$.) This presentation is unique.*
- (2) *If a is a root of p , then there is a monic polynomial $q(x)$ such that $p(x) = (x - a)q(x)$.*
- (3) *If q is a monic polynomial of degree at most n , and a_1, \dots, a_n are distinct numbers such that $p(a_i) = q(a_i) = 0$, then p and q are the same polynomial.*

Proof.

- (1) We write p_i, q_i, t_i and r_i for the coefficients of p, q, t and r , respectively. Thus we are given p_i and q_i , and we need to show that there exist unique t_i and r_i such that

$$x^n + \cdots + p_0 = (x^m + \cdots + q_0)(x^{n-m} + \cdots + t_0) + r_{m-1}x^{m-1} + \cdots + r_0 \quad (5)$$

$$= x^n + (t_{n-m-1} + q_{m-1})x^{n-1} + \cdots \quad (6)$$

$$+ (t_0 + q_{m-1}t_1 + \cdots + q_0t_m)x^m \quad (7)$$

$$+ (q_{m-1}t_0 + \cdots + q_0t_{m-1} + r_{m-1})x^{m-1} + \cdots \quad (8)$$

$$+ q_0t_0 + r_0 \quad (9)$$

Comparing coefficients, we get:

$$t_{n-m-1} + q_{m-1} = p_{n-1}$$

$$t_{n-m-2} + t_{n-m-1}q_{m-1} + q_{m-2} = p_{n-2}$$

...

$$\sum_{i=0}^m t_i q_{m-i} = p_m$$

$$\sum_{i=0}^{m-1} t_i q_{m-i-1} + r_{m-1} = p_{m-1}$$

...

$$q_0t_0 + r_0 = p_0$$

In this system of equations, the q_i and p_i are given, and the t_i and r_i are unknown. Thus it is a system of linear equations. Arranging the variable to have first the r_i and then the t_i , we get a matrix of the form

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & q_0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & q_1 & q_0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & q_{m-1} & q_{m-2} & q_{m-3} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & q_{m-1} & q_{m-2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

This an upper triangular matrix, with determinant 1. Therefore, there is a unique solution for any sets of coefficients p_i and q_i .

- (2) According to the first part, we can write $p(x) = (x - a)t(x) + r(x)$, with $\deg(r) < \deg(x - a) = 1$. Hence r is a constant r_0 . Setting $x = a$ on both sides, we get that $r_0 = 0$.
- (3) According to the previous part, $p(x) = (x - a_1)p_1(x)$. Since a_2 is a root of p and $a_2 \neq a_1$, a_2 is a root of $p_1(x)$. Hence $p_1(x) = (x - a_2)p_2(x)$. Continuing this way, we may write $p(x) = (x - a_1)(x - a_2) \cdots (x - a_n)$ (since the degree is n , there is nothing else left.) The same is true of $q(x)$, so the polynomials are equal.

□

Remark 69. Using example 67, we note that if all the coefficients involved are integers, then the polynomials t and r also have integer coefficients.

In applications, it is often the case that we are trying to find a function that passes through a given collection of points. In the following corollary, we find a minimal degree polynomial through any finite set of points in the plane.

Corollary 70. *Let a_1, \dots, a_n be numbers. Let*

$$A = \begin{bmatrix} 1 & a_1 & a_1^2 & \dots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \dots & a_2^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & a_n & a_n^2 & \dots & a_n^{n-1} \end{bmatrix}$$

Then $\det(A) = \prod_{i < j} (a_j - a_i)$

Proof. Both sides are monic polynomials of degree $n - 1$. Hence it is enough to check that they have $n - 1$ identical roots. The numbers a_1, \dots, a_{n-1} are roots of both sides. \square

Consider the equation $A\bar{x} = \bar{b}$. If $\bar{t} = (t_0, \dots, t_{n-1})$ is a solution, we get that $t_{n-1}a_i^{n-1} + \dots + t_0 = b_i$. Setting $p(x) = t_{n-1}x^{n-1} + t_{n-2}x^{n-2} + \dots + t_0$, we get that $p(x)$ is a polynomial with the property that $p(a_i) = b_i$ for each i . Thus we proved:

Corollary 71. *If a_1, \dots, a_n are distinct, then for any b_1, \dots, b_n there is a unique polynomial $p(x)$ of degree (at most) $n - 1$ with $p(a_i) = b_i$ for all i .*

This polynomial is the *interpolation polynomial* for the give points.

Example 72. For $n = 2$, we get the familiar statement that there is a line through any two points.

8. EIGENVALUES AND EIGENVECTORS

A *diagonal* matrix is a square matrix with all entries outside the main diagonal zero. Diagonal matrices are very simple: the transformation they define on the space operates separately in each coordinate, and it is easy to multiply them.

In some cases, a matrix is diagonal when considering a different basis. For example, consider the matrix $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$. Then $A(1, 1) = (2, 2)$ and $A(1, -1) = (-2, 2)$. Therefore, A acts by multiplication by 2 in one direction, and by -2 in another. Since $(1, 1)$ and $(1, -1)$ form a basis of the space of pairs, this behaviour determines the action of A on the whole space. It is “like” the diagonal matrix $\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$. In fact, consider the matrix $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. It maps $(1, 0)$ to $(1, 1)$ and $(0, 1)$ to $(1, -1)$. Hence $ABe_1 = (2, 2)$ and $ABe_2 = (-2, 2)$. Now, B^{-1} maps $(1, 1)$ to e_1 and $(1, -1)$ to e_2 . Hence $B^{-1}ABe_1 = 2e_1$ and $B^{-1}ABe_2 = -2e_2$. So $B^{-1}AB = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$. We call A diagonalisable:

definition 73. A square matrix A is *conjugate* (or *similar*) to another such matrix C if there is an invertible matrix B such that $A = B^{-1}CB$. A is *diagonalisable* if it is conjugate to a diagonal matrix.

It is easy to see that if A is conjugate to B , then B is conjugate to A , and if A is conjugate to B and B is conjugate to C , the A is conjugate to C .

Example 74. It is easy to compute A^n where A is a diagonal matrix: it is just the diagonal matrix whose entries are the n -th powers of the corresponding entries of A . It is equally easy to compute the A^n when A is diagonalisable: if $A = B^{-1}CB$, then $A^n = (B^{-1}CB)^n = B^{-1}C^nB$.

The question remains, how to find out whether A is diagonalisable, and if it is, how to find a matrix B such that $B^{-1}AB$ is diagonal? As in the example, a diagonalisable matrix is characterised by the existence of a basis v_1, \dots, v_n such that $Av_i = c_i v_i$ for some numbers c_i . In this case, the diagonal matrix conjugate to A has the c_i on the diagonal, and the conjugating matrix B has the v_i as its columns. We make the following definition:

definition 75. Let $T : U \rightarrow U$ be a linear transformation. If a number c and a non-zero element $u \in U$ have the property that $T(u) = cu$, then c is called an *eigenvalue* and u an *eigenvector* of T .

So the diagonal form of a diagonalisable matrix has its eigenvalues on the diagonal, and the diagonalising matrix has eigenvectors as its columns.

How do we find the eigenvalues and eigenvectors? We want to find a non-zero solution u to the equation $Au = cu$, for some number c . This is the same as solving $(cI - A)u = 0$. This system has a non-trivial solution precisely when $(cI - A)$ is not invertible, which happens precisely if $\det(cI - A) = 0$. We proved:

Corollary 76. *The eigenvalues of a matrix A are the roots of the characteristic polynomial $\det(xI - A)$ of A .*

Thus to find the eigenvalues of A we should compute the roots of the polynomial $\det(xI - A)$. Once we found them, to find the eigenvectors corresponding to a root a , we solve the system $(aI - A)x = 0$. Gaussian elimination provides a basis for the set of solutions of this system. A is diagonalisable if the whole space has a basis of eigenvectors. The diagonalising matrix then has the basis of eigenvectors as its columns. In particular:

Proposition 77. *If the characteristic polynomial of a matrix A of size n^2 has n distinct roots, then A is diagonalisable.*

Proof. Let B be the matrix whose columns are eigenvectors v_1, \dots, v_n , corresponding to the distinct roots (eigenvalues) a_1, \dots, a_n . We need to show that B is invertible. Assume that there is some vector $x = (x_1, \dots, x_n)$, such that $Bx = 0$. We may write $x = x_1 e_1 + \dots + x_n e_n$. Hence

$$\begin{aligned} 0 = Bx &= B(x_1 e_1 + \dots + x_n e_n) = Bx_1 e_1 + \dots + Bx_n e_n \\ &= x_1 B e_1 + \dots + x_n B e_n = x_1 v_1 + \dots + x_n v_n \end{aligned}$$

