Notes on linear algebra

Darij Grinberg

Wednesday 4th December, 2019 at 15:09

These notes are **frozen** in a (very) unfinished state. Currently, only the basics of matrix algebra have been completed (products, triangularity, row operations etc.).

Contents

1. Preface

These notes are accompanying a class on applied linear algebra (Math 4242) I am giving at the University of Minneapolis in Fall 2016 (the website of the class is `http://www.cip.ifi.lmu.de/~grinberg/t/16f/`). They contain both the material of the class (although with no promise of timeliness!) and the homework exercises (and possibly some additional exercises).

There will (probably) be no actual applications in these notes, but only the mathematical material used in these applications. If time allows, the notes will contain tutorials on the use of SageMath (a computer algebra system suited both for numerical and for algebraic computations).

Sections marked with an asterisk (*) are not a required part of the Math 4242 course.

Several good books have been written on linear algebra; these notes are not supposed to replace any of them. Let me just mention four sources I can recommend[1]:

- Olver's and Shakiban's [OlvSha06] is the traditional text for Math 4242 at UMN. It might be the best place to learn about the applications of linear algebra.
- Hefferon's [Heffer16] is a free text that does things slowly but rigorously (at least for the standards of an introductory linear-algebra text). It has plenty of examples (and exercises with solutions), fairly detailed proofs, and occasional applications. (Which is why it is over 500 pages long; I hope you can easily decide what to skip based on your preferences.) Altogether, I think it does a lot very well. The main drawback is its lack of the theory of bilinear forms (but I don't know if we will even have time for that).
- Lankham's, Nachtergaele's and Schilling's [LaNaSc16] is a set of notes for introductory linear algebra, doing the abstract side (vector spaces, linear maps) early on and in some detail.
- Treil's [Treil15] is another free text; this is written for a more mathematically mature reader, and has a slight bias towards the linear algebra useful for functional analysis.[2]

[Please let the authors know if you find any errors or unclarities. Feel free to ask me if you want your doubts resolved beforehand.]

Also, some previous iterations of Math 4242 have left behind interesting notes:

[1] I have **not** read any of the books myself (apart from fragments). My recommendations are based on cursory skimming and random appraisal of specific points; I therefore cannot guarantee anything.

[2] The title of the book is a play on Axler's "Linear Algebra Done Right", which is biased towards analysis (or, rather, against algebra) to a ridiculous extent. Axler seems to write really well, but the usefulness of this book is severely limited by its obstinate avoidance of anything that looks too explicit and algebraic.

- Stephen Lewis, Fall 2014, `http://www.stephen-lewis.net/4242/` (enable javascript!).
- Natalie Sheils, Fall 2015, `http://math.umn.edu/~nesheils/F15_M4242/LectureNotes.html` (yes, those are on dropbox).

There are countless other sets of lecture notes on the internet[3], books in the library, and even books on the internet if you know where to look. You can find an overview of (published, paper) books in [Drucker12] (but usually without assessing their quality), and another (with reviews) on the MAA website `http://www.maa.org/tags/linear-algebra` . (Reviews on Amazon and goodreads are usually just good for a laugh.)

The notes you are reading are under construction, and will remain so for at least the whole Fall term 2016. Please let me know of any errors and unclarities you encounter (my email address is `darijgrinberg@gmail.com`)[4]. Thank you!

1.1. Acknowledgments

I would like to thank Mark Richard for correcting a typo in the notes.

2. Introduction to matrices

In this chapter, we shall introduce matrices, define the basic operations with matrices (addition, scaling, multiplication and powers) and two fundamental families of matrices (zero matrices and identity matrices), and state the most fundamental of their properties (and even prove some of them). We shall not go very deep here (most of this chapter corresponds to a part of [LaNaSc16, §A.2]), but we will give plenty of examples and some detailed proofs that will (hopefully) help you get some experience with the material.

2.1. Matrices and entries

In the following, we shall study matrices filled with numbers. This is not the most general thing to study (we could also fill matrices with other things, such as polynomials – and in fact, such matrices are highly useful); nor will we be very precise about it. In fact, for most of this chapter, we shall not even specify what we mean by "numbers", even though the word "number" is far from being a well-defined notion. However, as soon as we start caring about (say) computer

[3] Let me mention some: Two good-looking advanced texts for a mathematically prepared reader are Cameron's [Camero08] and Kowalski's [Kowals16]; a reader bored with the present notes might want to take a look at them. On the other side, Wildon's notes [Wildon16] include a lot of examples and geometric illustrations (but are probably too brief to peruse as a standalone text), whereas Chen's notes [Chen08] boast numerous applications (and seem quite readable, though I have not looked at them in depth).

[4] The sourcecode of the notes is also publicly available at `https://github.com/darijgr/lina` .

calculations, we will have to spend some words specifying our "numbers" more precisely. See Section 2.16 for what the word "number" means.

We shall use the symbol $\mathbb{N}$ for the set $\{0,1,2,\ldots\}$. This is the set of all nonnegative integers.[5]

Definition 2.1. If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, then an $n \times m$-*matrix* simply means a rectangular table with n rows and m columns, such that each cell is filled with a number.

For example, $\begin{pmatrix} 1 & 7 & 2 \\ -\sqrt{2} & 6 & 1/3 \end{pmatrix}$ is a 2×3-matrix, whereas $\begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}$ is a 3×1-matrix.[6]

The word "$n \times m$-matrix" is pronounced "n-by-m-matrix" (and sometimes even written in this way). Notice that the sign "$\times$" in "$n \times m$-matrix" is a symbol whose purpose is to separate n from m; it does **not** mean multiplication (although, of course, an $n \times m$-matrix does have nm entries). Do not rewrite "2×3-matrix" as "6-matrix" (after all, a 2×3-matrix is not the same as a 6×1-matrix, although $2 \cdot 3 = 6 \cdot 1$).

Definition 2.2. The word "*matrix*" will encompass all $n \times m$-matrices for all possible values of n and m.

Definition 2.3. The *dimensions* of an $n \times m$-matrix A are the two integers n and m. When they are equal (that is, $n = m$), we say that A is a *square matrix*, and call n its *size*.

For a rectangular matrix A, we also will sometimes say that "A has size $n \times m$" when we mean that "A is an $n \times m$-matrix".

Definition 2.4. If A is an $n \times m$-matrix, and if $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, then $A_{i,j}$ will denote the entry of A in row i and column j. This entry is also called the (i,j)-*th entry of* A (or simply the (i,j)-*entry of* A).

For example, $\begin{pmatrix} 1 & 7 & 2 \\ -\sqrt{2} & 6 & 1/3 \end{pmatrix}_{1,3} = 2$. Note that this notation $A_{i,j}$ (for the (i,j)-th entry of a matrix A) is not standard in literature! Some authors instead

[5] Some authors use the symbol $\mathbb{N}$ for the set $\{1,2,3,\ldots\}$ (the set of all **positive** integers) instead. Unfortunately, there is no consensus here. If you want to avoid notational conflicts, use the notation $\mathbb{Z}_{\geq 0}$ for $\{0,1,2,\ldots\}$ and the notation $\mathbb{Z}_{>0}$ for $\{1,2,3,\ldots\}$. These notations, at least, are self-explanatory (once you know that $\mathbb{Z}$ denotes the set of all integers, i.e., $\{\ldots,-2,-1,0,1,2,\ldots\}$).

[6] For the friends of stupid examples (me?), here are two more: $\begin{pmatrix} \\ \\ \\ \end{pmatrix}$ is a 3×0-matrix (it contains no cells, and thus no numbers), and $\begin{pmatrix} & & \end{pmatrix}$ is a 0×3-matrix (again, with no numbers because it has no cells). For various technical reasons ([deBoor]), it is helpful to regard such empty matrices as different.

use A^i_j (the i on top is not an exponent, but just a superscript) or $a_{i,j}$ (where a is the lowercase letter corresponding to the uppercase letter A denoting the matrix[7]). Many authors often drop the comma between i and j (so they call it A_{ij} or a_{ij}); this notation is slightly ambiguous (does A_{132} mean $A_{13,2}$ or $A_{1,32}$?). Unfortunately, some authors use the notation $A_{i,j}$ for something else called a *cofactor* of A (which is, in a sense, quite the opposite of the (i,j)-th entry of A); but we will never do this here (and probably we will not really get into cofactors anyway).

2.2. The matrix builder notation

I would like to do something interesting, but I am forced to introduce more notations. Please have patience with me. Let me introduce a notation for building a matrix out of a bunch of entries:

Definition 2.5. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Assume that you are given a number $a_{i,j}$ for each pair (i,j) of an integer $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. Then, $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ shall denote the $n \times m$-matrix whose (i,j)-th entry is $a_{i,j}$ for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. (To say it differently: $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ shall denote the $n \times m$-matrix A such that $A_{i,j} = a_{i,j}$ for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. In other words,

$$\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{pmatrix}.$$

)

Some examples:

- What is the matrix $(i-j)_{1\leq i\leq 2,\ 1\leq j\leq 3}$? By definition, it is the 2×3-matrix whose (i,j)-th entry is $i-j$ for all $i \in \{1,2\}$ and $j \in \{1,2,3\}$. Thus, $(i-j)_{1\leq i\leq 2,\ 1\leq j\leq 3} = \begin{pmatrix} 1-1 & 1-2 & 1-3 \\ 2-1 & 2-2 & 2-3 \end{pmatrix} = \begin{pmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \end{pmatrix}$.
- We have $(j-i)_{1\leq i\leq 2,\ 1\leq j\leq 3} = \begin{pmatrix} 1-1 & 2-1 & 3-1 \\ 1-2 & 2-2 & 3-2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \end{pmatrix}$.
- We have $(i+j)_{1\leq i\leq 3,\ 1\leq j\leq 2} = \begin{pmatrix} 1+1 & 1+2 \\ 2+1 & 2+2 \\ 3+1 & 3+2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{pmatrix}$.

[7] This notation is bad for two reasons: First, it forces you to always denote matrices by uppercase letters; second, it doesn't let you write things like $\begin{pmatrix} 1 & 7 & 2 \\ -\sqrt{2} & 6 & 1/3 \end{pmatrix}_{1,3}$.

- We have $\left(\dfrac{i+1}{j}\right)_{1\leq i\leq 3,\ 1\leq j\leq 3} = \begin{pmatrix} \dfrac{1+1}{1} & \dfrac{1+1}{2} & \dfrac{1+1}{3} \\ \dfrac{2+1}{1} & \dfrac{2+1}{2} & \dfrac{2+1}{3} \\ \dfrac{3+1}{1} & \dfrac{3+1}{2} & \dfrac{3+1}{3} \end{pmatrix} = \begin{pmatrix} 2 & 1 & \dfrac{2}{3} \\ 3 & \dfrac{3}{2} & 1 \\ 4 & 2 & \dfrac{4}{3} \end{pmatrix}$.
- We have $\left(\dfrac{i-j}{i+j}\right)_{1\leq i\leq 3,\ 1\leq j\leq 2} = \begin{pmatrix} \dfrac{1-1}{1+1} & \dfrac{1-2}{1+2} \\ \dfrac{2-1}{2+1} & \dfrac{2-2}{2+2} \\ \dfrac{3-1}{3+1} & \dfrac{3-2}{3+2} \end{pmatrix} = \begin{pmatrix} 0 & -\dfrac{1}{3} \\ \dfrac{1}{3} & 0 \\ \dfrac{1}{2} & \dfrac{1}{5} \end{pmatrix}$.

The notation $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$ is fairly standard (you will be understood if you use it), though again there are variations in the literature.

We used the two letters i and j in the notation $\left(a_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m}$, but we could just as well have picked any other two letters (as long as they aren't already taken for something else). For example, $(xy)_{1\leq x\leq 2,\ 1\leq y\leq 2} = \begin{pmatrix} 1\cdot 1 & 1\cdot 2 \\ 2\cdot 1 & 2\cdot 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. For a more confusing example, $(i-j)_{1\leq i\leq 2,\ 1\leq j\leq 1} = \begin{pmatrix} 1-1 \\ 2-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ can be rewritten as $(j-i)_{1\leq j\leq 2,\ 1\leq i\leq 1}$ (we just renamed the letters i and j as j and i here). Do not confuse this with the 1×2-matrix $(j-i)_{1\leq i\leq 1,\ 1\leq j\leq 2} = \begin{pmatrix} 1-1 & 2-1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix}$. The difference between the two matrices $(j-i)_{1\leq j\leq 2,\ 1\leq i\leq 1}$ and $(j-i)_{1\leq i\leq 1,\ 1\leq j\leq 2}$ is the order in which j and i appear in the subscript ("$1\leq j\leq 2,\ 1\leq i\leq 1$" versus "$1\leq i\leq 1,\ 1\leq j\leq 2$"). If j comes first, then j is the number of the row and i the number of the column; but if i comes first, then it's the other way round!

Of course, if you decompose an $n\times m$-matrix A into its entries, and then assemble these entries back into an $n\times m$-matrix (arranged in the same way as in A), then you get back A. In other words: For every $n\times m$-matrix A, we have

$$\left(A_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq m} = A. \tag{1}$$

2.3. Row and column vectors

Here is some more terminology:

Definition 2.6. Let $n\in\mathbb{N}$. A *row vector of size n* means a $1\times n$-matrix. A *column vector of size n* means an $n\times 1$-matrix.

