Topic 2: Algebra

Seppe Staelens

July 5, 2024

Introduction

1 Lecture 1: Groups

A **group** $(G, \circ)$ is a set G on which we define an operation $\circ$ that maps two elements of the group to another element in the same group. In mathematical notation this is written as

$$\circ : G \times G \to G : (g, h) \mapsto g \circ h \tag{1}$$

The group has to satisfy the following:

1. There must exist a **neutral element** of the group, which will here be indicated with e (or more specifically e_G), which satisfies: $\forall g \in G : g \circ e = e \circ g = g$.
2. The operation $\circ$ must be **associative**: $\forall f, g, h \in G : f \circ (g \circ h) = (f \circ g) \circ h$.
3. For each element $g \in G$, there must exist an **inverse** element h such that $g \circ h = h \circ g = e$.

Sometimes the group operation is written as $+$, in which case the neutral element is often denoted as 0. This is called the additive notation. Alternatively, the group operation can be written as $\star$, in which case the neutral element is often denoted as 1. This is called the multiplicative notation.

Proposition 1. *The neutral element of a group is necessarily unique.*

Proof. Suppose the group $(G, \circ)$ has two neutral elements for the operation $\circ$, e_A and e_B. By definition of the neutral element, we then must have the following sequence of equalities:

$$e_A = e_A \circ e_B = e_B \,.$$

So, we find that the two neutral elements must be in fact the same element, which is therefore unique. □

Similarly, we find that the inverse of an element is also unique.

Proposition 2. *The inverse of an element in a group is unique.*

Proof. The proof is left as an exercise to the reader. **SJS: Can be done in lecture.** □

The inverse of an element g is often denoted as g^{-1} or $-g$ in the multiplicative and additive notation, respectively.

Exercise 1. *Let $(G, \circ)$ be a group. Show that $\forall g, h \in G$ we have $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$ and $(g^{-1})^{-1} = g$.* ***SJS: Can be done in lecture.***

An **Abelian group** is a group for which the operation is commutative, i.e.

$$\forall g, h \in G : g \circ h = h \circ g \,.$$

1.1 Examples

This section discusses some examples of well-known groups. **SJS: I will list some of the first examples in the lecture, but then we will start to derive the smallest groups. For 1,2,3 elements there are only cyclic groups. For 4 elements, we have the cyclic group of size 4, OR the Klein four-group.**

- $(\mathbb{N}, +), (\mathbb{Q}, +), (\mathbb{R}, +), (\mathbb{C}, +)$ are all Abelian groups with neutral element 0.
- $(\mathbb{R}_0, \cdot)$ is an Abelian group with neutral element 1.
- The real matrices form an Abelian group with respect to the addition $(\mathbb{R}^{m\times n}, +)$.
- The real matrices do not form a group with respect to the multiplication $(\mathbb{R}^{m\times n}, \cdot)$. Why? Instead, the square invertible matrices form a group with respect to the multiplication. This group is denoted as $(GL(n, \mathbb{R}), \cdot)$.
- The linear functions with $f(x) = ax + b, a \neq 0$, form a group with respect to the composition of functions.
- The cyclic group $(\mathbb{Z}_n, +)$ is a group with n elements that contains a neutral element e and a *generator* a, which generates all the other elements as $a + a, a + a + a, \ldots$, i.e. by repeatedly adding a to itself. Denoting these elements as $e, a^1 \equiv a, a^2 \equiv a + a, \ldots, a^{n-1} \equiv a + a + \cdots + a$, we finally have that $a^{n-1} + a = e$.
- The Klein four-group $(V, \oplus)$ is a group with four elements, defined by the following **Cayley table**:

$\oplus$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

This table should be read as e.g.

$\oplus$	b
a	$a \oplus b$

This group can be thought of as the symmetry group of a non-square rectangle, **SJS: as discussed in the lecture**.

1.2 Isomorphisms

It is possible to construct larger groups from known smaller groups. A way to do this is by means of the direct product.

Definition 1. *The **direct product** of two groups $(G, \circ)$ and $(H, \star)$ is the group $(G \times H, \square)$, containing elements (g, h) with $g \in G, h \in H$ .The operation $\square$ is defined as*

$$(g_1, h_1)\square(g_2, h_2) = (g_1 \circ g_2, h_1 \star h_2)\,. \tag{2}$$

However, this does not always mean we are making a "new group". **SJS: Here I will have them construct the Cayley table of $\mathbb{Z}_2 \times \mathbb{Z}_2$, equipped with regular addition.** Consider the Cayley table of $\mathbb{Z}_2 \times \mathbb{Z}_2$, both equipped with the regular addition. This table is given by

$+$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,1)$	$(0,1)$	$(0,0)$	$(1,1)$	$(1,0)$
$(1,0)$	$(1,0)$	$(1,1)$	$(0,0)$	$(0,1)$
$(1,1)$	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$

Comparing this to the Cayley table of the Klein four-group, we see that these have the same structure. This means that, in a way, these groups are the *same*, we just gave the elements different names. To make this more precise, we introduce the concept of an isomorphism.