Thus we need to prove that if $x_1 v_1 + \dots + x_n v_n = 0$ then $x_1 = \dots = x_n = 0$. Assume this is not the case, and let x_1, \dots, x_k be a minimal sequence of non-zero numbers, such that $x_1 v_{i_1} + x_2 v_{i_2} + \dots + x_k v_{i_k} = 0$ for some i_1, \dots, i_k . Since the v_i are non-zero, $k > 1$. Applying A to both sides, we get $0 = Ax_1 v_{i_1} + \dots + Ax_k v_{i_k} = x_1 a_{i_1} v_{i_1} + \dots + x_k a_{i_k} v_{i_k}$. Since all the a_i are distinct, there is some a_{i_j} which is non-zero. Assume for simplicity that a_{i_1} is non-zero. Dividing by it, and then subtracting the original equation, we get an equation $y_2 v_{i_2} + \dots + y_k v_{i_k} = 0$. The y_i are not all zero, since the a_i are distinct. This is a contradiction to the minimality. \square

However, A may be diagonalisable even if the polynomial has less roots. In any case, if A is diagonalisable, then any value that appears on the diagonal is an eigenvalue, hence a root of the polynomial.

Example 78 (Example 3.3-4). Diagonalise the matrix

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

Example 79 (Example 3.3-8). Show that $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is not diagonalisable.

Example 80. Is the matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ diagonalisable?

Example 81 (Example 3.3-7). Diagonalise the matrix

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

Example 82. The Fibonacci sequence is given by the rule $a_0 = a_1 = 1$, and for $n > 0$, $a_{n+1} = a_n + a_{n-1}$. Find a closed formula for the n -th term a_n .

9. LINEAR MAPS IN 3-SPACE

The following transformations of 3-space are linear:

- (1) Reflection or projection on a line or a plane through the origin
- (2) Rotations about a line through the origin

This can be checked either geometrically, or using coordinates, from the formulas for these operations. Once this is known we may compute the matrix for these transformation in the usual way.

Example 83. Let v be a vector in 3-space lying on the xy -plane, with angle α between it and the x -axis. Compute the matrix of the rotation about v in angle θ (counterclockwise when looking in the direction of v .)

As mentioned in remark 62, a linear transformation A maps a parallelepiped of volume V to one of volume $\det(A)V$.

10. LINEAR INDEPENDENCE, BASES AND DIMENSION

Recall that a *linear combination* of vectors v_1, \dots, v_n is an expression of the form $a_1v_1 + \dots + a_nv_n$ (where the a_i are numbers.) Recall also that a basis for a vector space U is a set B of vectors in U such that any vector in U can be written uniquely as a linear combination of vectors in B (where all the coefficients are non-zero.) This definition includes two conditions: first, any vector is a linear combination of vectors in B , and second, such a linear combination is unique. Since 0 can always be written with the empty linear combination, it implies that if $a_1v_1 + \dots + a_nv_n = 0$ for $v_i \in B$, then $a_1 = \dots = a_n = 0$. Conversely, if this last condition is satisfied, then the uniqueness is satisfied in general, since if $v = a_1v_1 + \dots + a_nv_n = b_1u_1 + \dots + b_ku_k$ for $v_i, u_j \in B$ are different expressions for v , then subtracting them gives a non-trivial expression for 0. All this motivates the following definition:

definition 84. Let U be a vector space, and B a subset of U .

- (1) The *span* of B is the set of linear combinations of elements of B .
- (2) The set B is *linearly independent* if 0 is not a non-trivial linear combination of elements of B .

Thus, a basis for U is a linearly independent set whose span is the whole of U .

Example 85 (Example 5.1-4). For $P = (0, -11, 8, 1)$ and $Q = (2, 3, 1, 2)$, determine if they are in the span of $(2, -1, 2, 1)$ and $(3, 4, -1, 1)$.

The span can be characterised as follows:

Proposition 86. *Let B be a subset of a vector space U . Then the span of B is a sub-space. It is the smallest subspace of U containing B .*

Proof. The first statement is easy, we prove the second. Let V be a subspace of U containing B . By definition, the span of B as a subset of U and as a subset of V are the same. Therefore, the span of B is contained in V . \square

In particular, B is a basis if it is linearly independent, and if there is no proper subspace containing B .

Example 87 (Example 5.1-5). If X and Y are two vectors, show that $X + Y$ and $X - Y$ span the same space as X and Y . If X and Y are independent, then so are $X + Y$ and $X - Y$.

10.1. Relation with linear maps. Let $T : U \rightarrow V$ be a linear transformation. The *null space* (or the *kernel*) of T is defined to be the set of vectors $u \in U$, such that $T(u) = 0$. The *image* of T is the set of vectors $v \in V$ of the form $v = T(u)$ for some $u \in U$. It is easy to see that both the kernel and the image of T are subspaces (of U and of V , respectively.)

Claim 88. *Let $T : U \rightarrow V$ be a linear map, and let $B \subseteq U$. We denote by $T(B)$ the set of elements of the form $T(u)$, where $u \in B$.*

- (1) *If B spans U , then $T(B)$ spans the image of T*
- (2) *If T is injective, and B is linearly independent, then $T(B)$ is also linearly independent.*
- (3) *If T is invertible, and B is a basis of U , then $T(B)$ is a basis of V .*

Proof. (1) Let v be in the image of T . Then there is some $u \in U$ such that $T(u) = v$. Since B spans U , we have $u = a_1u_1 + \dots + a_nu_n$ for some numbers a_i and vectors $v_i \in B$. Hence $v = T(u) = a_1T(u_1) + \dots + a_nT(u_n)$.
 (2) Assume that $a_1v_1 + \dots + a_nv_n = 0$ for some $v_i \in T(B)$. Since T is injective, there are unique $u_i \in B$ such that $T(u_i) = v_i$. Then $0 = a_1T(u_1) + \dots + a_nT(u_n) = T(a_1u_1 + \dots + a_nu_n)$. Since T is injective, $a_1u_1 + \dots + a_nu_n = 0$, hence, since B is linearly independent, $a_i = 0$ for all i .
 (3) Since B spans U and T is surjective, $T(B)$ spans V . Since B is linearly independent and T is injective, $T(B)$ is also linearly independent. \square

If $U = \mathbb{k}^m$ and $V = \mathbb{k}^n$, then T is represented by a unique matrix A . In these terms, the kernel of A is the set of solutions to the homogeneous system $A\bar{x} = 0$, and we already know that Gaussian elimination produces a basis for this set. By definition, the kernel is 0 if and only if the columns of A are linearly independent. Since the standard basis e_1, \dots, e_m spans U , we get from claim 88 that Ae_1, \dots, Ae_m spans the image. These are just the columns of A . We proved:

Corollary 89. *Let A be an $m \times n$ matrix, and let T_A be the corresponding linear map. Then T_A is injective if and only if the columns of A are linearly independent, and it is surjective if and only if the columns span \mathbb{k}^n .*

If A is a square matrix, then it is injective if and only if it is surjective (if and only if it is invertible.) Hence, the columns of A span \mathbb{k}^n if and only if they are linearly independent. In other words, the columns of A form a basis for \mathbb{k}^n if and only if A is invertible. Since A^T is invertible if and only if A is, we may exchange columns and rows throughout.

10.2. Dimension. We next want to show that the size of any two bases for a vector space U is the same. We will do this only for *finite dimensional* vector spaces.

definition 90. A vector space U is *finite dimensional* if it is spanned by a finite subset.

The terminology will be explained soon. Assume that B is a finite set of vectors spanning U . If B is linearly dependent, there is some vector $v \in B$ that is in the span of the other vectors in B . Therefore, if we remove it from B , the resulting set B_1 will still span U . Continuing this way, we get (since B is finite) a basis for U . Thus we prove:

Claim 91. *A finite dimensional space has a finite basis*

Let U be a vector space with basis v_1, \dots, v_n , and consider the map $T : \mathbb{k}^n \rightarrow U$ given by $T(x_1, \dots, x_n) = x_1v_1 + \dots + x_nv_n$. By definition, this is a linear map. We claim that it is invertible: indeed, if $T(x_1, \dots, x_n) = 0$, then $x_1v_1 + \dots + x_nv_n = 0$. Since the v_i form a basis, they are linearly independent, so $x_i = 0$ for all i . This proves that T is injective. On the other hand, $T(\{e_i\})$ span U , so by 88, T is surjective. Combining this with the last claim, we get:

Proposition 92. *If U is a finite dimensional space, then there is an invertible map from U to some \mathbb{k}^n .*

We are now ready to show that any two bases of a finite dimensional space have the same size.