For example, $\begin{pmatrix} a & b \end{pmatrix}$ is a row vector of size 2, while $\begin{pmatrix} a \\ b \end{pmatrix}$ is a column vector of size 2.

The following definition is common-sense:

Definition 2.7. Let $n \in \mathbb{N}$. If v is a row vector of size n, then the $(1,j)$-th entry of v (for $j \in \{1,2,\ldots,n\}$) will also be called the *j-th entry* of v (because v has only one row, so that we don't have to say which row an entry lies in). If v is a column vector of size n, then the $(i,1)$-th entry of v (for $i \in \{1,2,\ldots,n\}$) will also be called the *i-th entry* of v.

2.4. Transposes

Definition 2.8. The *transpose* of an $n \times m$-matrix A is defined to be the $m \times n$-matrix $\left(A_{j,i}\right)_{1 \leq i \leq m,\ 1 \leq j \leq n}$. It is denoted by A^T.

Let us unravel this confusing-looking definition! It says that the transpose of an $n \times m$-matrix A is the $m \times n$-matrix whose (i,j)-th entry (for $i \in \{1,2,\ldots,m\}$ and $j \in \{1,2,\ldots,n\}$) is the (j,i)-th entry of A. So the transpose of A has the very same entries as A, but in different position: namely, the entry in position (i,j) gets moved into position (j,i). In other words, the entry that was in row i and column j gets moved into column i and row j. So, visually speaking, the transpose of the matrix A is obtained by "reflecting A around the diagonal". Some examples should help clarify this:

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix}^T = \begin{pmatrix} a & a' \\ b & b' \\ c & c' \end{pmatrix};$$

$$\begin{pmatrix} a & b \\ a' & b' \end{pmatrix}^T = \begin{pmatrix} a & a' \\ b & b' \end{pmatrix};$$

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}^T = \begin{pmatrix} a & b & c \end{pmatrix}.$$

Transposes have many uses, but for now we stress one particular use: as a space-saving device. Namely, if you work with column vectors, you quickly notice that they take up a lot of vertical space in writing: just see by how much the column vector $\begin{pmatrix} 4 \\ -1 \\ 2 \\ 0 \end{pmatrix}$ has stretched the spacing between its line and the lines above and below[8]! It is much more economical to rewrite it as the transpose of a row vector:

[8] Additionally, column vectors of size 2 have the annoying property that they can get confused for binomial coefficients. To wit, $\begin{pmatrix} 4 \\ 2 \end{pmatrix}$ denotes a column vector, whereas $\binom{4}{2}$ denotes a binomial coefficient (which equals the number 6). The only way to tell them apart is by the amount of empty space between the parentheses and the entries; this is not a very reliable way to keep different notations apart.

$\begin{pmatrix} 4 & -1 & 2 & 0 \end{pmatrix}^T$. It is furthermore common to write row vectors as tuples (i.e., put commas between their entries instead of leaving empty space); thus, the row vector $\begin{pmatrix} 4 & -1 & 2 & 0 \end{pmatrix}$ becomes $(4,-1,2,0)$ (which takes up less space), and our column vector above becomes $(4,-1,2,0)^T$.

The transpose of a matrix A is also denoted by A^t or ${}^TA$ or ${}^tA$ by various authors (not me).

Here is a very simple fact about transposes: The transpose of the transpose of a matrix A is the matrix A itself. In other words:

Proposition 2.9. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Then, $\left(A^T\right)^T = A$.

Proof of Proposition 2.9. This is fairly clear, but let me give a formal proof just to get you used to the notations.

We have $A^T = \left(A_{j,i}\right)_{1 \leq i \leq m,\ 1 \leq j \leq n}$ (by the definition of A^T). Thus, A^T is an $m \times n$-matrix and satisfies

$$\left(A^T\right)_{i,j} = A_{j,i} \qquad \text{for all } i \in \{1,2,\ldots,m\} \text{ and } j \in \{1,2,\ldots,n\}. \tag{2}$$

Hence,

$$\left(A^T\right)_{j,i} = A_{i,j} \qquad \text{for all } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,m\}. \tag{3}$$

(Indeed, this follows by applying (2) to j and i instead of i and j.)

Now, the definition of $\left(A^T\right)^T$ yields

$$\left(A^T\right)^T = \left(\underbrace{\left(A^T\right)_{j,i}}_{\substack{=A_{i,j} \\ \text{(by (3))}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq m} = \left(A_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m} = A$$

(by (1)). This proves Proposition 2.9. □

2.5. Addition, scaling and multiplication

Matrices can (sometimes) be added, (always) be scaled and (sometimes) be multiplied. Let me explain:

Definition 2.10. Let A and B be two matrices of the same dimensions (that is, they have the same number of rows, and the same number of columns). Then, $A + B$ denotes the matrix obtained by adding each entry of A to the corresponding entry of B. Or, to write it more formally: If A and B are two $n \times m$-matrices, then

$$A + B = \left(A_{i,j} + B_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}.$$

For example, $\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} + \begin{pmatrix} a' & b' & c' \\ d' & e' & f' \end{pmatrix} = \begin{pmatrix} a+a' & b+b' & c+c' \\ d+d' & e+e' & f+f' \end{pmatrix}$. (I am increasingly using variables instead of actual numbers in my examples, because they make it easier to see what entry is going where.) On the other hand, the two matrices $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} c & d \end{pmatrix}$ cannot be added (since they have different dimensions[9]).

Definition 2.10 is often laconically summarized as follows: "Matrices are added entry by entry" (or "entrywise"). This simply means that each entry of the sum $A + B$ is the sum of the corresponding entries of A and B; nothing fancy is going on.

So we now know how to add two matrices.

Definition 2.11. Let A be a matrix, and λ be a number. Then, λA (or $\lambda \cdot A$) denotes the matrix obtained by multiplying each entry of A by λ. In other words: If A is an $n \times m$-matrix, then

$$\lambda A = \left(\lambda A_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}.$$

The matrix λA is often called "λ times A". The procedure of transforming A into λA is called *scaling the matrix A by λ*. (Sometimes we say "multiplying" instead of "scaling", but "scaling" is more precise.)

We write $-A$ for $(-1)\, A$.

For example, $\lambda \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b & \lambda c \\ \lambda d & \lambda e & \lambda f \end{pmatrix}$.

So now we know how to scale a matrix. ("To scale" means to multiply by a number.) Definition 2.11 is summarized as follows: "Matrices are scaled entry by entry".

"Scaling" is often called "scalar multiplication" (but this is confusing terminology, since "scalar product" means something completely different). If A is a matrix, then a *scalar multiple* of A is defined as a matrix of the form λA for some number λ.

With scaling and addition defined, we obtain subtraction for free:

Definition 2.12. Let A and B be two matrices of the same dimensions. Then, $A - B$ denotes the matrix $A + (-B) = A + (-1)\, B$.

For example, $\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} - \begin{pmatrix} a' & b' & c' \\ d' & e' & f' \end{pmatrix} = \begin{pmatrix} a-a' & b-b' & c-c' \\ d-d' & e-e' & f-f' \end{pmatrix}$.

Now, to the more interesting part: multiplying matrices. This is **not** done by multiplying corresponding entries! (Why not? Well, it wouldn't make for a particularly useful notion.) Instead, the definition goes as follows:

[9] The dimensions of the former matrix are 2 and 1, whereas the dimensions of the latter matrix are 1 and 2. Even though they are equal **up to order**, they do not count as equal.

Definition 2.13. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let B be an $m \times p$-matrix. (Thus, A has to have m columns, while B has to have m rows; other than this, the two matrices do not need to have any relation to each other.) The product AB of these two matrices is defined as follows:

$$AB = \left(\underbrace{A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}}_{\substack{\text{This is the sum of the } m \text{ terms of the form} \\ A_{i,k}B_{k,j}, \text{ for } k \text{ ranging over } \{1,2,\ldots,m\}}} \right)_{1 \leq i \leq n,\ 1 \leq j \leq p}.$$

This is an $n \times p$-matrix.

This definition is somewhat overwhelming, so let me rewrite it in words and give some examples:

It says that the product AB is well-defined whenever A has as many columns as B has rows. In this case, AB is the $n \times p$-matrix whose (i,j)-th entry is obtained by adding together:

- the product $A_{i,1}B_{1,j}$ of the $(i,1)$-th entry of A with the $(1,j)$-th entry of B;
- the product $A_{i,2}B_{2,j}$ of the $(i,2)$-th entry of A with the $(2,j)$-th entry of B;
- and so on;
- the product $A_{i,m}B_{m,j}$ of the (i,m)-th entry of A with the (m,j)-th entry of B.

In other words, AB is the matrix whose (i,j)-th entry is obtained by multiplying each entry of the i-th row of A with the corresponding entry of the j-th column of B, and then adding together all these products. The word "corresponding" means that the 1-st entry of the i-th row of A gets multiplied with the 1-st entry of the j-th column of B, the 2-nd entry with the 2-nd entry, etc.. In particular, for this to make sense, the i-th row of A and the j-th column of B have to have the same number of entries. This is why we required that A has as many columns as B has rows!

I promised examples. Here are four:

$$\begin{pmatrix} a & b \\ a' & b' \end{pmatrix} \begin{pmatrix} x & x' \\ y & y' \end{pmatrix} = \begin{pmatrix} ax+by & ax'+by' \\ a'x+b'y & a'x'+b'y' \end{pmatrix};$$

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax+by+cz \\ a'x+b'y+c'z \end{pmatrix};$$

$$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \end{pmatrix};$$

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} ax & ay \\ bx & by \end{pmatrix}$$

(note how in the fourth example, we don't see any plus signs, because each of the sums has only one addend).

We can also denote the product AB by $A \cdot B$ (though few people ever do this[10]).

We have thus learnt how to multiply matrices. Notice that the (i,j)-th entry of the product AB depends only on the i-th row of A and the j-th column of B. Why did we pick this strange definition, rather than something simpler, like multiplying entry by entry, or at least row by row? Well, "entry by entry" is too simple (you will see later what matrix multiplication is good for; "entry by entry" is useless in comparison), whereas "row by row" would be lacking many of the nice properties that we will see later (e.g., our matrix multiplication satisfies the associativity law $(AB)\,C = A\,(BC)$, while "row by row" does not).

Exercise 2.14. Let $A = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 1 & 6 \end{pmatrix}$.

(a) The matrix A is of size 3×2. What is the size of B ?

(b) Is AB defined? If it is, compute it.

(c) Is BA defined? If it is, compute it.

Exercise 2.15. (a) Compute

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

and

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(b) Compute $\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ for an arbitrary 3×3-matrix $\begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$.

[10] **Warning:** The notation $A \cdot B$ is somewhat nonstandard. Many authors (for example, Olver and Shakiban in [OlvSha06, §3.1]) define the "dot product" of two column vectors $v = (v_1, v_2, \ldots, v_n)^T$ and $w = (w_1, w_2, \ldots, w_n)^T$ (of the same size) to be the number $v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$; they furthermore denote this dot product by $v \cdot w$. This notation is in conflict with our notation $A \cdot B$, because the dot product of v and w is not what we call $v \cdot w$ (it is, in fact, what we call $v^T \cdot w$). The reason why I have picked the somewhat nonstandard convention to regard $A \cdot B$ as a synonym for AB is my belief that a dot should always denote the same multiplication as juxtaposition (i.e., that $A \cdot B$ should always mean the same as AB).

(c) Compute $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ for an arbitrary 4×1-matrix $\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$.

Exercise 2.16. **(a)** Let $A_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ and $B_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Compute A_3^2, B_3^2, A_3B_3 and B_3A_3.

(b) Let $A_4 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$ and $B_4 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$. Compute A_4^2, B_4^2, A_4B_4 and B_4A_4.

(c) For any $n \in \mathbb{N}$, define two "checkerboard-pattern" $n \times n$-matrices A_n and B_n by

$$A_n = ((i+j)\,\%2)_{1 \leq i \leq n,\ 1 \leq j \leq n}, \qquad B_n = ((i+j-1)\,\%2)_{1 \leq i \leq n,\ 1 \leq j \leq n},$$

where $k\%2$ denotes the remainder left when k is divided by 2 (so $k\%2 = \begin{cases} 1, & \text{if } k \text{ is odd;} \\ 0, & \text{if } k \text{ is even} \end{cases}$). (The matrices A_3 and B_3 in part **(a)** of this problem, as well as the matrices A_4 and B_4 in its part **(b)**, are particular cases of this construction.) Prove that each **even** $n \in \mathbb{N}$ satisfies $A_n^2 = B_n^2$ and $A_nB_n = B_nA_n$. Prove that each **odd** $n \geq 3$ satisfies $A_nB_n \neq B_nA_n$.

2.6. The matrix product rewritten

Let me show another way to restate our above definition of a product of two matrices. First, one more notation:

Definition 2.17. Let A be an $n \times m$-matrix.

(a) If $i \in \{1, 2, \ldots, n\}$, then $\operatorname{row}_i A$ will denote the i-th row of A. This is a row vector of size m (that is, a $1 \times m$-matrix), and is formally defined as

$$\left(A_{i,y}\right)_{1 \leq x \leq 1,\ 1 \leq y \leq m} = \begin{pmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,m} \end{pmatrix}$$

(notice how i is kept fixed but y is ranging from 1 to m here).