Definition 2. *Let $(G, \circ)$ and $(H, \star)$ be two groups. A homomorphism from G to H is a map $\phi : G \to H$ that satisfies*

$$\forall g_1, g_2 \in G : \phi(g_1 \circ g_2) = \phi(g_1) \star \phi(g_2)\,. \tag{3}$$

An isomorphism is a homomorphism that is also bijective. In this case, the groups are said to be isomorphic, and we write $G \cong H$.

Note that in equation (3) the operation $\star$ is used for the objects $\phi(g_1)$ and $\phi(g_2)$, as by definition they are elements of H.

Exercise 2. *Given the Cayley tables of $(V, \oplus)$ and $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$, show that these groups are the same by giving an explicit isomorphism between them.* ***SJS: Can be done in lecture.***

Theorem 1. *Given groups $(G, \circ)$ and $(H, \star)$ and a homomorphism $\phi : G \to H$. Denoting the neutral elements with e_G, e_H respectively, we have that*

1. $\phi(e_G) = e_H$,
2. $\phi(g^{-1}) = \phi(g)^{-1}$.

Furthermore, if ϕ is an isomorphism, then the inverse map $\phi^{-1} : H \to G$ is also an isomorphism.

Proof. The proof is left as an exercise to the reader. **SJS: Can be done in lecture.** □

1.3 Subgroups

Definition 3. *A subset H of a group G is called a* ***subgroup*** *if it is itself a group with respect to the operation of G.*

Exercise 3. *Show that the set of even integers is a subgroup of the group of integers with respect to addition.* ***SJS: Can be done in lecture.*** *Is the set of odd integers also a subgroup?*

Exercise 4. *Determine all subgroups of the cyclic group of order 12, $\mathbb{Z}_{12}$.* ***SJS: Can be done in lecture.***

Proposition 3. *Given two groups $(G, \circ)$ and $(H, \star)$, and a homomorphism $\phi : G \to H$. Given any subgroup K of G, the image of K under ϕ is a subgroup of H.*

Proof. We have to show that $\phi(K)$ is a subgroup of H, i.e. we have to show that it is a group. Since K is a group, $e_G \in K$, and therefore $e_H = \phi(e_G) \in \phi(K)$. Associativity of the operation $\star$ is inherited from the fact that H is a group. The set $\phi(K)$ is closed under the operation $\star$ as well: given any two elements of $\phi(K)$, say $\phi(k_1)$ and $\phi(k_2)$, we have that $\phi(k_1) \star \phi(k_2) = \phi(k_1 \circ k_2) \in \phi(K)$. Finally, we need to check that every element of $\phi(K)$ has an inverse in $\phi(K)$. Given an element $\phi(k) \in \phi(K)$, we have that $\phi(k^{-1}) \in \phi(K)$, as $k^{-1} \in K$ since K is a group. Therefore, $\phi(K)$ is a subgroup of H. □

Exercise 5. *Are the following groups isomorphic to $\mathbb{Z}_6$?*

- $\mathbb{Z}_2 \times \mathbb{Z}_3$
- $\mathbb{Z}_3 \times \mathbb{Z}_2$
- S_3. *This is the group of symmetries of an equilateral triangle. The elements of the groups are the reflections (around the axes of symmetry) and the rotations (by 120° and 240°). Start by listing and labelling these basic elements, and form the Cayley table.*

2 Lecture 2: Linear algebra

Having studied groups, we will now impose additional structure on sets to turn them into vector spaces. This is one of the central concepts in the field of linear algebra, which in itself forms the basis of any scientific / engineering degree.

Definition 4. *A **real vector space** is an Abelian group $(V, +)$ that is furthermore equipped with a scalar multiplication*

$$\cdot : \mathbb{R} \times V \to V : (\lambda, v) \mapsto \lambda \cdot v \,. \tag{4}$$

If the distinction between the scalars and group elements is clear, we will often simply denote $\lambda \cdot v$ as λv. The two operations have to satisfy the following properties $\forall v, w \in V, \lambda, \mu \in \mathbb{R}$:

1. *Identity element for scalar multiplication: $1v = v$ for all $v \in V$.*
2. *Distributivity w.r.t. the scalar multiplication: $\lambda \cdot (v + w) = \lambda v + \lambda w$.*
3. *Distributivity w.r.t. the vector addition: $(\lambda + \mu)v = \lambda v + \mu v$.*
4. *Compatibility of scalar multiplication with field multiplication: $(\lambda\mu)v = \lambda(\mu v)$.*

2.1 Examples

Some basic examples include:

- The set of real numbers $\mathbb{R}$ is a vector space with respect to the regular addition and multiplication.
- The set of real matrices $\mathbb{R}^{m\times n}$ is a vector space with respect to the regular matrix addition and scalar multiplication.
- The set of real functions $\mathcal{F}(\mathbb{R}, \mathbb{R})$ (i.e. functions $f : \mathbb{R} \to \mathbb{R}$) is a vector space with respect to the regular function addition and scalar multiplication, defined for all functions $f, g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ and $\lambda \in \mathbb{R}$ as
$$(f + g)(x) = f(x) + g(x)\,, \qquad (\lambda f)(x) = \lambda f(x)\,.$$
- In quantum mechanics, wave functions $|\psi\rangle$ are elements of a *Hilbert space*, which is a vector space with more structure added onto it. They can be added together and multiplied by complex numbers, as for example in the famous example of Schrödinger's cat who is "alive and dead" with wave function
$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\text{alive}\rangle + |\text{dead}\rangle)\,.$$

2.2 Linear transformations

Much like we had homomorphisms between groups, we can define linear transformations between vector spaces.

Definition 5. *A **linear transformation** from a vector space V to a vector space W is a map $L : V \to W$ that satisfies the following:*

$$\forall \lambda, \mu \in \mathbb{R}, \forall v, w \in V : L(\lambda v + \mu w) = \lambda L(v) + \mu L(w)\,. \tag{5}$$

Exercise 6. *Show that the composition of two linear transformations is again a linear transformation.*

Theorem 2. *If $L : V \to W$ is a linear transformation, then the following statements are true:*

1. *$L(0_V) = 0_W$.*
2. *$L(-v) = -L(v)$.*

Proof. The first statement follows from the linearity of L:

$$L(0_V) = L(0_V + 0_V) = L(0_V) + L(0_V) \Rightarrow L(0_V) = 0_W \,.$$

The second statement follows from the first, as

$$0_W = L(0_V) = L(v + (-v)) = L(v) + L(-v) \Rightarrow L(-v) = -L(v) \,.$$

□

Exercise 7. *Show that the set of all linear transformations from a vector space V to itself forms a vector space under the regular addition and scalar multiplication of functions. This vector space is denoted as $\mathcal{L}(V)$.*

2.3 Bases of vector spaces

It turns out that vector spaces have a well defined structure, which can be described by a basis. We first need to define the concepts of linear independence and span.

Definition 6. *A set of vectors $\{v_1, \ldots, v_N\}$ is called **linearly independent** if the only way to write the zero vector as a linear combination of these vectors is by setting all the coefficients to zero, i.e.*

$$\sum_{i=1}^{N} \lambda^i v_i = 0 \Rightarrow \forall 1 \leq i \leq N : \lambda^i = 0 \,. \tag{6}$$

*If a set of vectors is not linearly independent, it is called **linearly dependent**.*

For example, the vectors $(1,0), (0,1) \in \mathbb{R}^2$ are linearly independent, as the only way to write the zero vector $(0,0)$ as a linear combination of these vectors is by setting both coefficients to zero. However, the vectors $(1,0), (2,0)$ are not linearly independent, as the zero vector can be written as $1 \cdot (2,0) + (-2) \cdot (1,0) = (0,0)$.

Definition 7. *The **span** of a set of vectors $\{v_1, \ldots, v_N\}$ is the set of all possible linear combinations of these vectors, i.e.*

$$\operatorname{span}\{v_1, \ldots, v_N\} = \left\{ \sum_{i=1}^{N} \lambda^i v_i \mid \lambda^i \in \mathbb{R} \right\} \,. \tag{7}$$

For example, in $\mathbb{R}^3$, the span of the vectors $(1,0,0), (0,1,0)$ is the xy-plane, as any vector in the xy-plane can be written as a linear combination of these two vectors:

$$(x, y, 0) = x \cdot (1,0,0) + y \cdot (0,1,0) \,.$$

Exercise 8. *Show that the span of a set of vectors is a vector space.*

Definition 8. *A set of vectors $\{e_1, \ldots, e_N\}$ is called a **basis** of a vector space V if they are linearly independent and span V.*

Such a basis is in general not unique. However, it turns out that the number of basis vectors is fixed for a given vector space. Once a basis is chosen, any vector in the vector space can be written as a linear combination of the basis vectors, and this representation is unique.

Theorem 3. *Given a basis $\{e_1, \ldots, e_N\}$ of a vector space V, any vector $v \in V$ can be written as*

$$v = \sum_{i=1}^{N} v^i e_i \tag{8}$$

in a unique way.

Proof. Suppose that v can be written in two ways, i.e.

$$v = \sum_{i=1}^{N} v^i e_i \,,$$
$$v = \sum_{i=1}^{N} w^i e_i \,.$$

Subtracting the two expressions, we find that

$$0 = \sum_{i=1}^{N} (v^i - w^i) e_i \,.$$

Since the basis vectors are linearly independent, we must have that $v^i - w^i = 0$ for all i. □

The proof of the following theorem would take us too far, but forms a cornerstone of the field of linear algebra.

Theorem 4. *Every vector space V admits a basis. The **dimension** of a vector space V is the number of basis vectors in any basis of V. This number is denoted as $\dim(V)$. If the dimension is finite, we say that the vector space is **finite-dimensional**.*

The following theorem is a direct consequence of the definition of a basis.