Theorem 93. *The size of any basis for \mathbb{k}^n is n . All bases of a finite dimensional space have the same size.*

Proof. The columns of any basis form an invertible matrix. However, only square matrices are invertible. To prove the second claim, let U be finite dimensional, and let $T : \mathbb{k}^n \rightarrow U$ be an invertible linear map, as provided by proposition 92. Any basis of U is mapped to a basis of \mathbb{k}^n (by 88), whose size we just proved to be n . \square

The theorem allows us to make the following definition:

definition 94. Let U be a finite dimensional space. The *dimension* of U is the size of a basis for U .

Thus, a space is finite dimensional if it has a finite dimension.

These concept are of no much use unless we know that a space is finite dimensional:

Proposition 95. *Let U be a finite dimensional space, and let V be a subspace. Then V is also finite-dimensional, and $\dim(V) \leq \dim(U)$. If $\dim(V) = \dim(U)$, then $V = U$.*

Proof. Since U is finite-dimensional, there is an invertible map T from U to \mathbb{k}^n . The image $T(V)$ of V under T is a subspace of \mathbb{k}^n . If we find an invertible map S from $T(V)$ to some \mathbb{k}^m , then $S \circ T$ is an invertible map from V to \mathbb{k}^m . Furthermore, an invertible map maps bases to bases. Thus it is enough to prove the claim for $U = \mathbb{k}^n$.

We prove it by induction on n . Assume we proved it for all \mathbb{k}^i where $i < n$. If $V = U$ there is nothing to prove. Otherwise, there is at least one vector e_i in the standard basis that is not in V . Consider the (linear) map π_i from \mathbb{k}^n to \mathbb{k}^{n-1} that “forgets” the i -th coordinate. The fact that $e_i \notin V$ means that the restriction of π_i to V is injective. Therefore, π_i is an invertible map from V to $\pi(V) \subseteq \mathbb{k}^{n-1}$. By induction, $\pi(V)$ is finite-dimensional, of dimension at most $n - 1$. Hence so is V . \square

Remark 96. If V is the kernel of a matrix A , then Gaussian elimination can be viewed as the process of finding explicitly an invertible map from V to some \mathbb{k}^m .

Remark 97. We saw that any spanning set of a finite-dimensional space contains a basis. We can now show that any linearly-independent set B is contained in a basis. In fact, if there is a vector not in the span of B , then adding the vector to B will result in an independent set. Continuing the process, we are bound to stop when the size of B reaches the dimension (if we can do this one more time, the span of B is a subspace of dimension greater than the space containing it, contradicting the last proposition.) At this point, B is a basis.

11. SCALAR PRODUCT AND ORTHOGONALITY

We will consider an additional structure on a vector space U , a scalar product. The scalar product is a function from pairs of vectors in U to numbers, $(u, v) \mapsto u \cdot v$, that satisfies the following conditions:

- (1) It is symmetric: $u \cdot v = v \cdot u$
- (2) It is linear in each factor: $u \cdot (av_1 + bv_2) = au \cdot v_1 + bu \cdot v_2$.
- (3) If $v \cdot v = 0$ then $v = 0$.

The main example of such an operation is given in \mathbb{Q}^n (\mathbb{Q} is the set of rational numbers) by the formula $(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = x_1y_1 + \dots + x_ny_n$. Though this is the only scalar product we will use, the properties above are usually more useful than the formula.

Recall that in the 2 or 3 dimensional case, the scalar product has a geometric meaning. In particular, the square root of $v \cdot v$ is the length $\|v\|$ of v . And two vectors u, v are orthogonal if they are non-zero, and $u \cdot v = 0$. In the general context, we *define* $\|v\|^2 = v \cdot v$ and u, v are *orthogonal* if $u \cdot v = 0$. A set B of vectors is orthogonal if any two vectors in it are orthogonal. A set is *orthonormal* if it is orthogonal, and the (square of the) length of each vector in it is 1. If we are allowed to extract square roots, then any orthogonal set can be converted to an orthonormal one by dividing each vector v by $\|v\|$. For example, the standard basis is orthonormal.

Example 98 (Example 5.3-6). Show that the vectors $(1, 1, 1, -1)$, $(1, 0, 1, 2)$, $(-1, 0, 1, 0)$ and $(-1, 3, -1, 1)$ are orthogonal and normalise them.

The reason we are interested in orthogonal sets is:

Theorem 99. *Any orthogonal set is linearly independent. If v_1, \dots, v_n is an orthonormal basis, then for any v , $v = (v \cdot v_1)v_1 + \dots + (v \cdot v_n)v_n$.*

Proof. Assume that $a_1v_1 + \dots + a_nv_n = 0$, where the v_i are orthogonal. Taking scalar product with v_i on both sides, we get $a_i\|v_i\|^2 = 0$. Since $v_i \neq 0$, this forces $a_i = 0$. This proves that the v_i are linearly independent. If they form an orthonormal basis, write $v = a_1v_1 + \dots + a_nv_n$, and again form the scalar product with v_i . We get $a_i = v \cdot v_i$. \square

As mentioned above, the dot product conveys something about the geometry of the vectors involved. We saw that in the plane (or in 3-space), vectors u_1, \dots, u_n are orthogonal if and only if they are perpendicular. It follows from the Pythagorean theorem that $\|v_1 + \dots + v_n\|^2 = \|v_1\|^2 + \dots + \|v_n\|^2$. This holds for the dot product in general (expand the left side.) To explain further geometric properties, we mention another property of the dot product. This property only makes sense when the numbers we are using are ordered in such a way that $x^2 \geq 0$ for any number x (this is true for the rational or real numbers, but not for the complex numbers.) In this case the formula for the dot product shows directly that $\|x\|^2$ is a sum of squares, and hence non-negative. In this setting we can prove:

Theorem 100 (Cauchy inequality). *For any two vectors u and v , $(u \cdot v)^2 \leq \|u\|^2\|v\|^2$. Equality holds if and only if u and v are co-linear.*

Proof. We note that for any number a ,

$$0 \leq \|u - av\|^2 = (u - av) \cdot (u - av) = \|u\|^2 - 2a(u \cdot v) + a^2\|v\|^2$$

We may assume that $v \neq 0$. Then, setting $a = \frac{u \cdot v}{\|v\|^2}$, we get

$$0 \leq \|u\|^2 - 2\frac{(u \cdot v)^2}{\|v\|^2} + \frac{(u \cdot v)^2}{\|v\|^2} = \|u\|^2 - \frac{(u \cdot v)^2}{\|v\|^2}$$

Since $\|v\|^2 > 0$, we may multiply both sides by it, and get the result. Equality holds if and only if it holds in the first equation, if and only if $u = av$. \square

As a corollary, we get the triangle inequality, which states that the sum of lengths of two sides of a triangle is always at least the length of the third. In this statement we use an additional property of numbers: Every non-negative number has a non-negative square root (this holds for real numbers, but not for the rationals; it holds in the geometric context because of the Pythagorean theorem.) In this case we set $\|u\|$ to be the non-negative square root of $\|u\|^2$.

Corollary 101. *For any pair u, v of vectors, $\|u + v\|^2 \leq (\|u\| + \|v\|)^2$.*

Proof. By the Cauchy inequality,

$$\|u + v\|^2 = \|u\|^2 + 2u \cdot v + \|v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$$

\square

11.1. Scalar product and matrices. We note that $x \cdot y$ is nothing but the matrix product xy , where x is viewed as an $1 \times n$ matrix, and y as an $n \times 1$ one. Hence, if A is an $m \times n$ matrix, $x \in \mathbb{k}^m$ and $y \in \mathbb{k}^n$, then $x \cdot Ay = xA \cdot y$. As a vector, $xA = A^T x$. Hence we get $A^T x \cdot y = x \cdot Ay$. Setting $x = Ay$, we get $A^T Ay \cdot y = \|Ay\|^2$. Hence, if $A^T Ay = 0$, then $Ay = 0$. We proved that if A is injective, then so is $A^T A$. But $A^T A$ is a square matrix, hence is invertible. Conversely, of course, if $A^T A$ is invertible, then A is injective.