(b) If $j \in \{1, 2, \ldots, m\}$, then $\operatorname{col}_j A$ will denote the j-th column of A. This is a column vector of size n (that is, an $n \times 1$-matrix), and is formally defined as

$$\left(A_{x,j}\right)_{1 \leq x \leq n,\ 1 \leq y \leq 1} = \begin{pmatrix} A_{1,j} \\ A_{2,j} \\ \vdots \\ A_{n,j} \end{pmatrix}.$$

Example 2.18. If $A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$, then $\operatorname{row}_2 A = \begin{pmatrix} d & e & f \end{pmatrix}$ and $\operatorname{col}_2 A = \begin{pmatrix} b \\ e \end{pmatrix}$.

Now, we observe that if R is a row vector of some size m, and if C is a column vector of size m, then RC is a 1×1-matrix. More precisely: The product of a row vector $\begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix}$ and a column vector $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$ is given by

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} r_1 c_1 + r_2 c_2 + \cdots + r_m c_m \end{pmatrix}. \tag{4}$$

We shall often equate a 1×1-matrix with its (unique) entry; so the equality (4) rewrites as

$$\begin{pmatrix} r_1 & r_2 & \cdots & r_m \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = r_1 c_1 + r_2 c_2 + \cdots + r_m c_m. \tag{5}$$

Now I will show a little collection of formulas for the product of two matrices. They are all pretty straightforward to obtain (essentially, they are the definition of the product viewed from different angles), but they are helpful when it comes to manipulating products:

Proposition 2.19. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let B be an $m \times p$-matrix.

(a) For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, p\}$, we have

$$(AB)_{i,j} = A_{i,1} B_{1,j} + A_{i,2} B_{2,j} + \cdots + A_{i,m} B_{m,j}.$$

(b) For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, p\}$, the (i,j)-th entry of AB equals the product of the i-th row of A and the j-th column of B. In formulas:

$$(AB)_{i,j} = \operatorname{row}_i A \cdot \operatorname{col}_j B \tag{6}$$

for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$ (where the expression $\operatorname{row}_i A \cdot \operatorname{col}_j B$ should be read as $(\operatorname{row}_i A) \cdot (\operatorname{col}_j B)$). Thus,

$$\begin{aligned} AB &= \left(\operatorname{row}_i A \cdot \operatorname{col}_j B\right)_{1 \leq i \leq n,\ 1 \leq j \leq p} \\ &= \begin{pmatrix} \operatorname{row}_1 A \cdot \operatorname{col}_1 B & \operatorname{row}_1 A \cdot \operatorname{col}_2 B & \cdots & \operatorname{row}_1 A \cdot \operatorname{col}_p B \\ \operatorname{row}_2 A \cdot \operatorname{col}_1 B & \operatorname{row}_2 A \cdot \operatorname{col}_2 B & \cdots & \operatorname{row}_2 A \cdot \operatorname{col}_p B \\ \vdots & \vdots & \ddots & \vdots \\ \operatorname{row}_n A \cdot \operatorname{col}_1 B & \operatorname{row}_n A \cdot \operatorname{col}_2 B & \cdots & \operatorname{row}_n A \cdot \operatorname{col}_p B \end{pmatrix}. \end{aligned}$$

(c) For every $i \in \{1,2,\ldots,n\}$, we have

$$\operatorname{row}_i (AB) = (\operatorname{row}_i A) \cdot B.$$

(d) For every $j \in \{1,2,\ldots,p\}$, we have

$$\operatorname{col}_j (AB) = A \cdot \operatorname{col}_j B.$$

Proposition 2.19 **(c)** says that if A and B are two matrices (for which AB makes sense), then each row of AB equals the corresponding row of A multiplied by B. Similarly, Proposition 2.19 **(d)** says that each column of AB equals A multiplied by the corresponding column of B. These are fairly simple observations, but they are surprisingly useful.

Proof of Proposition 2.19. **(a)** By the definition of AB, we have

$$AB = \left(A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq p}.$$

In other words,

$$(AB)_{i,j} = A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j} \tag{7}$$

for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$. This proves Proposition 2.19 **(a)**.

(b) Now, let us prove Proposition 2.19 **(b)**. It is clearly enough to prove (6) (because all the other statements of Proposition 2.19 **(b)** are just restatements of (6)). So let's do this. Let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$. Then,

$$\operatorname{row}_i A = \begin{pmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,m} \end{pmatrix} \qquad \text{and} \tag{8}$$

$$\operatorname{col}_j B = \begin{pmatrix} B_{1,j} \\ B_{2,j} \\ \vdots \\ B_{m,j} \end{pmatrix}. \tag{9}$$

Hence,

$$\begin{aligned}\operatorname{row}_i A \cdot \operatorname{col}_j B &= \begin{pmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,m} \end{pmatrix} \begin{pmatrix} B_{1,j} \\ B_{2,j} \\ \vdots \\ B_{m,j} \end{pmatrix} \\ &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}.\end{aligned}$$

Comparing this with (7), we obtain $(AB)_{i,j} = \operatorname{row}_i A \cdot \operatorname{col}_j B$. Thus, we have proven (6). Hence, Proposition 2.19 **(b)** is proven.

(c) Let $i \in \{1, 2, \ldots, n\}$. Set $C = \operatorname{row}_i A$. Notice that C is a row vector of size m, thus a $1 \times m$-matrix. We can refer to any given entry of C either as "the j-th entry" or as "the $(1, j)$-th entry" (where j is the number of the column the entry is located in).

We have

$$C = \operatorname{row}_i A = \begin{pmatrix} A_{i,1} & A_{i,2} & \cdots & A_{i,m} \end{pmatrix}.$$

Thus,

$$C_{1,k} = A_{i,k} \qquad \text{for every } k \in \{1, 2, \ldots, m\}. \tag{10}$$

Let $j \in \{1, 2, \ldots, p\}$. Then,

$$\begin{aligned}&(\text{the } j\text{-th entry of } \operatorname{row}_i (AB)) \\ &= (\text{the } (i, j)\text{-th entry of } AB) = (AB)_{i,j} \\ &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j} \qquad (\text{by (7)}).\end{aligned}$$

Comparing this with

$$\begin{aligned}&(\text{the } j\text{-th entry of } CB) \\ &= (\text{the } (1, j)\text{-th entry of } CB) \qquad (\text{since } CB \text{ is a row vector}) \\ &= \underbrace{C_{1,1}}_{\substack{=A_{i,1} \\ (\text{by (10)})}} B_{1,j} + \underbrace{C_{1,2}}_{\substack{=A_{i,2} \\ (\text{by (10)})}} B_{2,j} + \cdots + \underbrace{C_{1,m}}_{\substack{=A_{i,m} \\ (\text{by (10)})}} B_{m,j} \\ &\qquad \left(\begin{array}{c} \text{by Proposition 2.19 } \textbf{(a)}\text{, applied to 1, } C \text{ and 1} \\ \text{instead of } n, A \text{ and } i \end{array} \right) \\ &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j},\end{aligned}$$

we obtain

$$(\text{the } j\text{-th entry of } \operatorname{row}_i (AB)) = (\text{the } j\text{-th entry of } CB). \tag{11}$$

Now, forget that we fixed j. We thus have shown that (11) holds for each $j \in \{1, 2, \ldots, p\}$. In other words, each entry of the row vector $\operatorname{row}_i (AB)$ equals the

corresponding entry of the row vector CB. Hence, $\operatorname{row}_i(AB)$ equals CB. Thus, $\operatorname{row}_i(AB) = \underbrace{C}_{=\operatorname{row}_i A} B = (\operatorname{row}_i A) \cdot B$. This proves Proposition 2.19 **(c)**.

(d) The proof of Proposition 2.19 **(d)** is similar to that of Proposition 2.19 **(c)**. Let me nevertheless show it, for the sake of completeness. (The proof below is essentially a copy-pasted version of the above proof of Proposition 2.19 **(c)**, with only the necessary changes made. This is both practical for me, as it saves me some work, and hopefully helpful for you, as it highlights the similarities.)

Let $j \in \{1, 2, \ldots, p\}$. Set $D = \operatorname{col}_j B$. Notice that D is a column vector of size m, thus an $m \times 1$-matrix. We can refer to any given entry of D either as "the i-th entry" or as "the $(i, 1)$-th entry" (where i is the number of the row the entry is located in).

We have

$$D = \operatorname{col}_j B = \begin{pmatrix} B_{1,j} \\ B_{2,j} \\ \vdots \\ B_{m,j} \end{pmatrix}.$$

Thus,

$$D_{k,1} = B_{k,j} \qquad \text{for every } k \in \{1, 2, \ldots, m\}. \tag{12}$$

Let $i \in \{1, 2, \ldots, n\}$. Then,

$$\begin{aligned} &\left(\text{the } i\text{-th entry of } \operatorname{col}_j(AB)\right) \\ &= (\text{the } (i,j)\text{-th entry of } AB) = (AB)_{i,j} \\ &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j} \qquad \text{(by (7))}. \end{aligned}$$

Comparing this with

$$\begin{aligned} &(\text{the } i\text{-th entry of } AD) \\ &= (\text{the } (i,1)\text{-th entry of } AD) \qquad (\text{since } AD \text{ is a column vector}) \\ &= A_{i,1} \underbrace{D_{1,1}}_{\substack{=B_{1,j} \\ \text{(by (12))}}} + A_{i,2} \underbrace{D_{2,1}}_{\substack{=B_{2,j} \\ \text{(by (12))}}} + \cdots + A_{i,m} \underbrace{D_{m,1}}_{\substack{=B_{m,j} \\ \text{(by (12))}}} \\ &\qquad \left(\begin{array}{c} \text{by Proposition 2.19 } \textbf{(a)}\text{, applied to 1, } D \text{ and 1} \\ \text{instead of } p, B \text{ and } j \end{array}\right) \\ &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}, \end{aligned}$$

we obtain

$$\left(\text{the } i\text{-th entry of } \operatorname{col}_j(AB)\right) = (\text{the } i\text{-th entry of } AD). \tag{13}$$

Now, forget that we fixed i. We thus have shown that (13) holds for each $i \in \{1, 2, \ldots, n\}$. In other words, each entry of the column vector $\operatorname{col}_j(AB)$ equals the corresponding entry of the column vector AD. Hence, $\operatorname{col}_j(AB)$ equals AD. Thus, $\operatorname{col}_j(AB) = A \underbrace{D}_{=\operatorname{col}_j B} = A \cdot \operatorname{col}_j B$. This proves Proposition 2.19 **(d)**. □

2.7. Properties of matrix operations

The operations of adding, scaling and multiplying matrices, in many aspects, "behave almost as nicely as numbers". Specifically, I mean that they satisfy a bunch of laws that numbers satisfy:

Proposition 2.20. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then:

(a) We have $A + B = B + A$ for any two $n \times m$-matrices A and B. (This is called "commutativity of addition".)

(b) We have $A + (B + C) = (A + B) + C$ for any three $n \times m$-matrices A, B and C. (This is called "associativity of addition".)

(c_1) We have $\lambda (A + B) = \lambda A + \lambda B$ for any number λ and any two $n \times m$-matrices A and B.

(c_2) We have $\lambda (\mu A) = (\lambda \mu) A$ and $(\lambda + \mu) A = \lambda A + \mu A$ for any numbers λ and μ and any $n \times m$-matrix A.

(c_3) We have $1A = A$ for any $n \times m$-matrix A.

Let furthermore $p \in \mathbb{N}$. Then:

(d) We have $A (B + C) = AB + AC$ for any $n \times m$-matrix A and any two $m \times p$-matrices B and C. (This is called "left distributivity".)

(e) We have $(A + B) C = AC + BC$ for any two $n \times m$-matrices A and B and any $m \times p$-matrix C. (This is called "right distributivity".)

(f) We have $\lambda (AB) = (\lambda A) B = A (\lambda B)$ for any number λ, any $n \times m$-matrix A and any $m \times p$-matrix B.

Finally, let $q \in \mathbb{N}$. Then:

(g) We have $A (BC) = (AB) C$ for any $n \times m$-matrix A, any $m \times p$-matrix B and any $p \times q$-matrix C. (This is called "associativity of multiplication".)

Example 2.21. Most parts of Proposition 2.20 are fairly easy to visualize and to prove. Let me give an example for the least obvious one: part **(g)**.

Part **(g)** essentially says that $A (BC) = (AB) C$ holds for any three matrices A, B and C for which the products AB and BC are well-defined (i.e., A has as many columns as B has rows, and B has as many columns as C has rows). For example, take $n = 1$, $m = 3$, $p = 2$ and $q = 3$. Set

$$A = \begin{pmatrix} a & b & c \end{pmatrix}, \qquad B = \begin{pmatrix} d & d' \\ e & e' \\ f & f' \end{pmatrix}, \qquad C = \begin{pmatrix} x & y & z \\ x' & y' & z' \end{pmatrix}.$$

Then,

$$AB = \begin{pmatrix} ad + be + cf & ad' + be' + cf' \end{pmatrix}$$

and thus

$$(AB)\,C = \begin{pmatrix} ad+be+cf & ad'+be'+cf' \end{pmatrix} \begin{pmatrix} x & y & z \\ x' & y' & z' \end{pmatrix}$$
$$= \begin{pmatrix} ad'x'+be'x'+cf'x'+bex+adx+cfx \\ ad'y'+be'y'+cf'y'+bey+ady+cfy \\ ad'z'+be'z'+cf'z'+bez+adz+cfz \end{pmatrix}^T$$

after some computation. (Here, we have written the result as a transpose of a column vector, because if we had written it as a row vector, it would not fit on this page.) But

$$BC = \begin{pmatrix} d & d' \\ e & e' \\ f & f' \end{pmatrix} \begin{pmatrix} x & y & z \\ x' & y' & z' \end{pmatrix} = \begin{pmatrix} d'x'+dx & d'y'+dy & d'z'+dz \\ e'x'+ex & e'y'+ey & e'z'+ez \\ f'x'+fx & f'y'+fy & f'z'+fz \end{pmatrix}$$

and as before

$$A\,(BC) = \begin{pmatrix} ad'x'+be'x'+cf'x'+bex+adx+cfx \\ ad'y'+be'y'+cf'y'+bey+ady+cfy \\ ad'z'+be'z'+cf'z'+bez+adz+cfz \end{pmatrix}^T.$$

Hence, $(AB)\,C = A\,(BC)$. Thus, our example confirms Proposition 2.20 **(g)**.