Theorem 5. *A linear transformation $L : V \to W$ is completely determined by its action on the basis vectors.*

Proof. Take a basis of the N-dimensional vector space V, $\{e_1, \ldots, e_N\}$. Then any vector $v \in V$ can be written as $v = \sum_{i=1}^{N} v^i e_i$. The linear transformation acting on v then satisfies

$$L(v) = \sum_{i=1}^{N} v^i L(e_i)$$

Therefore, once the linear transformation of the basis vectors is known, it is known for all the vectors. □

We can also create linear maps between different vector spaces.

Theorem 6. *Any vector space V is isomorphic to $\mathbb{R}^N$ for $N = \dim(V)$.*

Proof. Take a basis of V, $\{e_1, \ldots, e_N\}$. Then any vector $v \in V$ can be written as $v = \sum_{i=1}^{N} v^i e_i$. The map $\phi : V \to \mathbb{R}^N$ that sends v to the vector $(v^1, \ldots, v^N)$ is an isomorphism. This means we implicitly take the standard basis of $\mathbb{R}^N$, which is the set of vectors $(1, 0, 0, \ldots), (0, 1, 0, \ldots), \ldots$. □

3 Lecture 3: Matrices

Normally you are already familiar with vectors in $\mathbb{R}^n$. These consist of an ordered collection of n real numbers, and we usually denote these vectors as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} . \tag{9}$$

Matrices are similar, but add a dimension. A $m \times n$ matrix is a collection of $m \cdot n$ real numbers, ordered in m rows and n columns. Such a matrix looks like

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

The set of these matrices is denoted by $\mathbb{R}^{m\times n}$. The **diagonal** of a matrix is the collection of elements a_{ii}, i.e. the elements with the same row and column index. Matrices for which the only non-zero elements are found on the diagonal are called **diagonal matrices**.

3.1 Matrix operations

The addition of matrices is fairly straigthforward. Addition between matrices is only well-defined between matrices of the same shape. The resulting matrix is simply obtained by adding the elements in the corresponding positions. For example, the addition of two 2×3 matrices A, B is

$$A + B = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{pmatrix} .$$

Note that we can write this more compactly using element notation. Let's define the matric $C = A + B$, then we say that the element

$$c_{ij} = a_{ij} + b_{ij} , \qquad \forall 1 \leq i \leq m, \forall 1 \leq j \leq n .$$

In this notation, c_{ij} refers to the element on row i and column j of the matrix C.

The multiplication of matrices is less straigthforward. Even though an elementwise multiplication seems straigthforward, this turns out to not be very interesting. Instead, for many applications the *matrix multiplication*, defined as below, is way more useful.

The product of two matrices is not always well-defined, *but* allows for matrices of different types. First of all, it should be noted that the multiplication of matrices is *not* commutative. The product of A with B is only well-defined if the number of *columns* in A matches the number of *rows* in B. For example, $A \in \mathbb{R}^{2\times 3}$ and $B \in \mathbb{R}^{3\times 1}$ can be multiplied, but A and $C \in \mathbb{R}^{2\times 1}$ cannot. If we write $C = A \times B$ with $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times l}$, then the elements of $C \in \mathbb{R}^{m\times l}$ are given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{10}$$

Example 1. *Consider two matrices*

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \end{pmatrix} , \quad B = \begin{pmatrix} 2 & 0 \\ 1 & 1 \\ 3 & 1 \end{pmatrix} .$$

Note that $A \in \mathbb{R}^{2\times 3}, B \in \mathbb{R}^{3\times 2}$, and therefore we should have $C \equiv A \cdot B \in \mathbb{R}^{2\times 2}$. We calculate that

$$\begin{aligned} C &= A \cdot B \\ &\begin{pmatrix} 1\cdot 2+0\cdot 1+3\cdot 3 & 1\cdot 0+0\cdot 1+3\cdot 1 \\ 0\cdot 2+2\cdot 1+4\cdot 3 & 0\cdot 0+2\cdot 1+4\cdot 1 \end{pmatrix} \\ &= \begin{pmatrix} 11 & 3 \\ 14 & 6 \end{pmatrix} \end{aligned}$$

Alternatively, we can also calculate $D \equiv B \cdot A \in \mathbb{R}^{3\times 3}$. This is a different matrix indeed:

$$\begin{aligned} D &= B \cdot A \\ &= \begin{pmatrix} 2\cdot 1+0\cdot 0 & 2\cdot 0+0\cdot 2 & 2\cdot 3+0\cdot 4 \\ \vdots & \vdots & \vdots \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 & 6 \\ 1 & 2 & 7 \\ 3 & 2 & 13 \end{pmatrix} \end{aligned}$$

Two special types of matrices are the unit matrix $\mathbb{1}_n$ and the zero-matrix 0_n. The first one has zeroes everywhere, except for the *diagonal*, i.e. the elements with the same row and column index. The latter has zeroes everywhere. The n denotes the dimensions, as these matrices are square, i.e. elements of $\mathbb{R}^{n\times n}$. If the dimension is clear, these matrices are often simply denoted as $1, 0$ respectively.