12. RANK OF A LINEAR MAP

Recall that we defined (in section 3.3) the rank of a matrix A to be the number of non-zero lines in the (unique) equivalent reduced row echelon matrix A' . We note that the rows of A' are linearly independent: if $a_1R_1 + \cdots + a_mR_m = 0$, where R_i are the rows of A' , then, looking at the coordinates containing the leading 1, we see that the a_i must be 0. Therefore, the rank is the dimension of the space spanned by the rows of A' . On the other hand, considering the columns of A' , we see that the columns containing the leading 1s form a basis for the space spanned by the columns of A' . In other words, it is the dimension of the image of A' . Now, $A = CA'$ for some invertible matrix C (so $\text{im}(A) = \text{im}(CA')$.) Since C is invertible, it preserves dimensions, so the dimension of $\text{im}(CA')$ is equal to the dimension of $\text{im}(A')$, which is the rank of A . Thus we get an intrinsic definition of the rank:

definition 102. If $T : U \rightarrow V$ is a linear map between finite dimensional vector spaces, the *rank* of T is the dimension of the image of T .

In the case of a matrix, the rank is thus the dimension of the *column space*, the space spanned by the columns. Going back to the discussion above, this is also the dimension of the space spanned by the rows of A' , which is the image of $(A')^T$. Again we have $A^T = (A')^T C^T$ for the above C , so the images of A^T and $(A')^T$ are equal. Thus we see that the dimension of the *row space*, the space spanned by the rows of A , is also equal to the rank. We proved:

Theorem 103. *The dimensions of the row space and the column space of any matrix A are equal (and both are equal to the rank of A .) The ranks of A and A^T are equal.*

In particular, if A is an $m \times n$ matrix, the rank of A is at most n and at most m . Also, multiplying by an invertible matrix (on either side) does not change the rank. If A is a square matrix, it is invertible if and only if it is surjective, if and only if the rank is n (the size of the matrix.)

If A is an $m \times n$ matrix of rank r , we saw that the reduced row echelon form gives rise to an invertible map from the kernel of A (the set of solutions of the corresponding system of linear equations) to some \mathbb{k}^l . What is l ? l is the number of columns in the reduced row echelon that do not contain a leading 1. Hence it is $n - r$. We thus proved:

Corollary 104. *If A is an $m \times n$ matrix, then $\dim(\text{Im}(A)) + \dim(\ker(A)) = n$.*

We summarise what we know in the following theorem:

Theorem 105. *Let A be an $m \times n$ matrix*

- (1) *The following are equivalent:*
 - (a) $\text{rank}(A) = n$
 - (b) *The rows of A span \mathbb{k}^n*
 - (c) *The columns of A are linearly independent*
 - (d) $CA = I_n$ for some $n \times m$ matrix C
 - (e) *A is injective*
- (2) *The following are equivalent:*
 - (a) $\text{rank}(A) = m$
 - (b) *The columns of A span \mathbb{k}^m*

- (c) *The rows of A are linearly independent*
- (d) *$AC = I_m$ for some $n \times m$ matrix C*
- (e) *A is surjective*

Proof. (1) The first two items are equivalent by definition, The third is by corollary 104. The fourth follows since there is some C' such that $C'A$ is in reduced row echelon form. This form has the shape I_n with some rows of zeroes added. C is then C' with the last $m - n$ rows removed. Finally the fifth follows immediately from the fourth, and is a restatement of the third.

- (2) We replace A by A^T in the first part. Everything follows automatically, except the last item, which is equivalent to the second. □

We note that we showed in section 11.1, that if we have a scalar product, then the first set of conditions is also equivalent to $A^T A$ being invertible. Taking the transpose again, we see that the second set is equivalent to AA^T being invertible.

Example 106 (Example 5.4-4). Show that $\begin{bmatrix} 3 & x+y+z \\ x+y+z & x^2+y^2+z^2 \end{bmatrix}$ is invertible if at least two of x, y, z are distinct.

13. MORE ON SIMILARITY

We recall that two square matrices A and B are similar (or conjugate) if there is an invertible matrix P such that $A = P^{-1}BP$. We note that if A and B are similar, then so are A^T and B^T and A^k and B^k (for integer k if A^{-1} exists.) In particular, if A is diagonalisable, then so are A^T and A^k .

Some quantities we may compute for matrices are equal for similar matrices. For example, we saw that if A and B are two matrices, then $\det(AB) = \det(A)\det(B)$. It follows that $\det(P^{-1}AP) = \frac{1}{\det(P)}\det(A)\det(P) = \det(A)$. Hence if A and B are similar, they have the same determinant. For the same reason they have the same characteristic polynomial, and so the same eigenvalues. They also have the same rank, since multiplication by an invertible matrix on either side preserves it.

A more surprising example is the trace: the *trace* of a matrix is defined to be the sum of entries on the main diagonal. We note that $A \mapsto \text{tr}(A)$ is a linear map from the space of $n \times n$ matrices to \mathbb{k} . It is not the case that $\text{tr}(AB) = \text{tr}(A)\text{tr}(B)$. However:

Proposition 107. *For any two square matrices A and B , $\text{tr}(AB) = \text{tr}(BA)$*

Proof. Since $\text{tr}((cA_1 + dA_2)B) = c\text{tr}(A_1B) + d\text{tr}(A_2B)$ (for any fixed B , the map $A \mapsto \text{tr}(AB)$ is linear), if we prove the statement for A_1 and A_2 it also holds for A . A similar remark applies to B . Thus it is enough to prove the statement when A and B are the elements of some basis. A basis for the space of matrices is given by the matrices $E_{i,j}$, having 1 on the (i, j) -th place, and 0 elsewhere. But $E_{i,j}E_{l,m} = E_{i,m}$ if $j = l$, and 0 otherwise. Hence $\text{tr}(E_{i,j}E_{j,i}) = 1$ for all i, j , and is 0 for all other products. This condition is symmetric, and so the equality holds. □

Hence $\text{tr}((P^{-1}A)P) = \text{tr}(PP^{-1}A) = \text{tr}(A)$. In fact, both the determinant and the trace are coefficients in the characteristic polynomial: the determinant is a_0 and the trace is a_{n-1} . We note that $\begin{bmatrix} 1 & \\ & 1 \end{bmatrix}$ and the identity have the same characteristic polynomial and the same rank, but are not conjugate.

We have seen that A is diagonalisable if and only if there are n eigenvectors such that the matrix P with these eigenvectors as columns is invertible (in which case P is the conjugating matrix.) In our current terminology we may simply say that the space has a basis of eigenvectors. We showed in proposition 77 that eigenvectors corresponding to distinct eigenvalues are distinct (hence if the characteristic polynomial has n distinct roots, then A is diagonalisable.)

Recall (theorem 68) that a is a root of a polynomial $p(x)$ if and only if we may write $p(x) = (x - a)q(x)$ for some polynomial. a may further be a root of $q(x)$. If $p(x) = (x - a)^m q(x)$ for some polynomial $q(x)$ such that a is not a root of q , we call m the *multiplicity* of the root m . For example, $(x - 3)^2(x - 1)$ has the root 1 with multiplicity 1 and the root 3 with multiplicity 2. The number of roots *with multiplicities* of $p(x)$ is the sum of multiplicities of all roots of p (thus, it is just the number of roots if the multiplicity of each root is 1.)

If α is a number, we let $E_\alpha(A) = \ker(\alpha I - A)$. Thus, if α is an eigenvalue, E_α is the space of eigenvectors for α , and otherwise $E_\alpha = 0$. Let $c_A(x)$ be the characteristic polynomial of A . We have

Theorem 108. *An $n \times n$ matrix A is diagonalisable if and only if $c_A(x)$ has n roots with multiplicities, and for any eigenvalue α , the dimension of E_α is equal to the multiplicity of α as a root of $c_A(x)$.*

Proof. Assume that A is similar to the diagonal matrix B . The claim is certainly true for B . The diagonalising matrix P maps eigenspaces to eigenspaces, and since it is invertible, it preserves dimension.