The laws of Proposition 2.20 allow you to do many formal manipulations with matrices similarly to how you are used to work with numbers. For example, if you have n matrices $A_1, A_2, \ldots, A_n$ such that successive matrices can be multiplied (i.e., for each $i \in \{1, 2, \ldots, n-1\}$, the matrix A_i has as many columns as A_{i+1} has rows), then the product $A_1 A_2 \cdots A_n$ is well-defined: you can parenthesize it in any order, and the result will always be the same. For example, the product $ABCD$ of four matrices A, B, C, D can be computed in any of the five ways

$$((AB)\,C)\,D, \qquad (AB)\,(CD), \qquad (A\,(BC))\,D, \qquad A\,((BC)\,D), \qquad A\,(B\,(CD)),$$

and all of them lead to the same result. This is called *general associativity* and is not obvious (even if you know that Proposition 2.20 **(g)** holds)[11]. Let me state this result again as a proposition, just to stress its importance:

Proposition 2.22. Let $A_1, A_2, \ldots, A_n$ be n matrices. Assume that, for each $i \in \{1, 2, \ldots, n-1\}$, the number of columns of A_i equals the number of rows of A_{i+1}

[11] If you are curious about the proofs:

We shall prove Proposition 2.20 **(g)** further below (in Section 2.9). General associativity can be derived from Proposition 2.20 **(g)** in the general context of "binary operations"; see (for example) [Zuker14] for this argument.

(so that the product A_iA_{i+1} makes sense). Then, the product $A_1A_2\cdots A_n$ is well-defined: Any way to compute this product (by parenthesizing it) yields the same result. In particular, it can be computed both as $A_1\left(A_2\left(A_3\left(\cdots\left(A_{n-1}A_n\right)\right)\right)\right)$ and as $\left(\left(\left(\left(A_1A_2\right)A_3\right)\cdots\right)A_{n-1}\right)A_n$.

Please take a moment to appreciate general associativity! Without it, we could not make sense of products like ABC and $ABCDE$, because their values could depend on how we choose to compute them. This is one reason why, in the definition of AB, we multiply entries of the i-th row of A with entries of the j-th column of B. Using rows both times would break associativity![12]

There is also a general associativity law for addition:

Proposition 2.23. Let $A_1, A_2, \ldots, A_n$ be n matrices of the same size. Then, the sum $A_1 + A_2 + \cdots + A_n$ is well-defined: Any way to compute this sum (by parenthesizing it) yields the same result. In particular, it can be computed both as $A_1 + \left(A_2 + \left(A_3 + \left(\cdots + \left(A_{n-1} + A_n\right)\right)\right)\right)$ and as $\left(\left(\left(\left(A_1 + A_2\right) + A_3\right) + \cdots\right) + A_{n-1}\right) + A_n$.

There is also another variant of general associativity that concerns the interplay of matrix multiplication and scaling. It claims that products of matrices and numbers can be parenthesized in any order. For example, the product $\lambda\mu AB$ of two numbers λ and μ and two matrices A and B can be computed in any of the five ways

$$\left(\left(\lambda\mu\right)A\right)B, \qquad \left(\lambda\mu\right)\left(AB\right), \qquad \left(\lambda\left(\mu A\right)\right)B, \qquad \lambda\left(\left(\mu A\right)B\right), \qquad \lambda\left(\mu\left(AB\right)\right),$$

and all of them lead to the same result. This can be deduced from parts **(c_2)**, **(f)** and **(g)** of Proposition 2.20.

We shall give proofs of parts **(d)** and **(g)** of Proposition 2.20 in Section 2.9 below.

Various other identities follow from Proposition 2.20. For example, if A, B and C are three matrices of the same size, then $A - \left(B + C\right) = A - B - C$. For another example, if A and B are two $n \times m$-matrices (for some $n \in \mathbb{N}$ and $m \in \mathbb{N}$) and if C is an $m \times p$-matrix (for some $p \in \mathbb{N}$), then $\left(A - B\right)C = AC - BC$. These identities are proven similarly as the analogous properties of numbers are proven; we shall not linger on them.

[12] Of course, our formulation of general associativity was far from rigorous. After all, we have not defined what a "way to compute a product" means, or what "parenthesizing a product" means. There are several ways to make Proposition 2.22 rigorous. See [m.se709196] for a discussion of such ways. (Note that the simplest way actually avoids defining "parenthesizing". Instead, it defines the product $A_1A_2\cdots A_n$ by recursion on n, namely defining it to be A_1 when $n = 1$, and defining it to be $\left(A_1A_2\cdots A_{n-1}\right)A_n$ otherwise (where we are using the already-defined product $A_1A_2\cdots A_{n-1}$). Informally speaking, this means that the product $A_1A_2\cdots A_n$ is defined as $\left(\left(\left(\left(A_1A_2\right)A_3\right)\cdots\right)A_{n-1}\right)A_n$. Now, general associativity says that this product $A_1A_2\cdots A_n$ equals $\left(A_1A_2\cdots A_k\right)\left(A_{k+1}A_{k+2}\cdots A_n\right)$ for each $k \in \{1, 2, \ldots, n-1\}$. (This is not too hard to prove by induction over n.) Informally speaking, this shows that our product $A_1A_2\cdots A_n$ also equals the result of any way of computing it (not only the $\left(\left(\left(\left(A_1A_2\right)A_3\right)\cdots\right)A_{n-1}\right)A_n$ way).)

2.8. Non-properties of matrix operations

Conspicuously absent from Proposition 2.20 is one important law that is well-known to hold for numbers: commutativity of multiplication (that is, $ab = ba$). This has a reason: it is false for matrices. There are at least three reasons why it is false:

1. If A and B are matrices, then it can happen that AB is well-defined (i.e., A has as many columns as B has rows) but BA is not (i.e., B does not have as many columns as A has rows). For example, if $A = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $B = \begin{pmatrix} x & y \end{pmatrix}$, then AB is well-defined but BA is not.

2. If A and B are matrices such that both AB and BA well-defined, then AB and BA might still have different dimensions. Namely, if A is an $n \times m$-matrix and B is an $m \times n$-matrix, then AB is an $n \times n$-matrix, but BA is an $m \times m$-matrix. So comparing AB and BA makes no sense unless $n = m$.

3. Even if AB and BA are of the same dimensions, they can still be distinct. For example, if $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $B = A^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, then $AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ whereas $BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

Two matrices A and B are said to *commute* if $AB = BA$ (which, in particular, means that both AB and BA are well-defined). You will encounter many cases when matrices A and B happen to commute (for example, every $n \times n$-matrix commutes with the $n \times n$ identity matrix; see below for what this means); but in general there is no reason to expect two randomly chosen matrices to commute.

As a consequence of matrices refusing to commute (in general), we cannot reasonably define division of matrices. Actually, there are two reasons why we cannot reasonably define division of matrices: First, if A and B are two matrices, then it is not clear whether $\frac{A}{B}$ should mean a matrix C satisfying $BC = A$, or a matrix C satisfying $CB = A$. (The failure of commutativity implies that these are two different things.) Second, in general, neither of these matrices C is necessarily unique; nor is it guaranteed to exist. This is similar to the fact that we cannot divide by 0 (in fact, $\frac{0}{0}$ would not be unique, while $\frac{1}{0}$ would not exist); but with matrices, 0 is not the only forbidden denominator. Here is an example:

Example 2.24. (a) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = A$. Then, $BC = A$ holds for $C = A$, but also for $C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (and also for many other matrices C). So the matrix C satisfying $BC = A$ is not unique.

(b) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Then, there exists no matrix C satisfying $BC = A$. Indeed, if $C = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ is any matrix, then $BC = \begin{pmatrix} x & y \\ 0 & 0 \end{pmatrix}$ has its second row filled with zeroes, but A does not; so BC cannot equal A.

Exercise 2.25. **(a)** Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Show that the 2×2-matrices B satisfying $AB = BA$ are precisely the matrices of the form $\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$ (where a and d are any numbers). [**Hint:** Set $B = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$, and rewrite $AB = BA$ as a system of linear equations in x, y, z, w. Solve this system.]

(b) Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Characterize the 2×2-matrices B satisfying $AB = BA$.

2.9. (*) The summation sign, and a proof of $(AB)C = A(BC)$

We now take a break from studying matrices to introduce an important symbol: the summation sign ($\sum$). This sign is one of the hallmarks of abstract mathematics (and also computer science), and helps manipulate matrices comfortably. Here is a quick (but informal) definition of the summation sign $\sum$:

Definition 2.26. Let p and q be two integers such that $p \leq q+1$. Let $a_p, a_{p+1}, \ldots, a_q$ be some numbers. Then, $\sum_{k=p}^{q} a_k$ means the sum $a_p + a_{p+1} + \cdots + a_q$. The symbol $\sum$ is called the summation sign; we pronounce the expression $\sum_{k=p}^{q} a_k$ as "sum of a_k for all k ranging from p to q".

This definition needs some clarifications; but before I give them, let me show some examples:

- We have $\sum_{k=p}^{q} k = p + (p+1) + \cdots + q$. For example, $\sum_{k=1}^{n} k = 1 + 2 + \cdots + n$. (A well-known formula says that this sum $\sum_{k=1}^{n} k = 1 + 2 + \cdots + n$ equals $\frac{n(n+1)}{2}$. For a concrete example, $\sum_{k=1}^{3} k = 1 + 2 + 3 = \frac{3(3+1)}{2} = 6$.) For another example, $\sum_{k=-n}^{n} k = (-n) + (-n+1) + \cdots + n$. (This latter sum equals 0, because it contains, for each its addend, also its negative[13].)

[13] except for the addend 0, but this 0 doesn't change the sum anyway

- We have $\sum\limits_{k=p}^{q} k^2 = p^2 + (p+1)^2 + \cdots + q^2$. For example, $\sum\limits_{k=1}^{n} k^2 = 1^2 + 2^2 + \cdots + n^2$. (A well-known formula says that this sum $\sum\limits_{k=1}^{n} k^2 = 1^2 + 2^2 + \cdots + n^2$ equals $\dfrac{n(n+1)(2n+1)}{6}$.)

- We have $\sum\limits_{k=p}^{q} 1 = \underbrace{1+1+\cdots+1}_{q-p+1 \text{ times}} = q - p + 1$. This illustrates the fact that the a_k in a sum $\sum\limits_{k=p}^{q} a_k$ needs not depend on k (although the cases where it does not depend on k are fairly trivial).

- If $p = q$, then $\sum\limits_{k=p}^{q} a_k = a_p$. (A sum of only one number is simply this number.)

As I have said, a few remarks and clarifications on the summation sign are in order:

Definition 2.27. (a) In the expression $\sum\limits_{k=p}^{q} a_k$ (as defined in Definition 2.26), the letter k is called the *summation index*. It stands for the "moving part" in the sum (e.g., the part in which the addends differ). For example, $\sum\limits_{k=p}^{q} \dfrac{1}{k+3}$ is the sum of the fractions $\dfrac{1}{k+3}$ for k ranging from p to q; its addends all have the form $\dfrac{1}{k+3}$, but for different values of k.

The summation index doesn't have to be called k; any letter is legitimate (as long as it is not already used otherwise). For example, $\sum\limits_{i=p}^{q} a_i$ and $\sum\limits_{x=p}^{q} a_x$ are two synonymous ways to write $\sum\limits_{k=p}^{q} a_k$. Just make sure that you are using the same letter under the $\sum$ sign and to its right (so you should not write $\sum\limits_{i=p}^{q} a_k$, unless you mean the sum $\underbrace{a_k + a_k + \cdots + a_k}_{q-p+1 \text{ times}}$).

(b) You might be wondering what Definition 2.26 means in the case when $p = q + 1$; after all, in this case, there are no numbers $a_p, a_{p+1}, \ldots, a_q$, and the sum $a_p + a_{p+1} + \cdots + a_q$ has no addends. (For example, how should $\sum\limits_{k=2}^{1} k = 2 + 3 + \cdots + 1$ be understood?) However, there is a general convention in mathematics that a sum with no addends is always defined to be 0, and is called an *empty sum*.

Thus, $\sum\limits_{k=p}^{q} a_k = 0$ whenever $p = q+1$. For example, $\sum\limits_{k=2}^{1} k = 0$ and $\sum\limits_{k=0}^{1} \dfrac{1}{k} = 0$ (even though $\dfrac{1}{k}$ makes no sense for $k = 0$). (This convention might sound arbitrary, but is logically adequate: In fact, it ensures that the equality $a_p + a_{p+1} + \cdots + a_q = \left(a_p + a_{p+1} + \cdots + a_{q-1}\right) + a_q$ holds not only for $q > p$, but also for $q = p$.)