Exercise 9. *Check that for all $A \in \mathbb{R}^{n\times n}$ the following holds:*

1. $A \cdot 1 = 1 \cdot A = A$
2. $A \cdot 0 = 0 \cdot A = 0$

Another basic matrix operation is the **transposition** of a matrix. This is simply the operation of flipping the matrix over its diagonal. The element a_{ij} at position (i, j) is then placed at position (j, i) in the transposed matrix. This operation is denoted as A^T. It should be clear that if $A \in \mathbb{R}^{m\times n}$, then $A^T \in \mathbb{R}^{n\times m}$.

Example 2. *The transpose of the matrix* $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ *is* $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

Exercise 10. *Show that for all $A, B \in \mathbb{R}^{m\times n}$ the following holds:*

1. $(A^T)^T = A$
2. $(A + B)^T = A^T + B^T$
3. $(\lambda A)^T = \lambda A^T$
4. $(A \cdot B)^T = B^T \cdot A^T$

Matrices for which the transpose is equal to the original matrix are called **symmetric**. Matrices for which the transpose is equal to the negative of the original matrix are called **antisymmetric**.

3.1.1 Inversion and determinant

Given that we have addition and multiplication for matrices, one might wonder whether the inverse operations also exist. Subtraction of matrices is straigthforward: if we have a matrix A, we can define $-A$ as the matrix where every element of A gets a minus sign. With this definition, $A-B$ can simply be interpreted as $A+(-B)$.

Does division also work for matrices? Given that multiplication is defined in a way that is *not* element-wise, division isn't either. The defining property for division of real numbers is that it is the inverse of

multiplication, i.e. $\frac{a}{b} = c \Leftrightarrow a = b \cdot c$. Because of this property, as you should be well aware, division by 0 is not allowed. So in similar spirit, we would like to define division of matrices as an operation such that

$$\frac{A}{B} := C \Leftrightarrow A = B \cdot C$$

However, a couple subtleties arise here:

- First of all, we have to remember that the multiplication of matrices is not commutative. Therefore, the definition above is not equivalent to $\frac{A}{B} := C \Leftrightarrow A = C \cdot B$. It seems like we need two kinds of divisions.
- Do we have a problem with matrix division analogous to division by zero?

In this course we will not go into the details of this problem, as there is a lot to be said about it. We simply summarize the main points:

1. For matrices, "division" is replaced by "inversion". Instead of dividing by a matrix A, we multiply with the *inverse matrix* A^{-1}.
2. The inverse matrix satisfies $A \cdot A^{-1} = 1 = A^{-1} \cdot A$.
3. Not all matrices have a well-defined inverse. Inverse matrices are only defined for square matrices, and within this subset only those that are *invertible*. Inversion is restricted to square matrices such that the left and right inverse are the same matrix. Square matrices are invertible if and only if their **determinant** is non-zero.

The determinant of a matrix is a scalar value that is calculated from the elements of the matrix. It is denoted as $\det(A)$ or $|A|$. Matrices that have determinant equal to zero are said to be singular, and do not have an inverse. This is the matrix equivalent of not being able to divide by zero.

We will not go into the details of how to calculate the determinant of a general matrix, but will give the formula for a 2×2 and $3x3matrix$ matrix.

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the determinant is given by

$$\det(A) = ad - bc\,. \tag{11}$$

For a 3×3 matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, the determinant is given by

$$\det(A) = aei + bfg + cdh - ceg - bdi - afh\,. \tag{12}$$

This can be graphically remembered as the sum of the products of the elements along the diagonals, minus the sum of the products of the elements along the anti-diagonals, as illustrated in Figure 1. To illustrate how determinants of larger matrices are calculated, note that (12) can be rewritten as

$$\det(A) = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}\,. \tag{13}$$

This illustrates how the determinant of a 3×3 matrix can be calculated by taking the determinants of the 2×2 matrices that are formed by removing one row and one column.

3.2 Linear transformations as matrices

We already mentioned how essentially every vector space is isomorphic to $\mathbb{R}^n$ for some n. When we choose a basis $\{e_1, \ldots, e_n\}$ for V, where n is thus the dimension of V, we can write any vector $v \in V$ as $v =$

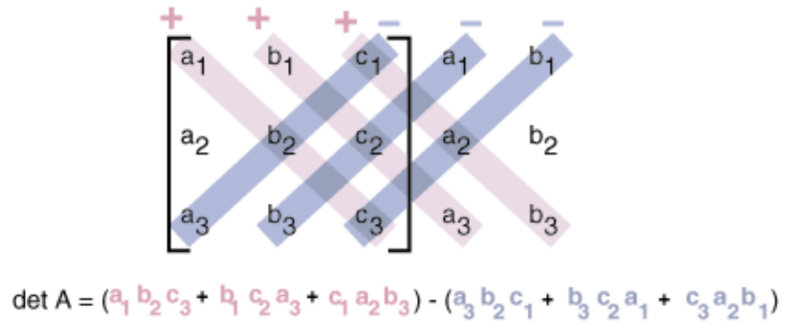


Figure 1: The graphical representation of the determinant of a 3×3 matrix.