Conversely, if the dimensions are equal to the multiplicities, pick a basis of each E_α , and consider the set consisting of all of these vectors. This is a set of n vectors, so to prove it is a basis, we only need to prove they are independent, which we already did. \square

If A is not diagonalisable, it is still true that the dimension of E_α is at most the multiplicity of α . To prove this we first make the following general remark.

Suppose that we have a quantity defined on matrices that does not change when passing to a similar matrix (such as determinant.) Then we may define it on any a general map $T : V \rightarrow V$, where V is a finite dimensional space. Indeed, since V is finite dimensional, there is an invertible map $S : V \rightarrow \mathbb{k}^n$ for some n . Then $S \circ T \circ S^{-1}$ is a map from \mathbb{k}^n to itself, and hence is represented by a matrix A_S , and we define the determinant (or trace, etc.) to be the determinant of A_S . These definition appears to depend on the choice of the invertible map S . However, if $R : V \rightarrow \mathbb{k}^n$ is another invertible map, the matrix A_R we get satisfies $A_R = (RS^{-1})A_S(SR^{-1}) = (RS^{-1})A_S(RS^{-1})^{-1}$. Hence A_R is similar and has the same determinant (or trace, etc.) Thus the *characteristic polynomial* of T is by definition the characteristic polynomial of any such matrix A_S .

Now let $T : V \rightarrow V$ be a linear map on a finite-dimensional space, and let $U \subseteq V$ be a subspace, such that $T(U) \subseteq U$ (we say that U is *invariant*.) Then T restricted to U is a linear map from U to itself, which we denote by T_U . We may relate the characteristic polynomials of T and T_U :

Theorem 109. *Let $T : V \rightarrow V$ be a linear map on a finite-dimensional space, and let $U \subseteq V$ be a subspace, such that $T(U) \subseteq U$. Let p be the characteristic polynomial of T , and q the characteristic polynomial of T_U . Then q divides p : there is a polynomial r such that $p = qr$.*

Proof. We may, according to the remark above, choose an arbitrary invertible map $S : V \rightarrow \mathbb{k}^n$, and compute the characteristic polynomials using it. Let v_1, \dots, v_m be a basis of U . Since the v_i are linearly independent, we may choose vectors v_{m+1}, \dots, v_n that complete it to a basis of V . Let S be the map defined by this basis (i.e., $S(a_1v_1 + \dots + a_nv_n) = (a_1, \dots, a_n)$.) Let A be the matrix of $S \circ T \circ S^{-1}$. The space U is mapped to the sub space $\mathbb{k}^m \subseteq \mathbb{k}^n$. Hence the matrix A has the shape $\begin{bmatrix} A' & B \\ 0 & C \end{bmatrix}$, where A' is the restriction of A to \mathbb{k}^m . Hence $p(x) = \det(xI_n - A) = \det(xI_m - A') \det(xI_{n-m} - C) = q(x)r(x)$, where $r(x)$ is the characteristic polynomial of C . \square

Now it is easy to deduce the inequality mentioned above:

Corollary 110. *Let $T : V \rightarrow V$ be a linear map on a finite dimensional space. For any number α , the dimension of the eigenspace E_α is at most the multiplicity of α in the characteristic polynomial of T .*

Proof. The space E_α is invariant: If $v \in E_\alpha$, then $Tv = \alpha v$, hence $T(Tv) = T(\alpha v) = \alpha T(v)$. Hence we may apply the theorem, and get that the characteristic polynomial p of T is divisible by the characteristic polynomial of the restriction S of T to E_α . However, S is simply αI , where I is the identity on the space E_α . Hence the characteristic polynomial of S is $(x - \alpha)^m$, where m is the dimension of E_α . Hence $p(x) = (x - \alpha)^m r(x)$, and the statement is proved. \square

14. ORTHOGONALITY

We assume that every positive number has a square root, and that we are given a scalar product on \mathbb{k}^n . An orthogonal matrix is a matrix whose columns form an orthonormal basis. It follows that if P is orthogonal, then it is invertible, and $P^{-1} = P^T$. Matrices A and B are orthogonally similar if A and B are similar via an orthogonal matrix. A is *orthogonally diagonalisable* if it is orthogonally similar to a diagonal matrix. The following theorem characterises such matrices.

Theorem 111. *A matrix is orthogonally diagonalisable if and only if it is symmetric.*

Example 112 (Example 8.2-5). Orthogonally diagonalise

$$\begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$$

15. VECTOR SPACES

We recall the definition of a vector space: it is a set V , an operation $u, v \mapsto u + v$ (*addition*) on elements of V , another operation $a, v \mapsto a \cdot v$ (scalar multiplication), where a is a number and $v \in V$, and an element $0 \in V$ (*zero*.) These operations satisfy the following axiom for any elements u, v, w of V and numbers a, b :

- (1) $u + v = v + u$
- (2) $(u + v) + w = u + (v + w)$
- (3) $u + 0 = u$
- (4) There is an element $-u$ such that $-u + u = 0$
- (5) $a(u + v) = au + av$
- (6) $(a + b)u = au + bu$

$$(7) a(bv) = (ab)v$$

$$(8) 1u = u$$

Remark 113. It follows from the first and third axioms that there is only one 0. Hence it would amount to the same thing to require that there *is* an element 0 with $u + 0 = 0$ for all $u \in V$, rather than specifying it. Similarly, given u , $-u$ is unique (check!), so we could require a *function* $- : V \rightarrow V$ with the required property.

A *subspace* of a vector space V is a subset U closed under the operations of V . The restriction of the structure to U then equips U with a vector space structure.

So far we have mainly seen the examples of \mathbb{k}^n and its subspaces. In this section, we will consider some more diverse examples.

15.1. Matrix spaces. We have already mentioned that the set of $m \times n$ matrices is a vector space, viewed as a set of mn -tuples. It has a basis consisting of the matrices $E_{i,j}$, having 1 in the i,j -th place, and 0 elsewhere. Assume that $m = n$. The additional structure allows us to define several subspaces. To do this we recall that the kernel of any linear map is a subspace. The trace is a linear map, hence its kernel, the set of matrices of trace 0 is a subspace. As another example, the map T sending a matrix to its transpose, is linear. Since this map is invertible, its kernel is 0. However, the map $T - I$ (where I is the identity) is also linear, and its kernel is precisely the space of symmetric matrices. Similarly the space of matrices with $A^T = -A$ (anti-symmetric matrices) is the kernel of $T + I$.

Yet another example occurs as follows: Let A be a matrix. The map that takes a matrix X to $AX - XA$ is linear. The kernel of this map is thus a sub space. It is the space of all matrices that commute with A . More generally, if P is any set of matrices, the set C_P of matrices commuting with all the matrices in P is a subspace. In particular, if we take P to be the set of all matrices, we get a subspace called the centre (in fact, it is the set of scalar matrices cI .)

The map $A, B \mapsto \text{tr}(A^T B)$ is a scalar product on the set of $m \times n$ matrices (assuming, as usual, that a sum of squares of numbers is positive.) It is equal to $A, B \mapsto \text{tr}(AB^T)$. The standard basis is an orthonormal basis with this scalar product. The set of symmetric matrices is orthogonal to the set of anti-symmetric ones.

We finish this sub section by exhibiting some sets of matrices that are not subspaces: the set of invertible matrices, the set of diagonalisable matrices (check!), the set of orthogonal matrices.

Example 114 (Example 6.3-5). Let A be a matrix such that $A^k = 0$ but $A^{k-1} \neq 0$. Show that I, A, \dots, A^{k-1} are independent. Conclude that $k \leq n^2$.

15.2. Spaces of polynomials. Recall that a polynomial is an expression of the form $a_n x^n + \dots + a_0$, where x is an indeterminate, and the a_i are numbers. We denote by P the set of all polynomials. Recall that the degree of the polynomial is the highest i such that $a_i \neq 0$. We denote the set of polynomials of degree at most n by P_n . Both P and the P_n are vector spaces.

Example 115 (Example 6.3-1). Show that $1 + x, 3x + x^2, 2 + x - x^2$ are linearly independent.

Example 116 (Example 6.3-4). Show that any set of polynomials with distinct degrees is independent.

It follows from the last example that the dimension of P_n is $n + 1$: the set of monomials x^i , for $0 \leq i \leq n$, is a basis. Since P contains a space of arbitrary large dimension, it can not itself be finite dimensional. It has, however, an infinite basis: the set of all monomials is a basis for it.