Many authors define the sum $\sum\limits_{k=p}^{q} a_k$ to be 0 in the case when $p > q+1$ as well; thus, the sum $\sum\limits_{k=p}^{q} a_k$ is defined for **any** two integers p and q (without the requirement that $p \leq q+1$). However, this convention is somewhat slippery: for instance, it entails $\sum\limits_{k=1}^{n} k = 0$ for all negative k, and thus the equality $\sum\limits_{k=1}^{n} k = \dfrac{n(n+1)}{2}$ does **not** hold for negative n.

(c) From a fully rigorous point of view, Definition 2.26 did not define $\sum\limits_{k=p}^{q} a_k$ at all. Indeed, it defined $\sum\limits_{k=p}^{q} a_k$ to be $a_p + a_{p+1} + \cdots + a_q$, but what does $a_p + a_{p+1} + \cdots + a_q$ mean? The rigorous way to define $\sum\limits_{k=p}^{q} a_k$ is as follows (by recursion):

- If $q = p-1$, then $\sum\limits_{k=p}^{q} a_k$ is defined to be 0.
- If $q > p-1$, then $\sum\limits_{k=p}^{q} a_k$ is defined to be $\left(\sum\limits_{k=p}^{q-1} a_k\right) + a_q$.

This is a recursive definition (since it defines $\sum\limits_{k=p}^{q} a_k$ in terms of $\sum\limits_{k=p}^{q-1} a_k$), and provides an algorithm to compute $\sum\limits_{k=p}^{q} a_k$. From a formal point of view, "$a_p + a_{p+1} + \cdots + a_q$" is just a colloquial way to say "$\sum\limits_{k=p}^{q} a_k$".

Notice that the expression "$\sum\limits_{k=p}^{q} a_k$" is both a more compact and a more rigorous way to say "$a_p + a_{p+1} + \cdots + a_q$". A computer would not understand the expression "$a_p + a_{p+1} + \cdots + a_q$" (it could only guess what the "$\cdots$" means, and computers are bad at guessing); but the expression "$\sum\limits_{k=p}^{q} a_k$" has a well-defined meaning that

can be rigorously defined and can be explained to a computer[14]. Thus, if you want to tell a computer to compute a sum, the command you have to use will be closer to "$\sum_{k=p}^{q} a_k$" than to "$a_p + a_{p+1} + \cdots + a_q$". For example, in Python, you would have to write "`sum(a[k] for k in range(p, q+1))`" (where "`a[k]`" is understood to return a_k) [15].

Using the summation sign, we can rewrite the product AB of two matrices A and B (see Definition 2.13) more nicely:

Proposition 2.28. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let B be an $m \times p$-matrix. Then,

$$(AB)_{i,j} = \sum_{k=1}^{m} A_{i,k} B_{k,j} \qquad \text{for all } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,p\}.$$

Proof of Proposition 2.28. For all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$, we have

$$\begin{aligned}(AB)_{i,j} &= A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j} \qquad \text{(by Proposition 2.19 \textbf{(a)})} \\ &= \sum_{k=1}^{m} A_{i,k} B_{k,j}\end{aligned}$$

(because $\sum_{k=1}^{m} A_{i,k}B_{k,j}$ is exactly $A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}$, by its definition). Proposition 2.28 is proven. □

Here are two properties of sums that are fairly clear if you understand how sums are defined:

Proposition 2.29. Let p and q be two integers such that $p \leq q+1$. Let $a_p, a_{p+1}, \ldots, a_q$ be some numbers. Let b be a number. Then,

$$\sum_{k=p}^{q} b a_k = b \sum_{k=p}^{q} a_k.$$

(The expression $\sum_{k=p}^{q} b a_k$ has to be read as $\sum_{k=p}^{q} (b a_k)$.)

[14] See Definition 2.27 **(c)** for the rigorous definition of $\sum_{k=p}^{q} a_k$.

[15] Why "`q+1`" and not "`q`"? Because Python defines `range(u, v)` as the list $(u, u+1, \ldots, v-1)$ (that is, the list that starts at u and ends **just before** v). So `range(p, q)` would be $(p, p+1, \ldots, q-1)$, but we want $(p, p+1, \ldots, q)$.

Proof of Proposition 2.29. By the definition of $\sum$, we have

$$\sum_{k=p}^{q} ba_k = ba_p + ba_{p+1} + \cdots + ba_q = b\underbrace{\left(a_p + a_{p+1} + \cdots + a_q\right)}_{\substack{=\sum_{k=p}^{q} a_k \\ \text{(by the definition of } \sum_{k=p}^{q} a_k\text{)}}} = b\sum_{k=p}^{q} a_k.$$

□

Proposition 2.30. Let p and q be two integers such that $p \leq q+1$. Let $a_p, a_{p+1}, \ldots, a_q$ be some numbers. Let $b_p, b_{p+1}, \ldots, b_q$ be some numbers. Then,

$$\sum_{k=p}^{q} (a_k + b_k) = \sum_{k=p}^{q} a_k + \sum_{k=p}^{q} b_k.$$

(The expression $\sum_{k=p}^{q} a_k + \sum_{k=p}^{q} b_k$ has to be read as $\left(\sum_{k=p}^{q} a_k\right) + \left(\sum_{k=p}^{q} b_k\right)$.)

Proof of Proposition 2.30. By the definition of $\sum$, we have

$$\begin{aligned}
\sum_{k=p}^{q} (a_k + b_k) &= \left(a_p + b_p\right) + \left(a_{p+1} + b_{p+1}\right) + \cdots + \left(a_q + b_q\right) \\
&= \underbrace{\left(a_p + a_{p+1} + \cdots + a_q\right)}_{\substack{=\sum_{k=p}^{q} a_k \\ \text{(by the definition of } \sum_{k=p}^{q} a_k\text{)}}} + \underbrace{\left(b_p + b_{p+1} + \cdots + b_q\right)}_{\substack{=\sum_{k=p}^{q} b_k \\ \text{(by the definition of } \sum_{k=p}^{q} b_k\text{)}}} \\
&= \sum_{k=p}^{q} a_k + \sum_{k=p}^{q} b_k.
\end{aligned}$$

□

Our goal in this section is to prove Proposition 2.20 **(g)**, illustrating the use and manipulation of the $\sum$ sign. However, as a warmup, let us first prove Proposition 2.20 **(d)** (which is simple enough that you can easily check it without $\sum$ signs, but is nevertheless worth proving using the $\sum$ sign just to demonstrate how to work with the $\sum$ sign):

*Proof of Proposition 2.20 **(d)**.* Let A be an $n \times m$-matrix. Let B and C be two $m \times p$-matrices.

We shall show that $(A(B+C))_{i,j} = (AB + AC)_{i,j}$ for all $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, p\}$. Once this is proven, this will entail that corresponding entries

of the two $n \times p$-matrices $A(B+C)$ and $AB+AC$ are equal; and thus, these two matrices have to be equal.

Proposition 2.28 yields

$$(AB)_{i,j} = \sum_{k=1}^{m} A_{i,k} B_{k,j} \tag{14}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$.

Proposition 2.28 (applied to C instead of B) yields

$$(AC)_{i,j} = \sum_{k=1}^{m} A_{i,k} C_{k,j} \tag{15}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$.

Finally, Proposition 2.28 (applied to $B+C$ instead of B) yields

$$\begin{aligned}
(A(B+C))_{i,j} &= \sum_{k=1}^{m} A_{i,k} \underbrace{(B+C)_{k,j}}_{\substack{=B_{k,j}+C_{k,j} \\ \text{(since matrices are} \\ \text{added entry by entry)}}} = \sum_{k=1}^{m} \underbrace{A_{i,k}\left(B_{k,j}+C_{k,j}\right)}_{=A_{i,k}B_{k,j}+A_{i,k}C_{k,j}} \\
&= \sum_{k=1}^{m} \left(A_{i,k}B_{k,j} + A_{i,k}C_{k,j}\right) = \underbrace{\sum_{k=1}^{m} A_{i,k}B_{k,j}}_{\substack{=(AB)_{i,j} \\ \text{(by (14))}}} + \underbrace{\sum_{k=1}^{m} A_{i,k}C_{k,j}}_{\substack{=(AC)_{i,j} \\ \text{(by (15))}}} \\
&\qquad \left(\begin{array}{c} \text{by Proposition 2.30, applied to 1, } m, \\ A_{i,k}B_{k,j} \text{ and } A_{i,k}C_{k,j} \text{ instead of } p, q, a_k \text{ and } b_k \end{array} \right) \\
&= (AB)_{i,j} + (AC)_{i,j} \\
&= (AB+AC)_{i,j} \qquad \left(\begin{array}{c} \text{again because matrices} \\ \text{are added entry by entry} \end{array} \right)
\end{aligned}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$. In other words, each entry of the $n \times p$-matrix $A(B+C)$ equals the corresponding entry of $AB+AC$. Thus, the matrix $A(B+C)$ equals $AB+AC$. This proves Proposition 2.20 **(d)**. □

Before we can prove Proposition 2.20 **(g)**, we need another fact about sums:

Proposition 2.31. Let $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Assume that a number $a_{k,\ell}$ is given for every $k \in \{1,2,\ldots,m\}$ and $\ell \in \{1,2,\ldots,p\}$. Then,

$$\sum_{k=1}^{m} \sum_{\ell=1}^{p} a_{k,\ell} = \sum_{\ell=1}^{p} \sum_{k=1}^{m} a_{k,\ell}.$$

(Note that an expression like $\sum_{k=1}^{m} \sum_{\ell=1}^{p} a_{k,\ell}$ has to be understood as $\sum_{k=1}^{m} \left(\sum_{\ell=1}^{p} a_{k,\ell} \right)$.

It is a "nested sum", i.e., a sum of sums. For example,

$$\sum_{k=1}^{3} \underbrace{\sum_{\ell=1}^{4} k \cdot \ell}_{=k\cdot 1+k\cdot 2+k\cdot 3+k\cdot 4} = \sum_{k=1}^{3} (k\cdot 1 + k\cdot 2 + k\cdot 3 + k\cdot 4)$$
$$= (1\cdot 1 + 1\cdot 2 + 1\cdot 3 + 1\cdot 4) + (2\cdot 1 + 2\cdot 2 + 2\cdot 3 + 2\cdot 4) + (3\cdot 1 + 3\cdot 2 + 3\cdot 3 + 3\cdot 4).$$

)

Example 2.32. For $m = 2$ and $p = 3$, Proposition 2.31 says that

$$\sum_{k=1}^{2} \sum_{\ell=1}^{3} a_{k,\ell} = \sum_{\ell=1}^{3} \sum_{k=1}^{2} a_{k,\ell}.$$

In other words,

$$(a_{1,1} + a_{1,2} + a_{1,3}) + (a_{2,1} + a_{2,2} + a_{2,3}) = (a_{1,1} + a_{2,1}) + (a_{1,2} + a_{2,2}) + (a_{1,3} + a_{2,3}).$$

Proof of Proposition 2.31. Comparing

$$\sum_{k=1}^{m} \underbrace{\sum_{\ell=1}^{p} a_{k,\ell}}_{=a_{k,1}+a_{k,2}+\cdots+a_{k,p}}$$
(by the definition of the $\sum$ sign)
$$= \sum_{k=1}^{m} \left(a_{k,1} + a_{k,2} + \cdots + a_{k,p}\right)$$
$$= \left(a_{1,1} + a_{1,2} + \cdots + a_{1,p}\right) + \left(a_{2,1} + a_{2,2} + \cdots + a_{2,p}\right) + \cdots + \left(a_{m,1} + a_{m,2} + \cdots + a_{m,p}\right)$$
(by the definition of the $\sum$ sign)
$$= (\text{the sum of all possible numbers } a_{k,\ell})$$

with

$$\sum_{\ell=1}^{p} \underbrace{\sum_{k=1}^{m} a_{k,\ell}}_{\substack{=a_{1,\ell}+a_{2,\ell}+\cdots+a_{m,\ell} \\ \text{(by the definition of the } \sum \text{ sign)}}}$$
$$= \sum_{\ell=1}^{p} \left(a_{1,\ell} + a_{2,\ell} + \cdots + a_{m,\ell}\right)$$
$$= \left(a_{1,1} + a_{2,1} + \cdots + a_{m,1}\right) + \left(a_{1,2} + a_{2,2} + \cdots + a_{m,2}\right) + \cdots + \left(a_{1,p} + a_{2,p} + \cdots + a_{m,p}\right)$$
$$\left(\text{by the definition of the } \sum \text{ sign}\right)$$
$$= \left(\text{the sum of all possible numbers } a_{k,\ell}\right),$$

we obtain $\sum_{k=1}^{m} \left(\sum_{\ell=1}^{p} a_{k,\ell} \right) = \sum_{\ell=1}^{p} \left(\sum_{k=1}^{m} a_{k,\ell} \right)$.