$\sum_{i=1}^{n} v^i e_i$. We then define the linear transformation $L_e : V \to \mathbb{R}^n$ defined by mapping e_1 to $(1, 0, \ldots, 0)$, e_2 to $(0, 1, 0, \ldots, 0)$, and so on. This means that $L_e(v) = (v^1, v^2, \ldots, v^n)$ by linearity. Therefore, we can view any vector v in V as an n-dimensional vector in $\mathbb{R}^n$ that contains its coefficients in the basis $\{e_1, \ldots, e_n\}$.

However, this basis of V is not unique (nor is the standard basis of $\mathbb{R}^n$). Therefore, the linear transformation L is not unique either. With respect to a different basis $\{f_1, \ldots, f_n\}$ of V, the vector $v \in V$ would have different coefficients ν^i. The new linear transformation, which we will specify as L_f, would then map v to $(\nu^1, \nu^2, \ldots, \nu^n)$.

We can wonder what the relation is between the coefficients of v in the basis $\{e_1, \ldots, e_n\}$ and the coefficients of v in the basis $\{f_1, \ldots, f_n\}$. Take the basis vector f_1. Since it is a vector in V, we can write it as $f_1 = \sum_{i=1}^{n} f_1^i e_i$ with $f_1^i \in \mathbb{R}$. This is possible with any of the basis vectors in f. Observe now the following:

$$\begin{aligned} v &= \sum_{i=1}^{n} \nu^i f_i \\ &= \sum_{i=1}^{n} \nu^i \left(\sum_{j=1}^{n} f_i^j e_j \right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \nu^i f_i^j e_j \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \nu^i f_i^j \right) e_j \,. \end{aligned}$$

The coefficients of v in the basis $\{e_1, \ldots, e_n\}$ are thus given by $\sum_{i=1}^{n} \nu^i f_i^j$. But, we already defined these to be v^j. Therefore, we have that $v^i = \sum_{j=1}^{n} \nu^j f_j^i$.

We can rewrite this in matrix form as

$$\begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} f_1^1 & f_2^1 & \cdots & f_n^1 \\ f_1^2 & f_2^2 & \cdots & f_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ f_1^n & f_2^n & \cdots & f_n^n \end{pmatrix} \begin{pmatrix} \nu^1 \\ \nu^2 \\ \vdots \\ \nu^n \end{pmatrix} . \tag{14}$$

Alternatively, in matrix notation, we can write this as

$$\mathbf{v} = F\boldsymbol{\nu} \,. \tag{15}$$

What we are actually doing here, is defining a linear transformation F on V, defined by mapping the basis $\{f_1, \ldots, f_n\}$ to the basis $\{e_1, \ldots, e_n\}$. Indeed, the vector $\boldsymbol{\nu} = (1, 0, \ldots, 0)$, representing the vector f_1, is mapped to the first column of F, which is $(f_1^1, f_1^2, \ldots, f_1^n)$. These are exactly the coordinates of f_1 in the basis $\{e_1, \ldots, e_n\}$.

This illustrates how linear transformations can be represented by matrices. The matrix F represents the linear transformation $F : V \to V$ that maps the basis $\{f_1, \ldots, f_n\}$ to the basis $\{e_1, \ldots, e_n\}$.

3.3 Eigenvectors and Eigenvalues

In this section we discuss a more advanced topic in linear algebra. It combines knowledge of our previous discussions of matrices and linear transformations. For a moment, we forget about abstract vector spaces and focus on $\mathbb{R}^n$. Given a matrix A, we are interested in vectors that are mapped to a scalar multiple of themselves by the matrix A, i.e. vectors that satisfy

$$A\mathbf{v} = \lambda\mathbf{v}\,. \tag{16}$$

Definition 9. *A vector* $\mathbf{v}$ *is called an* ***Eigenvector***[1] *of a matrix* A *if it satisfies the equation* $A\mathbf{v} = \lambda\mathbf{v}$ *for some scalar* λ*. The scalar* λ *is called the* ***Eigenvalue*** *of the eigenvector* $\mathbf{v}$*.*

Equations of the form (16) often appear in physics and engineering, and are therefore thoroughly studied. The collection of eigenvalues of a matrix is called the **spectrum** of the matrix.

The paragraph below is rather advanced, but included for background information. The eigenvectors and eigenvalues of a matrix can be found by solving the equation $A\mathbf{v} = \lambda\mathbf{v}$. This equation can be rewritten as $(A-\lambda\mathbb{1})\mathbf{v} = 0$. This equation has non-trivial solutions if and only if the matrix $A-\lambda\mathbb{1}$ is singular, i.e. has a determinant of 0. Therefore, the eigenvalues of a matrix are the solutions to the equation $\det(A-\lambda\mathbb{1}) = 0$. The fundamental theorem of algebra states that a polynomial of degree n has n complex roots, counting multiplicities, meaning that a matrix has n complex eigenvalues, counting multiplicities. Usually, however, we are interested in the real eigenvalues. A corollary of the fundamental theorem of algebra is that any real matrix of odd dimension must have at least one real eigenvalue.