If a is a number, and $p(x)$ is a polynomial, we may evaluate p at a (i.e., substitute a for x .) We thus get a function $T_a : p \mapsto p(a)$, from the set P (or P_n) to the set of numbers. It is obviously linear. The kernel K_a of T_a , the space of polynomials that have a root at a , is a subspace. For $n > 0$, this space is non-empty. This space can also be obtained as follows: we saw that if a is a root of $p \in P_n$, then it has the form $(x - a)q(x)$ for some $q \in P_{n-1}$. The map $q \mapsto (x - a)q$ is a linear map from P_{n-1} to P_n , and K_a is its image.

Example 117 (Example 6.4-3). Show that $x - a, (x - a)^2, \dots, (x - a)^n$ is a basis for K_a .

Example 118. Let $d : P \rightarrow P$ be the linear map given by $d(x^n) = nx^{n-1}$ for $n > 0$, and $d(1) = 0$. Let $m : P \rightarrow P$ be the map $m(q) = xq$. Compute $dm - md$.

15.3. Function spaces. The most general example is given by function spaces. If X is any set, and V is any vector space, the set $F_X(V)$ of all functions from X to V is again a vector space. All examples we have seen so far can be seen as subspaces of such spaces.

Let X be the set of numbers $1, \dots, n$, and let V be the set of numbers. Then $F_X(V)$ is (isomorphic to) \mathbb{k}^n : to function $f : X \rightarrow V$ we assign the tuple $f(1), \dots, f(n)$. An (open) interval (a, b) is the set of numbers x such that $a < x < b$. If X is a non-empty open interval, and V is again \mathbb{k} , we may evaluate a polynomial on any point of the interval, and thus the polynomial gives a function on X . Hence we get a map from the set P of polynomials to $F_X(V)$. This map is linear and injective, so P can be viewed as a subspace.

If $X = U$ is also a vector space, the set $\underline{Hom}(U, V)$ of linear maps from U to V is a subspace of $F_X(V)$. If $U = \mathbb{k}^m$ and $V = \mathbb{k}^n$, we saw that linear maps can be identified with $m \times n$ matrices, so $\underline{Hom}(U, V)$ is identified with $Mat_{m,n}$, which we already studied. In general, most of what we said for matrices also applies to linear maps.

If $f : X \rightarrow V$ is a function, the *support* of f is the set of elements $x \in X$ such that $f(x) \neq 0$. The set $F_X^f(V)$ of functions whose support is a finite set, is a subspace of $F_X(V)$. If $v \in V$ and $a \in X$, the function $f_{a,v} : X \rightarrow V$ defined by $f_{a,v}(a) = v$ and $f_{a,v}(x) = 0$ for $x \neq a$, is supported on a (at most), and so belongs to $F_X^f(V)$. It is clear that the set $\{f_{a,v} | a \in X\}$ is independent. Hence, if X is infinite (and V is non-zero), then $F_X^f(V)$ (and $F_X(V)$) is infinite dimensional.

The space of polynomials can be viewed as $F_X^f(V)$, where $V = \mathbb{k}$ and $X = \mathbb{N}$ is the set of natural numbers: the function $a : \mathbb{N} \rightarrow \mathbb{k}$ corresponds to the polynomials $a(0) + a(1)x + a(2)x^2 + \dots$ (the sum is finite since a has finite support.)

Other examples of function spaces are obtained from analysis. For instance, if X is an open interval, and $V = \mathbb{k} = \mathbb{R}$, the subset $C^n \subset F_X(V)$ of functions with n continuous derivatives is a subspace, and so is the subset of integrable functions, etc.

16. COMPLEX NUMBERS

16.1. Fields. The definition of a vector space involves numbers — a vector can be multiplied by a number. However, so far we almost neglected the question of what do we mean by a number. This was intentional — most of the theory works with an arbitrary notion of numbers that form a certain structure, called a *field*. Namely, a field is a set, together with operations of addition and multiplication, both of which are commutative ($x + y = y + x$ and $xy = yx$), there are elements 0 and 1 satisfying $x + 0 = x$ and $1x = x$ for all x , for any x there is an element $-x$ with $-x + x = 0$, and if $x \neq 0$, there is an element $\frac{1}{x}$ with $\frac{1}{x}x = 1$. Also, for any x , y and z , $x + (y + z) = (x + y) + z$, $x(yz) = (xy)z$ and $x(y + z) = xy + xz$.

Example 119. The set \mathbb{Q} of rational numbers is a field (with the usual operations.)

Example 120. The set \mathbb{N} of natural numbers is not a field: if n is a positive number, there is no $-n$ in \mathbb{N} with $-n + n = 0$.

Example 121. The set \mathbb{Z} of integers is still not a field: it contains $-n$ for any element n , but not $\frac{1}{n}$ (for most n .)

Example 122. The set of 2×2 matrices is not a field: it has operations of addition and multiplication, but the multiplication is not commutative.

Example 123. A field can be finite: let \mathbb{F}_2 be the set $\{0, 1\}$, with usual multiplication, but with addition mod 2: $1 + 1 = 0$. Then \mathbb{F}_2 is a field (similarly, there is, for any prime number p , a field \mathbb{F}_p with p elements, where the operation are performed mod p .)

As mentioned above, any field can be used as a set of numbers in the definition (and theory) of vector spaces. This includes linear maps, dimension, determinants, eigenvectors and eigenvalues, etc.

However, for parts of the theory, additional assumptions on the field are required. For example, in the discussion of scalar products, we required the field to be ordered, with some conditions on the squares.

We note that for any field, we may consider polynomials with coefficients from this field. If p is such a polynomial, and a is an element of the field, $p(a)$ is again an element of the field. As usual, if $p(a) = 0$, then a is called a root of p . It then makes sense to ask, for a given polynomial $p(x)$, does p have a root in the field? The most important question about a field is: which polynomials have roots in the field? The simplest possible answer is: all of them! Such fields exist, and they are called *algebraically closed*.

16.2. Real numbers. When considering plane geometry, we have identified the points of a line (after fixing a point and length unit) with numbers. The field of numbers obtained in this way is called the *real numbers*, and is denoted by \mathbb{R} . The precise description of \mathbb{R} is outside the scope of this subject, but we will list properties that, in a sense, characterise it.

The field of real numbers is not algebraically closed: the square of any real number is positive, so the polynomials $x^2 + 1$ does not have a root. However, this field can be described as follows:

- (1) Any polynomial of *odd* degree has a root.
- (2) For any non-zero number x , exactly one of x , $-x$ has a square root.
- (3) Any integer is real

16.3. Complex numbers. The field \mathbb{C} of complex numbers is obtained from the reals by adding the square root of -1 , which is denoted by i . Thus the complex numbers include the real numbers, and also an element i that satisfies $i^2 = -1$. Since a field should be closed under the multiplication, it should contain all expressions of the form $a_0 + a_1i + \cdots + a_ni^n$ (where the a_i are real), and their inverses. However, since $i^2 = -1$, any such expression can be written as $z = a + bi$, where a and b are real. The numbers a and b are called the *real and imaginary parts* of z , respectively. Furthermore, inverses are also included in this description:

Proposition 124. *For any real numbers a, b , not both 0, there are real numbers c, d such that $(a + bi)(c + di) = 1$. Any complex number can be represented uniquely as $a + bi$ for some real a, b .*

Proof. $(a + bi)(c + di) = ac - bd + (ad + bc)i$. Consider the matrix $\begin{bmatrix} a & -b \\ c & d \end{bmatrix}$. The real part of the product is the scalar product of the rows of the matrix, and the imaginary part is the determinant. Thus we are looking for a vector (c, d) such that the determinant is 0, i.e., (c, d) is colinear with (a, b) . The fact that the scalar product is 1 then translate to saying that the length of (c, d) is the inverse of the length of (a, b) , and that they are in the same direction. This determines (c, d) uniquely. Explicitly, $c + di = \frac{1}{a^2 + b^2}(a - bi)$.

The existence in the second statement follows from the first. For the uniqueness, assume that $a + bi = c + di$. Then $a - c = (d - b)i$. The square of the left side is non-negative, and the square of the right is non-positive. Hence both are 0. \square

Example 125. Present $\frac{2+3i}{4-i}$ in the form $a + bi$.