(A more rigorous proof could be given using induction; but I don't want to move to that level of formalism in these notes. Notice the visual meaning of the above proof: If we place the mp numbers $a_{k,\ell}$ into a matrix $(a_{k,\ell})_{1 \leq k \leq m,\ 1 \leq \ell \leq p}$, then

- the number $\sum_{k=1}^{m} \sum_{\ell=1}^{p} a_{k,\ell}$ is obtained by summing the entries in each row of the matrix, and then summing the resulting sums;
- the number $\sum_{\ell=1}^{p} \sum_{k=1}^{m} a_{k,\ell}$ is obtained by summing the entries in each column of the matrix, and then summing the resulting sums.

Thus, clearly, both numbers are equal (namely, equal to the sum of all entries of the matrix).) □

Now, we can prove Proposition 2.20 **(g)**:

Proof of Proposition 2.20 (g). Let A be an $n \times m$-matrix. Let B be an $m \times p$-matrix. Let C be a $p \times q$-matrix.

We must show that $A(BC) = (AB)C$. In order to do so, it suffices to show that $(A(BC))_{i,j} = ((AB)C)_{i,j}$ for all $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, q\}$ (because this will show that respective entries of the two $n \times q$-matrices $A(BC)$ and $(AB)C$ are equal, and thus the two matrices are equal).

We know that A is an $n \times m$-matrix, and that BC is an $m \times q$-matrix. Hence, we can apply Proposition 2.28 to n, m, q, A and BC instead of n, m, p, A and B. We thus obtain

$$(A(BC))_{i,j} = \sum_{k=1}^{m} A_{i,k} (BC)_{k,j} \tag{16}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,q\}$. Also, we can apply Proposition 2.28 to m, p, q, B and C instead of n, m, p, A and B. We thus obtain

$$(BC)_{i,j} = \sum_{k=1}^{p} B_{i,k} C_{k,j} = \sum_{\ell=1}^{p} B_{i,\ell} C_{\ell,j} \tag{17}$$
$$\left(\text{here, we renamed the summation index } k \text{ as } \ell\right)$$

for all $i \in \{1,2,\ldots,m\}$ and $j \in \{1,2,\ldots,p\}$. Furthermore, we can apply Proposition 2.28 to n, p, q, AB and C instead of n, m, p, A and B. We thus find

$$((AB)\,C)_{i,j} = \sum_{k=1}^{p} (AB)_{i,k}\, C_{k,j} = \sum_{\ell=1}^{p} (AB)_{i,\ell}\, C_{\ell,j} \tag{18}$$
$$\left(\text{here, we renamed the summation index } k \text{ as } \ell\right)$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,q\}$. Finally, Proposition 2.28 (applied verbatim) yields

$$(AB)_{i,j} = \sum_{k=1}^{m} A_{i,k} B_{k,j} \tag{19}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$.

Now that we have found formulas for the entries of all matrices involved, we can perform our computation: For all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,q\}$, we have

$$\begin{aligned}
(A\,(BC))_{i,j} &= \sum_{k=1}^{m} A_{i,k} \underbrace{(BC)_{k,j}}_{\substack{=\sum_{\ell=1}^{p} B_{k,\ell} C_{\ell,j} \\ \text{(by (17), applied to } k \text{ instead of } i\text{)}}} && \text{(by (18))} \\
&= \sum_{k=1}^{m} \underbrace{A_{i,k} \left(\sum_{\ell=1}^{p} B_{k,\ell} C_{\ell,j} \right)}_{\substack{=\sum_{\ell=1}^{p} A_{i,k} B_{k,\ell} C_{\ell,j} \\ \text{(by an application of} \\ \text{Proposition 2.29)}}} = \sum_{k=1}^{m} \sum_{\ell=1}^{p} A_{i,k} B_{k,\ell} C_{\ell,j} \\
&= \sum_{\ell=1}^{p} \sum_{k=1}^{m} A_{i,k} B_{k,\ell} C_{\ell,j} && \left(\text{by Proposition 2.31, applied to } a_{k,\ell} = A_{i,k} B_{k,\ell} C_{\ell,j}\right)
\end{aligned}$$

and

$$\begin{aligned}
((AB)\,C)_{i,j} &= \sum_{\ell=1}^{p} \underbrace{(AB)_{i,\ell}}_{\substack{=\sum_{k=1}^{m} A_{i,k}B_{k,\ell} \\ \text{(by (19), applied to } \ell \\ \text{instead of } j)}} C_{\ell,j} \qquad \text{(by (16))} \\
&= \sum_{\ell=1}^{p} \underbrace{\left(\sum_{k=1}^{m} A_{i,k}B_{k,\ell}\right) C_{\ell,j}}_{\substack{=C_{\ell,j}\sum_{k=1}^{m} A_{i,k}B_{k,\ell} = \sum_{k=1}^{m} C_{\ell,j}A_{i,k}B_{k,\ell} \\ \text{(by an application of} \\ \text{Proposition 2.29)}}} = \sum_{\ell=1}^{p}\sum_{k=1}^{m} \underbrace{C_{\ell,j}A_{i,k}B_{k,\ell}}_{=A_{i,k}B_{k,\ell}C_{\ell,j}} \\
&= \sum_{\ell=1}^{p}\sum_{k=1}^{m} A_{i,k}B_{k,\ell}C_{\ell,j}.
\end{aligned}$$

Comparing these two equalities, we obtain

$$(A\,(BC))_{i,j} = ((AB)\,C)_{i,j}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,q\}$. In other words, each entry of the matrix $A\,(BC)$ equals the corresponding entry of the matrix $(AB)\,C$. Thus, the matrices $A\,(BC)$ and $(AB)\,C$ are equal. This proves Proposition 2.20 **(g)**. □

We have just proven the hardest part of Proposition 2.20. The rest is fairly straightforward.

2.10. The zero matrix

Definition 2.33. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, the $n \times m$ *zero matrix* means the matrix $(0)_{1\leq i\leq n,\ 1\leq j\leq m}$. This is the $n \times m$-matrix filled with zeroes. It is called $0_{n\times m}$. (When no confusion with the number 0 can arise, we will just call it 0.)

For example, the 2×3 zero matrix is $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Some authors (for example, Olver and Shakiban in [OlvSha06]) denote the zero matrix $0_{n\times m}$ by $\mathrm{O}_{n\times m}$ or $O_{n\times m}$ (thus using the letter O instead of the number 0), or simply by O.

The zero matrix behaves very much like the number 0:

Proposition 2.34. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then:
(a) We have $0_{n\times m} + A = A + 0_{n\times m} = A$ for each $n \times m$-matrix A.
(b) We have $0_{n\times m}A = 0_{n\times p}$ for each $p \in \mathbb{N}$ and each $m \times p$-matrix A.
(c) We have $A0_{n\times m} = 0_{p\times m}$ for each $p \in \mathbb{N}$ and each $p \times n$-matrix A.
(d) We have $0A = 0_{n\times m}$ for each $n \times m$-matrix A.
(e) We have $\lambda 0_{n\times m} = 0_{n\times m}$ for each number λ.

Remark 2.35. Numbers are known to be zero-divisor-free: If a product ab of two numbers a and b is 0, then one of a and b must be 0. This fails for matrices: If $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, then $AB = 0_{2\times 2}$ is the zero matrix, although neither A nor B is the zero matrix.

2.11. The identity matrix

Definition 2.36. Let $n \in \mathbb{N}$. The *diagonal entries* of an $n \times n$-matrix A are its entries $A_{1,1}, A_{2,2}, \ldots, A_{n,n}$. In other words, they are the entries $A_{i,j}$ for $i = j$.

For example, the diagonal entries of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ are a and d. The name "diagonal entries" comes from the visualization of an $n \times n$-matrix as a square table: When we say "diagonal", we always mean the diagonal of the square that connects the upper-left corner with the lower-right corner[16]; the diagonal entries are simply the entries along this diagonal. (The other diagonal is called the "antidiagonal" in linear algebra.)

Definition 2.37. If i and j are two objects (for example, numbers or sets or functions), then we set

$$\delta_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases}. \tag{20}$$

For example, $\delta_{3,3} = 1$ (since $3 = 3$) but $\delta_{1,2} = 0$ (since $1 \neq 2$). For another example, $\delta_{(1,2),(1,3)} = 0$ (because $(1,2) \neq (1,3)$); here we are using the notation $\delta_{i,j}$ in a situation where i and j are pairs of numbers.

The notation $\delta_{i,j}$ defined in (20) is called the *Kronecker delta*; it is extremely simple and yet highly useful. It has the property that $\delta_{i,j} = \delta_{j,i}$ for any i and j (because $i = j$ holds if and only if $j = i$).

Definition 2.38. Let $n \in \mathbb{N}$. Then, the $n \times n$ *identity matrix* means the matrix $\left(\delta_{i,j}\right)_{1\leq i\leq n,\ 1\leq j\leq n}$. This is the $n \times n$-matrix whose diagonal entries all equal 1, and whose all other entries equal 0. It is denoted by I_n. (Other people call it I or E or E_n.)

The $n \times n$ identity matrix I_n looks as follows:

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

[16] Often, this diagonal is also called the "main diagonal".

(with n rows and n columns). For example, the 3×3 identity matrix is $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

The identity matrix behaves very much like the number 1:

Proposition 2.39. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$.
(a) We have $I_n A = A$ for each $n \times m$-matrix A.
(b) We have $AI_m = A$ for each $n \times m$-matrix A.

Proposition 2.39 says that multiplying a matrix A by an identity matrix (from either side) does not change A. Thus, identity matrices have no effect inside a product, and so can be "cancelled" (or, more precisely, dropped). For example, if A, B, C and D are four $n \times n$-matrices, then $I_n A B I_n I_n C I_n D = ABCD$. (Of course, this is similar to dropping 1's from products of numbers: $1ab \cdot 1 \cdot 1c \cdot 1d = abcd$.)

2.12. (*) Proof of $AI_n = A$

Let me give a proof of Proposition 2.39 **(b)**, to illustrate the following simple, yet important point about summations and the Kronecker delta[17]:

Proposition 2.40. Let p and q be two integers such that $p \leq q$. Let $r \in \{p, p+1, \ldots, q\}$. Let $a_p, a_{p+1}, \ldots, a_q$ be some numbers. Then,

$$\sum_{k=p}^{q} a_k \delta_{k,r} = a_r.$$

Example 2.41. For $p = 1$, $q = 5$ and $r = 4$, Proposition 2.40 says that $\sum_{k=1}^{5} a_k \delta_{k,4} = a_4$. This is easy to check:

$$\begin{aligned} \sum_{k=1}^{5} a_k \delta_{k,4} &= a_1 \underbrace{\delta_{1,4}}_{\substack{=0 \\ \text{(since } 1 \neq 4\text{)}}} + a_2 \underbrace{\delta_{2,4}}_{\substack{=0 \\ \text{(since } 2 \neq 4\text{)}}} + a_3 \underbrace{\delta_{3,4}}_{\substack{=0 \\ \text{(since } 3 \neq 4\text{)}}} + a_4 \underbrace{\delta_{4,4}}_{\substack{=1 \\ \text{(since } 4 = 4\text{)}}} + a_5 \underbrace{\delta_{5,4}}_{\substack{=0 \\ \text{(since } 5 \neq 4\text{)}}} \\ &= a_1 0 + a_2 0 + a_3 0 + a_4 1 + a_5 0 = a_4 1 = a_4. \end{aligned}$$

What you should see on this example is that all but one addends of the sum $\sum_{k=p}^{q} a_k \delta_{k,r}$ are zero, and the remaining one addend is $a_r \underbrace{\delta_{r,r}}_{=1} = a_r 1 = a_r$. The proof below is just writing this down in the general situation.

[17] This is something that often comes up in computations (particularly in physics and computer science, where the use of the Kronecker delta is widespread).

Proof of Proposition 2.40. Let us first notice something simple: For any $k \in \{p, p+1, \ldots, q\}$ such that $k \neq r$, we have

$$a_k \underbrace{\delta_{k,r}}_{\substack{=0 \\ (\text{since } k \neq r)}} = a_k 0 = 0.$$

In other words, all terms of the form $a_k \delta_{k,r}$ with $k \neq r$ are 0. Hence, the sum of all these terms is 0 as well. In other words,

$$(\text{the sum of all terms of the form } a_k \delta_{k,r} \text{ with } k \neq r) = 0. \tag{21}$$

By the definition of the $\sum$ sign, we have

$$\begin{aligned}
\sum_{k=p}^{q} a_k \delta_{k,r} &= a_p \delta_{p,r} + a_{p+1} \delta_{p+1,r} + \cdots + a_q \delta_{q,r} \\
&= (\text{the sum of all terms of the form } a_k \delta_{k,r}) \\
&= a_r \underbrace{\delta_{r,r}}_{\substack{=1 \\ (\text{since } r=r)}} + \underbrace{(\text{the sum of all terms of the form } a_k \delta_{k,r} \text{ with } k \neq r)}_{\substack{=0 \\ (\text{by (21)})}} \\
&\qquad (\text{here, we have pulled out the addend } a_r \delta_{r,r} \text{ out of the sum}) \\
&= a_r 1 + 0 = a_r.
\end{aligned}$$

□

*Proof of Proposition 2.39 **(b)**.* We have $I_m = \left(\delta_{i,j}\right)_{1 \leq i \leq m,\ 1 \leq j \leq m}$ (this is how we defined I_m), and thus

$$(I_m)_{u,v} = \delta_{u,v} \qquad \text{for all } u \in \{1, 2, \ldots, m\} \text{ and } v \in \{1, 2, \ldots, m\}. \tag{22}$$

Let A be an $n \times m$-matrix. For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, we have

$$\begin{aligned}
(AI_m)_{i,j} &= \sum_{k=1}^{m} A_{i,k} \underbrace{(I_m)_{k,j}}_{\substack{=\delta_{k,j} \\ (\text{by (22)})}} \qquad (\text{by Proposition 2.28, applied to } p = m \text{ and } B = I_m) \\
&= \sum_{k=1}^{m} A_{i,k} \delta_{k,j} = A_{i,j}
\end{aligned}$$

(by Proposition 2.40, applied to $p = 1$, $q = m$, $r = j$ and $a_k = A_{i,k}$). In other words, each entry of the $n \times m$-matrix AI_m equals the corresponding entry of the $n \times m$-matrix A. In other words, AI_m equals A. This proves Proposition 2.39 **(b)**. □

Proposition 2.39 **(a)** can be proven similarly (but this time, instead of the sum $\sum_{k=1}^{m} A_{i,k} \delta_{k,j}$, we must consider the sum $\sum_{k=1}^{n} \delta_{i,k} A_{k,j} = \sum_{k=1}^{n} A_{k,j} \delta_{i,k} = A_{i,j}$).