[1] The reason that these are written with a capital letter originates from German, where nouns are written with capital letters. *Eigen* is German for *own*, or *proper*. In practice, the capital letter is often omitted, however, as I will do as well.

4 Lecture 4: PageRank

4.1 A brief history of PageRank

//Generated by ChatGPT// PageRank, developed by Larry Page and Sergey Brin at Stanford University in 1996, is a foundational algorithm that significantly influenced the development of web search engines. It was designed to rank web pages in search engine results by measuring the importance of each page. The key idea behind PageRank is that a page's significance can be inferred from the number and quality of links pointing to it. Essentially, a page linked to by many high-ranking pages receives a higher rank itself. Page and Brin's approach provided a novel way of leveraging the web's link structure to improve search accuracy, which was a major advancement over the keyword-based search algorithms prevalent at the time.

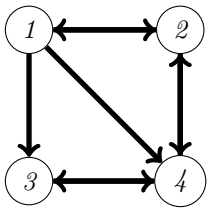
The algorithm became the core of Google's search engine, which Page and Brin co-founded in 1998. PageRank's ability to deliver more relevant and reliable search results quickly propelled Google to the forefront of the search engine market. The success of Google demonstrated the practical value of PageRank and influenced the development of other link-based ranking algorithms. Over time, while Google's ranking algorithms have become far more complex and incorporate hundreds of factors, PageRank's basic principle of link-based importance remains a cornerstone of modern search engine optimization (SEO) and digital marketing strategies.

4.2 The mathematics of PageRank

The basic idea, as stated above, is that the importance of a page can be inferred from the number and quality of links pointing to it. We want to capture this in a number, *the rank* of a webpage, which is a real number between 0 and 1. Suppose our internet consists of N webpages, and we denote the rank of webpage i as r_i. We require that all the ranks sum to 1, i.e. $\sum_{i=1}^{N} r_i = 1$. We now store all these ranks in a vector $\mathbf{r} \in \mathbb{R}^N$.

Suppose now that all these webpages have links between them, i.e. they refer to each other. This can be represented in a directed graph, where the nodes are the webpages and the edges are the links.

Example 3. *Consider a simple internet with 4 webpages, as shown in the figure below. The arrows indicate the links between the pages, i.e. webpage 1 links to webpage 2, 3 and 4, whereas 2 only links to 4.*



Exercise 11. *According to the above graph, which webpage seems the most important? Which one seems the least important?*

Imagine now that we have an enthousiastic web surfer, who starts at a random webpage and then clicks on a random link on that page. He keeps doing this for a while, and keeps track of which pages he visits.

Exercise 12. *Suppose the web surfer is on website 1. What is the probability that he will visit website 2, 3 and 4 next? What about the other websites? If the surfer keeps clicking, which website will he visit most often?*

We are going to represent these probabilities in a matrix, the linking matrix L. For this, we define the linking vector $\mathbf{l}_i$ for each webpage i. We first count the number of outgoing links from webpage i, and denote this number as n_i. The linking vector $\mathbf{l}_i$ is then a vector of length N, with $1/n_i$ at position j if there is a link from i to j, and 0 otherwise. Therefore, the vector $\mathbf{l}_i$ represents the probabilities of going from webpage i to all other webpages.

Exercise 13. *Calculate the linking vectors for the internet in Example 3. As an example. we give* $\mathbf{l}_1 = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}$.

We then define the linking matrix L as the matrix with the linking vectors as columns, i.e. $L = \begin{pmatrix} \mathbf{l}_1^T & \mathbf{l}_2^T & \dots & \mathbf{l}_N^T \end{pmatrix}$. This matrix clearly determines the structure of the internet, and we can use it to calculate the rank of the webpages.

Exercise 14. *How could one use this linking matrix to assign a level of importance to every website?*

Clearly, important websites will be those that are linked to by many other websites. In the linking matrix, this is visible as the rows that have few or no zeroes. Therefore, a good first idea to calculate the rank of a webpage is the following formula

$$\mathbf{r}_i = \sum_{j=1}^{N} L_{ij} \,. \tag{17}$$

However, this formula is prone to an issue.

Exercise 15. *What is the problem with simply defining the importance of a webpage by the number of links pointing to it? How can I easily create a website that is ranked very high? What could we do to circumvent this issue?*

If someone would want to create a website that is ranked very high, they could simply create a lot of websites that link to their own website. This would artificially inflate the rank of their website, and would not be a good representation of the actual importance of the website. This issue can be resolved by making a webiste important if it is linked to by many *important* websites.