It follows that the complex numbers form a vector space of dimension 2 over \mathbb{R} , with $1, i$ as a basis. In particular, we may identify complex numbers with points in the plane. The number $\sqrt{a^2 + b^2}$ whose square appears in the above proof is then nothing but the length of the vector determined by the point $z = a + bi$. It is denoted by $\|z\|$. The number $a - bi$ is the *complex conjugate* of z , denoted \bar{z} . Geometrically, it corresponds to reflection through the real axis. In particular, the complex number z is in fact real if and only if $z = \bar{z}$. And the above proof shows that for $z \neq 0$, $\frac{1}{z} = \frac{\bar{z}}{\|z\|^2}$. We list some more properties of the complex conjugation and the length: $\overline{x + y} = \bar{x} + \bar{y}$, $\overline{xy} = \bar{x}\bar{y}$, $\bar{\bar{x}} = x$, $\|xy\| = \|x\|\|y\|$, $\|x\| \geq 0$, with equality only if $x = 0$.

In addition the vector space structure, we can multiply complex numbers. To study the multiplication, it is convenient to consider a different representation of complex numbers. As mentioned above, a complex number z has a length $\|z\|$. The set of all complex numbers of a given positive length forms a circle. Thus a non-zero complex number is given by its length, and its position on the circle. The position on the circle is determined by the angle, but only up to 2π : α and $\alpha + 2k\pi$ determine the same number (of a given radius), where k is an integer. The complex number determined by length r and angle α is denoted by $re^{i\alpha}$. If $z = a + bi$, the angle α is determined by the equation $\cos(\alpha) = \frac{a}{\|z\|}$ (equivalently, $\sin(\alpha) = \frac{b}{\|z\|}$). The reason for the exponential notation is:

Claim 126. *For any r, s, α and β , $(re^{i\alpha})(se^{i\beta}) = rse^{i(\alpha+\beta)}$.*

Proof. The fact that lengths are multiplied is easy. We assume that $r = s = 1$. Then

$$\begin{aligned} e^{\alpha} e^{\beta i} &= (\cos(\alpha) + i \sin(\alpha))(\cos(\beta) + i \sin(\beta)) = \\ &= (\cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta)) + i(\sin(\alpha) \cos(\beta) + \cos(\alpha) \sin(\beta)) \end{aligned}$$

The last expression is equal to $\cos(\alpha + \beta) + i \sin(\alpha + \beta)$. \square

In particular, for any integer n , $z^n = (re^{i\alpha})^n = r^n e^{in\alpha}$ (*DeMoivre's law*).

Example 127. Find the complex 3-rd roots of 8.

We noted above that there are fields where any polynomial has a root. It turns out that the complex numbers satisfy this condition: the field of complex numbers is algebraically closed. It follows that any polynomial is a product of linear factors $u(x - a_1) \dots (x - a_n)$ (with a_i the roots.) In this sense it is simpler than the reals. For example, the characteristic polynomial of any linear map has a root, so any linear map is guaranteed to have an eigenvector.

The decomposition above implies a decomposition for real polynomials:

Proposition 128. *Let $p(x)$ be a polynomial over the real numbers.*

- (1) *If z is a root of $p(x)$, then so is \bar{z}*
- (2) *$p(x)$ is a product of linear and quadratic factors*

Proof. (1) If $p(z) = 0$, then $\overline{p(z)} = 0$. But $\overline{p(z)} = p(\bar{z})$ for any z , since the coefficients of p are real.

- (2) Let a be a (complex) root of p . If a is real, then $p(x) = (x - a)q(x)$, where q has real coefficients. Otherwise, \bar{a} is also a (different) root, so $p(x) = (x - a)(x - \bar{a})q(x)$. However, $(x - a)(x - \bar{a})$ is in fact real (and quadratic), so $q(x)$ is also real. In both cases, $q(x)$ has by induction the required form. \square

17. REVIEW

Here is a summary of the main point of the theory outlined above.

17.1. Linear equation. The initial motivation for linear algebra is the study of linear equations (section 3.) Gaussian elimination is an algorithm that produces, for each system, its unique equivalent system in reduced row echelon form (theorem 20.) The reduced row echelon form allows us to “solve” the system.

17.2. Matrices and linear maps. The product AB of two matrices A and B is defined if the number of columns of A is equal to the number of rows of B (section 4.) Products and sums of matrices behave similarly to these operation with numbers, with the main differences being that products of matrices don't commute. In particular, if A is an $m \times n$ matrix, and x is an n -tuple, Ax is an m tuple, and so A defines a linear map from \mathbb{k}^n to \mathbb{k}^m . Conversely, any linear map between these spaces can be represented by such a matrix, which can be computed by applying the map to the standard basis (subsection 5.1.) Multiplication of matrices corresponds to composition of linear maps.

A system of linear equations can be written as $Ax = b$, where x is the tuple of variables. The set of solution is the set of tuples mapped by A to b . If $b = 0$, the system is called homogeneous, and the set of solutions is a subspace. For a general

b , given one solution x_0 , any other solution can be written as $x_0 + x$, where x solves the corresponding homogeneous equation (sub section 5.2.)

17.3. Invertible maps and matrices. A linear map $T : V \rightarrow U$ is invertible if there is $S : U \rightarrow V$ such that $T \circ S = I_U$ and $S \circ T = I_V$. In this case, we write $S = T^{-1}$. This is equivalent to saying that T is injective and surjective (theorem 49.) Such a map is also called an isomorphism between V and U . To check that T is injective, it is enough to show that $x = 0$ is the only solution to $T(x) = 0$.

If T is invertible and corresponds to the matrix A , the matrix of T^{-1} is denoted by A^{-1} . By definition, A is invertible if and only if the system $Ax = b$ has a unique solution for every b . In this case, the solution is $A^{-1}b$. A is invertible if and only if its reduced row echelon form is the identity matrix (sub section 6.1.) In particular, only square matrices can be invertible. To find the inverse, we perform the row operations that bring A to its reduced row echelon form on the identity matrix.

If A is a square matrix, then it is invertible if and only if it is injective, if and only if it is surjective, if and only if it has an inverse on either side (all this is false for non-square matrices.) A is invertible if and only if A^T is invertible, and in this case, $(A^T)^{-1} = (A^{-1})^T$.

The elementary row operations are linear maps on the columns. The corresponding matrices are called elementary matrices. Since a matrix A is invertible if and only if it can be brought by elementary row operations to the identity matrix, a matrix is invertible if and only if it is a product of elementary matrices (sub section 6.2.)

17.4. Determinants and adjugates. The determinant $\det(A)$ of a square matrix A is a number that can be computed from A . It satisfies $\det(AB) = \det(A)\det(B)$ (theorem 61), and $\det(A) \neq 0$ if and only if A is invertible (theorem 60.) It is linear in each column. The determinants of A and A^T are equal.

The adjugate $\text{adj}(A)$ of a matrix A is computed from its minors. It has the property that $\text{adj}(A)A = A\text{adj}(A) = \det(A)I$. From this we get Cramer's rule, which gives an expression for each coordinate of the solution of $Ax = b$ separately.

17.5. Bases, Dimension and Rank. A subset B of a vector space V spans V if any vector in V is a linear combination of vectors in B . B is linearly independent if no vector in B is a linear combination of the other. B is a basis of V if it is linearly independent and spans V . V is finite dimensional if there is a finite set B spanning V . In this case, V has a finite basis (claim 91.) Any two bases of V have the same size, and this size is called the dimension of V . Since \mathbb{k}^n has a basis of n vectors (the standard basis), its dimension is n .

A space has dimension n if and only if there is an invertible map from it to \mathbb{k}^n . Invertible maps preserve all linear structure: dimension, subspaces, bases, etc. A subspace of a finite dimensional space is itself finite dimensional, of smaller dimension if it is a proper subspace. Any linearly independent subset can be extended to a basis, and any spanning subset contains a basis.

If $T : U \rightarrow V$ is a linear map, the null space of T is the subspace of U of vectors u such that $T(u) = 0$. The image of T is the subspace of V of elements of the form $T(v)$. If U and V are finite dimensional, the rank of T is the dimension of the image. In this case, the sum of the rank of T and the dimension of the null space is equal to the dimension of U (section 12.)

If T is represented by a matrix A , the image $\text{col}(A)$ of A is spanned by the columns of A . The rank of A is also equal to the dimension of the row space (in other words, the ranks of A and A^T are the same, theorem 103) A basis for the null space of A can be computed via Gaussian elimination.