2.13. Powers of a matrix

The k-th power of a number a (where $k \in \mathbb{N}$) is defined by repeated multiplication: We start with $a^0 = 1$ [18], and we define each next power of a by multiplying the previous one by a. In formulas: $a^{k+1} = a \cdot a^k$ for each $k \in \mathbb{N}$. Thus,

$$a^1 = a \cdot \underbrace{a^0}_{=1} = a \cdot 1 = a;$$
$$a^2 = a \cdot \underbrace{a^1}_{=a} = a \cdot a;$$
$$a^3 = a \cdot \underbrace{a^2}_{=a \cdot a} = a \cdot a \cdot a,$$

etc.. We can explicitly write

$$a^k = \underbrace{a \cdot a \cdot \cdots \cdot a}_{k \text{ times } a} \qquad \text{for each } k \in \mathbb{N},$$

where we understand $\underbrace{a \cdot a \cdot \cdots \cdot a}_{0 \text{ times } a}$ to mean 1 [19].

We can play the same game with square matrices, but instead of the number 1 we now take the $n \times n$ identity matrix I_n:

Definition 2.42. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then, the k-th power of the matrix A (where $k \in \mathbb{N}$) is defined by repeated multiplication: We start with $A^0 = I_n$, and we define each next power of A by multiplying the previous one by A. In formulas: $A^{k+1} = A \cdot A^k$ for each $k \in \mathbb{N}$. Explicitly, $A^k = \underbrace{A \cdot A \cdot \cdots \cdot A}_{k \text{ times } A}$ for each $k \in \mathbb{N}$, where the empty product of $n \times n$-matrices is defined to be I_n. (An "empty product" is a product with no factors. Thus, $A^0 = \underbrace{A \cdot A \cdot \cdots \cdot A}_{0 \text{ times } A}$ is an empty product.)

Notice that we have been a bit sloppy when we said "multiplying the previous one by A": When we multiply a matrix B by A, we might mean either AB or BA, and as we know, these two products can be different (matrices don't always commute!). However, in the above definition, this makes no matter, because both definitions lead to the same explicit formula $A^k = \underbrace{A \cdot A \cdot \cdots \cdot A}_{k \text{ times } A}$ (which is well-defined because of general associativity).

We now have a first moderately interesting example of commuting matrices: Any two powers of a square matrix commute. (In other words: For any $n \times n$-matrix A, and any $u \in \mathbb{N}$ and $v \in \mathbb{N}$, the two matrices A^u and A^v commute. This follows by observing that $A^u A^v = A^{u+v} = A^v A^u$.)

[18] Yes, this is how a^0 is defined, for all a. Anyone who tells you that the number 0^0 is undefined is merely spreading their confusion.

[19] This is a standard convention: An empty product of numbers always means 1.

2.14. (*) The summation sign for matrices

In Definition 2.26, we have introduced a notation for sums of arbitrary (finite) lists of numbers. The same notation can be used for sums of arbitrary (finite) lists of matrices:

Definition 2.43. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let p and q be two integers such that $p \leq q+1$. Let $A_p, A_{p+1}, \ldots, A_q$ be some $n \times m$-matrices. Then, $\sum_{k=p}^{q} A_k$ means the sum $A_p + A_{p+1} + \cdots + A_q$. (This sum is well-defined due to Proposition 2.23. Moreover, the order of its addends does not matter, as can be proven using Proposition 2.20 **(a)**.)

The notation $\sum_{k=p}^{q} A_k$ is analogous to the notation $\sum_{k=p}^{q} a_k$ introduced in Definition 2.26. The same remarks and clarifications done for the latter notation in Definition 2.27 apply to the former notation. There is only one difference: When the sum $\sum_{k=p}^{q} A_k$ has no addends (i.e., when $p \geq q+1$), its value is defined to be the zero matrix $0_{n \times m}$ rather than the number 0. (This is not much of a difference, seeing that the zero matrix $0_{n \times m}$ and the number 0 behave similarly; see, e.g., Proposition 2.34.)

Recall that addition of matrices was defined entry by entry. Let me restate this in terms of a single entry of a sum:

Proposition 2.44. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A and B be two $n \times m$-matrices. For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, we have

$$(A+B)_{i,j} = A_{i,j} + B_{i,j}.$$

Proof of Proposition 2.44. We have $A + B = \left(A_{i,j} + B_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq m}$ (by the definition of $A+B$). Thus, $(A+B)_{i,j} = A_{i,j} + B_{i,j}$ for each $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$. This proves Proposition 2.44. □

Proposition 2.44 concerns the sum of two matrices. Using the summation sign, we can state an analogue of Proposition 2.44 for the sum of **several** matrices:

Proposition 2.45. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $h \in \mathbb{N}$. Let $A_1, A_2, \ldots, A_h$ be some $n \times m$-matrices. For every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, we have

$$\left(\sum_{k=1}^{h} A_k \right)_{i,j} = \sum_{k=1}^{h} \left(A_k\right)_{i,j}. \tag{23}$$

Notice that the $\sum$ sign on the left hand side of (23) stands for a sum of several matrices, while the $\sum$ sign on the right hand side stands for a sum of several numbers.

Proof of Proposition 2.45. We could say that proving Proposition 2.45 is just a matter of applying Proposition 2.44 several times (since a sum of finitely many matrices can be obtained by repeatedly adding one matrix to another). This single sentence might not be a rigorous proof in itself, but it would pass for a proof in any mathematical paper or textbook, because any mathematician can easily make it as rigorous as she wants to have it by filling in the missing (straightforward) details.[20]

However, these notes are supposed to double as an introduction to proofs, so let me actually show how a rigorous proof of Proposition 2.45 looks like (even though it is exactly as straightforward and dull as you would expect it to be).

Let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. We claim that

$$\left(\sum_{k=1}^{\ell} A_k\right)_{i,j} = \sum_{k=1}^{\ell} \left(A_k\right)_{i,j} \tag{24}$$

for every $\ell \in \{0,1,\ldots,h\}$. (Once this is proven, then we will be able to obtain (23) simply by applying (24) to $\ell = h$.)

We will prove (24) by *induction over* ℓ. (If you have never seen a proof by induction: this here is an example.) This means that we shall prove the following two claims:

> *Claim 1:* (24) holds for $\ell = 0$.
>
> *Claim 2:* If $L \in \{0,1,\ldots,h-1\}$ is such that (24) holds for $\ell = L$, then (24) also holds for $\ell = L+1$.

Once these two claims are proven, the *principle of mathematical induction* will yield that (24) holds for all $\ell \in \{0,1,\ldots,h\}$. In fact:

- Claim 1 shows that (24) holds for $\ell = 0$;
- thus, Claim 2 (applied to $L = 0$) shows that (24) holds for $\ell = 1$;
- thus, Claim 2 (applied to $L = 1$) shows that (24) holds for $\ell = 2$;
- thus, Claim 2 (applied to $L = 2$) shows that (24) holds for $\ell = 3$;
- and so on, applying Claim 2 for higher and higher L, until we arrive at $\ell = h$.

[20] Actually, there is one more subtlety involved: Namely, if $h = 0$, then the sums appearing in (23) are not obtained by addition (in fact, they are empty), and therefore we cannot use Proposition 2.44 to prove (23) in this case. However, (23) is completely obvious in this case anyway (since it just says that $\left(0_{n\times m}\right)_{i,j} = 0$).

(See [LeLeMe16, Chapter 5] for an introduction to proofs by induction.)
Of course, we still have to prove the two claims.

1. *Proof of Claim 1:* For $L = 0$, the statement (24) claims that $\left(\sum_{k=1}^{0} A_k\right)_{i,j} = \sum_{k=1}^{0} (A_k)_{i,j}$. But

$$\sum_{k=1}^{0} A_k = (\text{an empty sum of } n \times m\text{-matrices}) = 0_{n\times m}$$

(since an empty sum of $n \times m$-matrices was **defined** to be $0_{n\times m}$). Hence,

$$\left(\sum_{k=1}^{0} A_k\right)_{i,j} = (0_{n\times m})_{i,j} = 0$$

(since every entry of the zero matrix $0_{n\times m}$ is 0). Compared with

$$\sum_{k=1}^{0} (A_k)_{i,j} = (\text{an empty sum of numbers}) = 0,$$

this yields $\left(\sum_{k=1}^{0} A_k\right)_{i,j} = \sum_{k=1}^{0} (A_k)_{i,j}$. In other words, (24) holds for $\ell = 0$. This proves Claim 1.

2. *Proof of Claim 2:* Let $L \in \{0, 1, \ldots, h-1\}$ be such that (24) holds for $\ell = L$. We must show that (24) holds for $\ell = L + 1$.

   Since (24) holds for $\ell = L$, we have

$$\left(\sum_{k=1}^{L} A_k\right)_{i,j} = \sum_{k=1}^{L} (A_k)_{i,j}. \tag{25}$$

   Now,

$$\sum_{k=1}^{L+1} A_k = \sum_{k=1}^{L} A_k + A_{L+1}.$$

Thus,

$$\left(\sum_{k=1}^{L+1} A_k\right)_{i,j} = \left(\sum_{k=1}^{L} A_k + A_{L+1}\right)_{i,j} = \underbrace{\left(\sum_{k=1}^{L} A_k\right)_{i,j}}_{\substack{=\sum_{k=1}^{L}(A_k)_{i,j} \\ \text{(by (25))}}} + (A_{L+1})_{i,j}$$

$$\left(\text{by Proposition 2.44, applied to } A = \sum_{k=1}^{L} A_k \text{ and } B = A_{L+1}\right)$$

$$= \sum_{k=1}^{L} (A_k)_{i,j} + (A_{L+1})_{i,j} = \sum_{k=1}^{L+1} (A_k)_{i,j}.$$

In other words, (24) holds for $\ell = L+1$. This proves Claim 2.

Now, both Claims 1 and 2 are proven, and thus (as we have explained above) the proof of (24) is complete.

(The way we organized our proof of (24) is typical for a proof by mathematical induction. Usually, the proof of Claim 1 is called the "induction base", and the proof of Claim 2 is called the "induction step". In the induction step, we have made the assumption that (24) holds for $\ell = L$; this assumption is called the "induction hypothesis".)

Now that (24) is proven, we can simply apply (24) to $\ell = h$, and conclude that $\left(\sum_{k=1}^{h} A_k\right)_{i,j} = \sum_{k=1}^{h} (A_k)_{i,j}$. This proves Proposition 2.45. □

The sums in Proposition 2.45 range from $k = 1$ to h; but the same statement holds for arbitrary sums:

Proposition 2.46. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let p and q be two integers such that $p \leq q+1$. Let $A_p, A_{p+1}, \ldots, A_q$ be some $n \times m$-matrices. For every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, we have

$$\left(\sum_{k=p}^{q} A_k\right)_{i,j} = \sum_{k=p}^{q} (A_k)_{i,j}. \tag{26}$$

As a further illustration of manipulating sums, let me derive this proposition from Proposition 2.45:

Proof of Proposition 2.46. Let $h = q+1-p$. Then, $h = q+1-p \geq 0$ (since $p \leq q+1$), so that $h \in \mathbb{N}$. Also,

$$p - 1 + \underbrace{h}_{=q+1-p} = p - 1 + q + 1 - p = q.$$

Notice that $A_p, A_{p+1}, \ldots, A_q$ are $q+1-p$ matrices. In other words, $A_p, A_{p+1}, \ldots, A_q$ are h matrices (since $h = q+1-p$). Denote these h matrices by $B_1, B_2, \ldots, B_h$. Thus,

$$B_k = A_{p-1+k} \qquad \text{for every } k \in \{1,2,\ldots,h\}. \tag{27}$$

Let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. We have

$$\begin{aligned}
\sum_{k=1}^{h} \underbrace{B_k}_{\substack{=A_{p-1+k} \\ \text{(by (27))}}} &= \sum_{k=1}^{h} A_{p-1+k} = A_p + A_{p+1} + \cdots + A_{p-1+h} \\
&= A_p + A_{p+1} + \cdots + A_q \qquad (\text{since } p-1+h = q) \\
&= \sum_{k=p}^{q} A_k
\end{aligned} \tag{28}$$

and

$$\begin{aligned}
\sum_{k=1}^{h} \left(\underbrace{B_k}_{\substack{=A_{p-1+k} \\ \text{(by (27))}}} \right)_{i,j} &= \sum_{k=1}^{h} \left(A_{p-1+k}\right)_{i,j} = \left(A_p\right)_{i,j} + \left(A_{p+1}\right)_{i,j} + \cdots + \left(A_{p-1+h}\right)_{i,j} \\
&= \left(A_p\right)_{i,j} + \left(A_{p+1}\right)_{i,j} + \cdots + \left(A_q\right)_{i,j} \qquad (\text{since } p-1+h = q) \\
&= \sum_{k=p}^{q} \left(A_k\right)_{i,j}.
\end{aligned} \tag{29}$$

But Proposition 2.45 (applied to $B_1, B_2, \ldots, B_h$ instead of $A_1, A_2, \ldots, A_h$) shows that

$$\left(\sum_{k=1}^{h} B_k \right)_{i,j} = \sum_{k=1}^{h} \left(B_k\right)_{i,j}.$$

In view of (28) and (29), this rewrites as

$$\left(\sum_{k=p}^{q} A_k \right)_{i,j} = \sum_{k=p}^{q} \left(A_k\right)_{i,j}.$$

Thus, Proposition 2.46 is proven. □

2.15. (*) Application: Fibonacci numbers

Here is a simple application of matrix multiplication to elementary mathematics.

