SCA Topic 2: Algebra
Oxbridge Academic Bootcamp 2025

Seppe J. Staelens

August 15, 2025

Introduction

Algebra is the branch of mathematics that studies the structure of sets, and in particular the operations that can be defined on them. We will introduce the concept of matrices, which are ubiquitous in mathematics and physics. They can be used to solve systems of linear equations, and can be used to represent linear transformations between vector spaces. We will put them to use by discussing the Google PageRank algorithm, which is based on the concept of eigenvectors of matrices.

Lecture 1: Matrices

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

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} . \tag{1}$$

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**.

1.1 Matrix operations

The addition of matrices is fairly straightforward. 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 matrix $C = A + B$, 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 straightforward. Even though an elementwise multiplication seems evident, 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

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{2}$$

Example 1. *Consider two matrices*

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

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 $\mathbb{1}, 0$ respectively.

Exercise 1. *Check that for all* $A \in \mathbb{R}^{n\times n}$ *the following holds:*

1. $A \cdot \mathbb{1} = \mathbb{1} \cdot A = A$
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 2. *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**.

1.1.1 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 straightforward: 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.

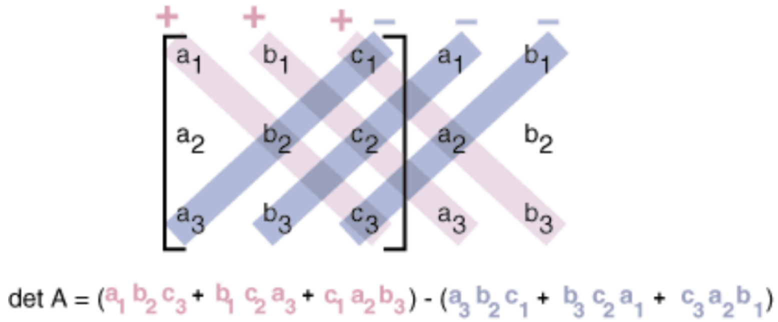


Figure 1: The graphical representation of the determinant of a 3×3 matrix.

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

$$\det(A) = ad - bc\,. \tag{3}$$

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

$$\det(A) = aei + bfg + cdh - ceg - bdi - afh\,. \tag{4}$$

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 (4) can be rewritten as

$$\det(A) = a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}\,. \tag{5}$$

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.

1.2 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

$$A\mathbf{v} = \lambda \mathbf{v}\,. \tag{6}$$

Definition 1. *A vector* $\mathbf{v}$ *is called an* ***Eigenvector***[a] *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* $\mathbf{v}$.

[a] 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.

Equations of the form (6) 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.

Exercise 3. *Def. 1 may seem arbitrary, but Eigenvectors really are 'special'. As an example, take the matrix* $A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$ *and try to find an Eigenvector* $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$*? Through trial and error, is it easy to find an Eigenvector? How many families of Eigenvectors can you find? What are their Eigenvalues.*

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.

Lecture 2: PageRank

2.1 A brief history of PageRank

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.

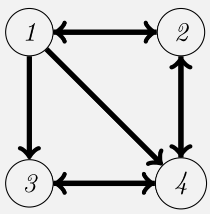
The algorithm became the core of Google's search engine, which Page and Brin co-founded in 1998. PageRank's ability to deliver more relevant and reliable search results quickly propelled Google to the forefront of the search engine market. The success of Google demonstrated the practical value of PageRank 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.

2.2 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 1 and 4.*



Exercise 4. *According to the above graph, which webpage seems the most important? Which one seems the least important?*

Imagine now that we have an enthusiastic 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 5. *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 6. *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. The linking matrix for the internet in Example 3 is then

$$L = \begin{pmatrix} 0 & . & . & . \\ 1/3 & . & . & . \\ 1/3 & . & . & . \\ 1/3 & . & . & . \end{pmatrix}. \tag{7}$$

Exercise 7. *How could one use this linking matrix to assign a level of importance to every website?*

Therefore, a good first idea to calculate the rank of a webpage is the following formula

$$\mathbf{r}_i = \sum_{j=1}^{N} L_{ij}\,. \tag{8}$$

However, this formula is prone to an issue.

Exercise 8. *What is the problem with simply defining the importance of a webpage as above? How can I easily create a website that is ranked very high? What could we do to circumvent this issue?*

This issue can be resolved by making a website important if it is linked to by many *important* websites.

Exercise 9. *This is clearly a circular definition. Why?*

$$\mathbf{r}_i = \sum_{j=1}^{N} L_{ij} \cdot \mathbf{r}_j\,. \tag{9}$$

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

$$\mathbf{r} = L \cdot \mathbf{r}\,. \tag{10}$$

Equation (10) reflects the core idea of the PageRank algorithm.

2.2.1 Solving equation (10)

As mentioned earlier, equation (10) is a circular definition.

Exercise 10. *Do you recognize what kind of equation this is?*

Theorem 1. *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. $\square$

To solve equation (10), 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.

2.2.2 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 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. If we multiply this vector with L, we get

$$L\mathbf{r}_0 = \sum_{i=1}^{N} a_i L\mathbf{e}_i = \sum_{i=1}^{N} a_i \lambda_i \mathbf{e}_i \,. \tag{11}$$

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 $k \to \infty$.

Exercise 11. *When could this method fail? How can we prevent this failure? When is this method (in)efficient?*

Therefore, after many multiplications with L, we have that

$$L^k\mathbf{r}_0 \approx \lambda_1^k a_1 \mathbf{e}_1 \,. \tag{12}$$

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.

2.2.3 The damping factor

With what we have discussed so far, two more problems can still arise.

Exercise 12. *Consider an internet with following linking matrix L:*

$$L = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Draw the corresponding graph, and discuss why we say that this graph has a "closed loop". Hint: think in terms of the random surfer, that follows links (arrows) at random.

Such a closed loop is a problem. The linking matrix is still *stochastic*, but if we apply a power method the resulting eigenvector will not be what we want the rank vector to represent. It turns out that the vector we obtain through power iteration in Exercise 12 will be $(0, 0, 0.42, 0.58)$.

Exercise 13. *Explain why this is not the desired outcome.*

A similar problem arises when we have a "dead end", i.e. a webpage that has no outgoing links. This will manifest itself in the linking matrix as a column of zeroes.

Exercise 14. *Draw a graph of a small internet with a dead end, and convince yourself that the linking matrix has a column of zeroes.*

In this case, it turns out that the power iteration method will give us a vector with all zeroes in the long run. This is because our linking matrix is no longer stochastic.

Luckily, there is a simple solution to this problem, called the *damping factor*. The damping factor is a parameter α that we introduce in the PageRank algorithm, and is usually set to a number around 0.8 or

0.9. The idea is that with probability α, the surfer will follow a link on the webpage he is on, and with probability $1-\alpha$, he will jump to a random webpage. The jumping to a random webpage is done to prevent the problems with closed loops and dead ends. It is represented by a matrix D, where all elements are $1/N$.

The new linking matrix is then given by

$$\mathcal{L} = \alpha L + (1-\alpha)D\,. \tag{13}$$

It turns out that the power method still converges with the adjusted linking matrix (13), and the PageRank vector can be calculated as before.

Exercise 15. *The adjusted linking matrix $\mathcal{L}$ can be thought of as introducing new links with small probability between the webpages. Calculate the new linking matrix for the internet in Example 12 with $\alpha = 0.85$. Interpreting the entries in the matrix as probabilities, show that this interpretation of introducing new links makes sense.*

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.