17.6. Similarity, Diagonalisation and Eigenvectors. Two square matrices A and B are similar if there is an invertible matrix P such that $A = P^{-1}BP$. If A and B are similar, then so are A^n and B^n , and A^T and B^T . A matrix is diagonalisable if it is similar to a diagonal matrix.

If a function F on matrices satisfies that $F(A) = F(B)$ when A and B are similar, then it can be defined for linear maps $T : U \rightarrow U$, where U is a finite dimensional space. In particular, the determinant, the characteristic polynomial, the rank and the trace (defined as the sum of entries on the main diagonal) are all well defined for such linear maps.

If a linear map $T : U \rightarrow U$, a number a , and $u \in U$ a non-zero vector satisfy $Tu = au$, then a is an eigenvalue of T , and u is an eigenvector for the eigenvalue a . The set E_a of vectors u satisfying $Tu = au$ is a subspace (it is the null space of $T - aI$), called the eigenspace of a . Thus a is an eigenvalue if E_a is non-zero. a is an eigenvalue if and only if it is a root of the characteristic polynomial $c_T(x) = \det(xI - T)$ (corollary 76.)

A matrix A is diagonalisable if and only if the space has a basis of eigenvectors of A (section 13.) In that case, the diagonalising matrix has the eigenvectors as columns, and the similar diagonal matrix has the corresponding eigenvalues on the diagonal. A is diagonalisable if and only if the dimension of each eigenspace E_a is equal to the multiplicity of a as a root of the characteristic polynomial $c_A(x)$. In general the dimension of E_a is at most this multiplicity.

17.7. Scalar products and Orthogonality. A scalar product is an additional structure on a vector space: it is a map taking a pair of vectors u, v to a number $u \cdot v$. It exists (at least) over the real numbers. If v is a vector, $\sqrt{v \cdot v}$ (positive square root) is the length of v . Vectors u and v are orthogonal if $u \cdot v = 0$.

An orthonormal basis is a basis of the space, such that the length of each vector is 1, and any two distinct vectors are orthogonal. If v_1, \dots, v_n is an orthonormal basis, the representation of v in this basis can be computed by taking the scalar product of v with the corresponding basis vector. The standard basis e_i in \mathbb{k}^n is orthonormal with respect to the usual scalar product. The Gram-Schmidt algorithm converts an arbitrary basis to an orthonormal one.

The orthogonal complement A^\perp of a subset A of a vector space with a scalar product is defined to be the set of all vectors orthogonal to every vector in A . It is always a subspace. $(A^\perp)^\perp$ is equal to the span of A (in particular, if A is a subspace, it is equal to A itself.) If A is a subspace, any vector in U can be written uniquely as a sum of a vector in A and a vector in A^\perp . If B is an orthonormal basis of A and C is an orthonormal basis of A^\perp then $B \cup C$ is an orthonormal basis of the whole space.

A square matrix P is orthogonal if its columns form an orthonormal basis. In this case, P^T is the inverse of P . A matrix A is orthogonally diagonalisable if it is diagonalised by an orthonormal matrix. This is equivalent to saying that the space has an orthonormal basis of eigenvectors, or that A is symmetric. In this

case, distinct eigenspaces are orthogonal, and the Gram-Schmidt algorithm can be used to find an orthonormal basis of each eigenspace.

17.8. Linear geometry. The set of vectors in the plane (or 3-space) forms a vector space. Addition in these spaces is given by the parallelogram law. Fixing axes and a unit of length, we get a linear isomorphism of these spaces with \mathbb{R}^2 and \mathbb{R}^3 . Linear subspaces correspond to sub-lines and sub-planes.

The scalar product in \mathbb{R}^2 and \mathbb{R}^3 determines lengths and angles in the space. The length of a vector is given by the scalar product length, and the angle α between two non-zero vectors u and v is given by $\cos(\alpha) = \frac{u \cdot v}{\|u\| \|v\|}$. This definition makes sense in an arbitrary space with scalar product, because of Cauchy's inequality: $u \cdot v \leq \|u\| \|v\|$. Projections on a line or a plane are given by the scalar product projection. A matrix is orthogonal if and only if it preserves lengths and angles.

The determinant of a matrix is the (oriented) volume of the parallelepiped determined by the columns of the matrix.

Examples of linear maps are projections, rotations, reflections, and rescaling in arbitrary directions.

17.9. Complex Numbers. Any complex number can be presented uniquely in the form $a + bi$, where a and b are real numbers, and $i^2 = -1$. The complex numbers form a field: any non-zero complex number $z = a + bi$ has an inverse, which can be computed as $\frac{\bar{z}}{\|z\|^2}$, where \bar{z} is the complex conjugate $a - bi$, and $\|z\|^2$ is the square of the length $a^2 + b^2$.

A non-zero complex number can also be presented as $z = r e^{i\alpha}$, where r is a unique real positive number, the length of z , and α is the angle, which is determined up to an integer multiple of 2π . $e^{i\alpha} = \cos(\alpha) + i \sin(\alpha)$. This presentation is convenient when multiplying, since $r_1 e^{i\alpha_1} r_2 e^{i\alpha_2} = r_1 r_2 e^{i(\alpha_1 + \alpha_2)}$.

The field of complex numbers is algebraically closed: any non-constant polynomial with complex coefficients has a root. In other words, any complex polynomial is a product of linear factors. It follows that any real polynomial is a product of linear and quadratic factors.

REFERENCES

- [Lan89] Serge Lang, *Linear algebra*, third ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1989. MR MR996636 (90b:15001) 1.2
- [Nic06] W. Keith Nicholson, *Linear algebra with applications*, fifth ed., McGraw-Hill Ryerson Higher Education, 2006. (document)
- [Wik07] Wikipedia, *Cartesian coordinate system* — *wikipedia, the free encyclopedia*, 2007, [Online; accessed 9-September-2007]. 1.2

INDEX

- characteristic polynomial
 - of a linear map, 33
 - of a matrix, 25
- column vector, 12
- complex number, 38
 - conjugate, 38
 - imaginary part of, 38
 - real part of, 38
- coordinates, 2
- Cramer's Rule, 22

- DeMoivre's law, 39
- dot product, 6

- eigenvalue, 25
- eigenvector, 25
- elementary operations, 8

- field, 37
 - algebraically closed, 37
 - of real numbers, 37

- Gauss elimination, 9

- identity matrix, 13
- indeterminates, 1
- inner product, 6
- interpolation polynomial, 24

- law of cosines, 4
- linear equation, 1, 8
- linear isomorphism, 17
- linear map, 14
 - image of, 27
 - injective, 16
 - invariant subspace, 33
 - inverse of, 16
 - invertible, 16
 - kernel of, *see* null space
 - null space of, 27
 - one-to-one, *see* injective
 - onto, *see* surjective
 - rank, 31
 - surjective, 16
- linear transformation, *see* linear map

- matrix, 11
 - adjugate of, 18, 22
 - cofactor of, 20
 - column space, 31
 - conjugate, 24
 - determinant of, 18
 - diagonal, 24
 - diagonalisable, 24
 - orthogonally diagonalisable, 34
 - product, 12
 - rank of, 10
- reduced row echelon form, *see* system of linear equations
- row echelon form, *see* system of linear equations
- row space, 31
- square, 12
- symmetric, 12
- upper triangular, 20

- parameter variables, 9
- polynomial, 22
 - degree of, 22
 - monic, 22
 - root of, 22
 - multiplicity of, 33

- rank
 - of a linear map, *see* linear map
 - of a matrix, *see* matrix
 - of a system, *see* system of linear equations

- scalar product, 6, 29–30
- similar matrix, *see* matrix, conjugate
- standard basis, 14
- subspace, 35
- system of linear equations, 8
 - equivalent, 8
 - homogeneous, 8
 - rank, 10
 - reduced row echelon form, 9
 - row echelon form, 8
 - solution of, 1, 8

- trace, 32

- unknown variables, 1

- vector space, 3, 12, 13, 17
 - basis, 14
 - dimension of, 28
 - finite dimensional, 28
- vectors
 - geometric, 2, 3
 - linear combination of, 14, 26
 - linearly independent, 26
 - located, 2
 - orthogonal, 29
 - orthonormal, 29
 - parallel, 4
 - projection of, 7
 - span of, 26
 - unit, 5

DEPARTMENT OF PURE MATH, UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA
E-mail address: <mailto:mkamensky@math.uwaterloo.ca>