Exercise 16. *This is clearly a circular definition. Why?*

The importance of webpages is now based on the importance of other webpages. It turns out that this will not be a problem, but we keep this in mind. We will incorporate this idea by calculating the rank of a page by weighing the links to it by the rank of the originating page. This is done as follows:

$$\mathbf{r}_i = \sum_{j=1}^{N} L_{ij} \cdot \mathbf{r}_j \,. \tag{18}$$

Remember that the elements L_{ij} signify whether there is a link from j to i. Therefore, this equation checks which websites link to website i, and incorporates how important they are. Note that the above equation can be written in matrix form as

$$\mathbf{r} = L \cdot \mathbf{r} \,. \tag{19}$$

Equation (19) reflects the core idea of the PageRank algorithm.

4.2.1 Solving equation (19)

As mentioned earlier, equation (19) is a circular definition.

Exercise 17. *Do you recognize what kind of equation this is? How can we solve it?*

Equation (19) is an Eigenvalue equation for the matrix L and an Eigenvector $\mathbf{r}$ with Eigenvalue 1.

The matrix L is called a *stochastic matrix*, as the sum of the elements in each column is 1.

Theorem 7. *A stochastic matrix, one whose rows or columns sum to 1, has an eigenvalue equal to 1. No other larger eigenvalue exists.*

Proof. Take a matrix M whose rows sum to 1. Given that the eigenvalues of a matrix are the same as those of its transpose, this is sufficient to prove the theorem. The vector $v = (1, 1, \ldots, 1)$ is an eigenvector of M with eigenvalue 1, as

$$M \cdot v = \begin{pmatrix} \sum_{j=1}^{N} M_{1j} \\ \sum_{j=1}^{N} M_{2j} \\ \vdots \\ \sum_{j=1}^{N} M_{Nj} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = v\,.$$

Therefore, the matrix M has an eigenvalue 1.

The proof that no other eigenvalue has a modulus larger than 1 can be done by means of the Gershgorin circle theorem. This however is beyond the scope of this course. □

To solve equation (19), we need to find an eigenvector with eigenvalue 1. We can do this by solving the equation $(L - \mathbb{1})\mathbf{r} = 0$, but this can be computationally redundant for large matrices.

There exists an iterative method to do this more efficiently, called the *power iteration method*. This method is based on the idea that if we multiply a vector $\mathbf{r}$ with a matrix L many times, it will converge to the eigenvector with eigenvalue the largest eigenvalue of L.

4.2.2 The power iteration method

We illustrate the power iteration method briefly, ignoring some of the subtleties that can arise.

Start with a vector $\mathbf{r}_0$, which can be chosen randomly. Suppose the vector $\mathbf{r}_0$ can be written as a a linear combination of the eigenvectors of L, i.e. $\mathbf{r}_0 = \sum_{i=1}^{N} a_i \mathbf{e}_i$, where the eigenvectors are ordered such that $\mathbf{e}_1$ has the largest eigenvalue. Then, after multiplying this vector with L many times, we get If we multiply this vector with L, we get

$$L\mathbf{r}_0 = \sum_{i=1}^{N} a_i L\mathbf{e}_i = \sum_{i=1}^{N} a_i \lambda_i \mathbf{e}_i\,. \tag{20}$$

If we repeat this process, we get

$$\begin{aligned} L^k\mathbf{r}_0 &= \sum_{i=1}^{N} a_i \lambda_i^k \mathbf{e}_i \\ &= \lambda_1^k \left(a_1\mathbf{e}_1 + \sum_{i=2}^{N} a_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \mathbf{e}_i \right)\,. \end{aligned}$$

If we assume that λ_1 is strictly larger than all other eigenvalues, we note that all the factors $\left(\frac{\lambda_i}{\lambda_1}\right)^k$ will go to zero for $k \to \infty$.

Therefore, after many multiplications with L, we have that

$$L^k\mathbf{r}_0 \approx \lambda_1^k a_1 \mathbf{e}_1\,. \tag{21}$$

In our case, $\lambda_1 = 1$, such that $L^k\mathbf{r}_0 \approx a_1\mathbf{e}_1$, i.e. the obtained vector is parallel with the eigenvector with eigenvalue 1. We can then simply calculate the rank vector $\mathbf{r}$ by normalizing the vector $L^k\mathbf{r}_0$, to find the PageRank vector that we were looking for.

Note that this method requires many matrix multiplications, and is therefore only efficient for sparse matrices, such as the linking matrix L.

4.2.3 The damping factor

Address problems of closed loops, pages without links.

5 References

These notes are based on my own knowledge of these basic mathematical concepts, and the writing has been accelerated by the use of *Github copilot* and its implementation in VSCode. Inspiration has been taken from the course notes for *"Algebraische Structuren"* (Algebraic Structures), used in the first year of the Bachelor of Mathematics at the KU Leuven, at the time taught by Prof. Raf Cluckers - also the author of the lecture notes. The section on PageRank has been based on (among others) this and this link. Proofs with respect to the spectrum of the linking matrix have been inspired by this link.