Definition 2.47. The *Fibonacci sequence* is the sequence $(0,1,1,2,3,5,8,13,21,34,55,\ldots)$ that is defined as follows: Its first two entries are 0 and 1, and each further entry is the sum of the previous two entries. In more formal terms, it is the sequence $(f_0, f_1, f_2, f_3, \ldots)$ (we start the labelling at 0) defined recursively by

$$\begin{aligned} &f_0 = 0, \qquad f_1 = 1, \qquad \text{and} \\ &f_n = f_{n-1} + f_{n-2} \qquad \text{for every } n \geq 2. \end{aligned}$$

The elements of this sequence are called the *Fibonacci numbers*.

Definition 2.47 gives a straightforward way to compute each particular Fibonacci number f_n, by computing the first $n+1$ Fibonacci numbers $f_0, f_1, \ldots, f_n$ one after the others. For example, it gives

$$\begin{aligned} &f_0 = 0; \qquad f_1 = 1; \qquad f_2 = 1+0 = 1; \qquad f_3 = 1+1 = 2; \qquad f_4 = 2+1 = 3; \\ &f_5 = 3+2 = 5; \qquad f_6 = 5+3 = 8; \qquad f_7 = 8+5 = 13; \qquad f_8 = 13+8 = 21, \end{aligned}$$

and so on. However, when n is large, computing f_n by this method is time-consuming (each of the $n+1$ first Fibonacci numbers has to be computed!). Is there a faster way to compute Fibonacci numbers?

It turns out that there is. It is based on the following fact:

Proposition 2.48. Let B be the 2×2-matrix $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Then, for every positive integer n, we have

$$B^n = \begin{pmatrix} f_{n+1} & f_n \\ f_n & f_{n-1} \end{pmatrix}. \tag{30}$$

Proof of Proposition 2.48. We shall prove Proposition 2.48 by induction over n:

Induction base: We have $B^1 = B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Comparing this with

$$\begin{pmatrix} f_{1+1} & f_1 \\ f_1 & f_{1-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \qquad (\text{since } f_{1+1} = f_2 = 1,\ f_1 = 1 \text{ and } f_{1-1} = f_0 = 0),$$

we obtain $B^1 = \begin{pmatrix} f_{1+1} & f_1 \\ f_1 & f_{1-1} \end{pmatrix}$. In other words, Proposition 2.48 holds for $n = 1$. This completes the induction base. (This was a completely straightforward computation. In the future, we will often leave such computations to the reader.)

Induction step: Let N be a positive integer. Assume that Proposition 2.48 holds for $n = N$. We must show that Proposition 2.48 also holds for $n = N+1$.

The definition of the Fibonacci sequence shows that $f_{N+2} = f_{N+1} + f_N$ and $f_{N+1} = f_N + f_{N-1}$.

We have assumed that Proposition 2.48 holds for $n = N$. In other words,

$$B^N = \begin{pmatrix} f_{N+1} & f_N \\ f_N & f_{N-1} \end{pmatrix}.$$

Now,

$$\begin{aligned} B^{N+1} &= \underbrace{B^N}_{=\begin{pmatrix} f_{N+1} & f_N \\ f_N & f_{N-1} \end{pmatrix}} \underbrace{B}_{=\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}} = \begin{pmatrix} f_{N+1} & f_N \\ f_N & f_{N-1} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} f_{N+1} \cdot 1 + f_N \cdot 1 & f_{N+1} \cdot 1 + f_N \cdot 0 \\ f_N \cdot 1 + f_{N-1} \cdot 1 & f_N \cdot 1 + f_{N-1} \cdot 0 \end{pmatrix} \\ &\qquad \text{(by the definition of a product of two matrices)} \\ &= \begin{pmatrix} f_{N+1} + f_N & f_{N+1} \\ f_N + f_{N-1} & f_N \end{pmatrix} = \begin{pmatrix} f_{N+2} & f_{N+1} \\ f_{N+1} & f_N \end{pmatrix} \end{aligned}$$

(since $f_{N+1} + f_N = f_{N+2}$ and $f_N + f_{N-1} = f_{N+1}$). In other words, Proposition 2.48 holds for $n = N + 1$. This completes the induction step; hence, Proposition 2.48 is proven. □

How does Proposition 2.48 help us compute f_n quickly? Naively computing B^n by multiplying B with itself n times is not any faster than computing f_n directly using Definition 2.47 (in fact, it is slower, since multiplying matrices takes longer than adding numbers). However, there is a trick for computing powers quickly, called *binary exponentiation*; this trick works just as well for matrices as it does for numbers. The trick uses the following observations:

- For every $m \in \mathbb{N}$, we have $B^{2m} = (B^m)^2$.
- For every $m \in \mathbb{N}$, we have $B^{2m+1} = B\,(B^m)^2$.

These observations allow us to quickly compute B^{2m} and B^{2m+1} using only B^m; thus, we can "jump up" from B^m directly to B^{2m} and to B^{2m+1} without the intermediate steps $B^{m+1}, B^{m+2}, \ldots, B^{2m-1}$. Let us use this to compute B^{90} (and thus f_{90}) quickly (without computing 91 Fibonacci numbers):

- We want to find B^{90}. Since $90 = 2 \cdot 45$, we have $B^{90} = \left(B^{45}\right)^2$ (by the formula $B^{2m} = (B^m)^2$).
- We thus want to find B^{45}. Since $45 = 2 \cdot 22 + 1$, we have $B^{45} = B\left(B^{22}\right)^2$ (by the formula $B^{2m+1} = B\,(B^m)^2$).
- We thus want to find B^{22}. Since $22 = 2 \cdot 11$, we have $B^{22} = \left(B^{11}\right)^2$ (by the formula $B^{2m} = (B^m)^2$).

- We thus want to find B^{11}. Since $11 = 2 \cdot 5 + 1$, we have $B^{11} = B\left(B^5\right)^2$ (by the formula $B^{2m+1} = B\left(B^m\right)^2$).
- We thus want to find B^5. Since $5 = 2 \cdot 2 + 1$, we have $B^5 = B\left(B^2\right)^2$ (by the formula $B^{2m+1} = B\left(B^m\right)^2$).
- We thus want to find B^2. Since $2 = 2 \cdot 1$, we have $B^2 = \left(B^1\right)^2$ (by the formula $B^{2m} = \left(B^m\right)^2$, but this was obvious anyway).

We know what B^1 is: $B^1 = B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Hence, $B^2 = \left(B^1\right)^2$ becomes $B^2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Hence, $B^5 = B\left(B^2\right)^2$ becomes $B^5 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^2 = \begin{pmatrix} 8 & 5 \\ 5 & 3 \end{pmatrix}$. Hence, $B^{11} = B\left(B^5\right)^2$ becomes $B^{11} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 8 & 5 \\ 5 & 3 \end{pmatrix}^2 = \begin{pmatrix} 144 & 89 \\ 89 & 55 \end{pmatrix}$. Hence, $B^{22} = \left(B^{11}\right)^2$ becomes $B^{22} = \begin{pmatrix} 144 & 89 \\ 89 & 55 \end{pmatrix}^2 = \begin{pmatrix} 28657 & 17711 \\ 17711 & 10946 \end{pmatrix}$. Hence, $B^{45} = B\left(B^{22}\right)^2$ becomes

$$B^{45} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 28657 & 17711 \\ 17711 & 10946 \end{pmatrix}^2 = \begin{pmatrix} 1836311903 & 1134903170 \\ 1134903170 & 701408733 \end{pmatrix}.$$

Hence, $B^{90} = \left(B^{45}\right)^2$ becomes

$$B^{90} = \begin{pmatrix} 1836311903 & 1134903170 \\ 1134903170 & 701408733 \end{pmatrix}^2 = \begin{pmatrix} 4660046610375530309 & 2880067194370816120 \\ 2880067194370816120 & 1779979416004714189 \end{pmatrix}.$$

Since f_{90} is the $(2,1)$-th entry of B^{90} (indeed, (30) shows that f_n is the $(2,1)$-th entry of B^n for all positive integers n), we thus obtain

$$f_{90} = 2880067194370816120.$$

Exercise 2.49. Show that, for any two positive integers n and m, we have

$$f_{n+m} = f_n f_{m+1} + f_{n-1} f_m.$$

[**Hint:** Begin with the equality $B^n B^m = B^{n+m}$. Rewrite it using Proposition 2.48, and compare entries.]

2.16. (*) What is a number?

So far, our notion of a matrix relies on a (somewhat vague) notion of a "number". What does the word "number" mean? There are several possible candidates for a meaning of this word: for example, "number" might mean "rational number", but might also mean "real number" or "complex number". For what we have been doing so far, the precise choice of meaning does not matter. However, it eventually

will matter, so let me discuss it briefly. (See any book on abstract algebra for a more detailed and systematic discussion.)

First, let me introduce some well-known sets:

- As explained above, $\mathbb{N}$ means the set of all nonnegative integers: $\mathbb{N} = \{0, 1, 2, \ldots\}$.
- Furthermore, $\mathbb{Z}$ means the set of all integers: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
- Moreover, $\mathbb{Q}$ means the set of all rational numbers: $\mathbb{Q} = \left\{ \frac{a}{b} \mid a \in \mathbb{Z},\ b \in \mathbb{Z} \setminus \{0\} \right\}$.
- Furthermore, $\mathbb{R}$ means the set of all real numbers. This contains rational numbers such as -2 and $\frac{5}{3}$, but also irrational numbers such as $\sqrt{2}$ and $\frac{\sqrt{3}}{1+\sqrt[3]{5}}$ and π (and various others, many of which cannot even be described in words[21]).
- Finally, $\mathbb{C}$ means the set of all complex numbers. They will be rarely used in these notes (indeed, most of linear algebra can be done without them, except for eigenvalue/eigenvector theory), so you do not actually have to know them in order to read these notes. However, let me give a quick briefing on complex numbers (probably more of a reminder for those who have already seen them):

  Complex numbers can be formally defined as pairs of real numbers (a, b) with an entrywise addition (that is, $(a, b) + (a', b') = (a + a', b + b')$, exactly like 1×2-matrices) and a somewhat strange-looking multiplication (namely, $(a, b)(a', b') = (aa' - bb', ab' + ba')$). [22] But the way everyone thinks about complex numbers (informally) is that they are an extension of real numbers (so $\mathbb{R}$ is a subset of $\mathbb{C}$) by adding a new "imaginary number" i which satisfies $i^2 = -1$. They are supposed to behave like real numbers as far as laws of addition and multiplication are concerned (thus, for instance, $a(b + c) = ab + ac$ and $a(bc) = (ab)c$); using these laws and the requirement that $i^2 = -1$, one can easily see how to multiply and add arbitrary complex numbers. These two definitions (the formal one as pairs of real numbers, and the informal one as "extended real numbers") are equivalent, and the complex number (a, b)

[21] This is not a poetic metaphor. What I am saying is that there are real numbers which cannot be described by any finite formula or computed (to arbitrary precision) by any finite algorithm. The reason is simply that there are uncountably many real numbers, but only countably many formulas and algorithms. If you find this unintuitive, imagine an immortal monkey typing an infinite decimal number on an infinite-memory computer: `9.461328724290054`... If the monkey is typing truly at random, then no finite rule or formula will suffice to predict every single digit he types; thus, the real number he defines is undescribable. (This, of course, is not a proof.)

[22] This sort of definition is not unlike our definition of matrices: they are also tables with entrywise addition and a less simple multiplication.

(according to the first definition) corresponds to the complex number $a + bi$ (according to the second definition).

For a detailed introduction to complex numbers, see [LaNaSc16, §2].

- We shall occasionally use another set: the set $\overline{\mathbb{Q}}$ of all *algebraic numbers*. These are complex numbers that are roots of nonzero polynomials with rational coefficients. For instance, $\sqrt{2}$ is an algebraic number[23], and i is an algebraic number[24], and $\frac{5}{2}$ is an algebraic number[25], and $\sqrt{2}+\sqrt{3}$ is an algebraic number[26]; but π (for example) is not an algebraic number[27].

  Again, you can read these notes without understanding algebraic numbers, but let me explain why they are useful.

  First, let me explain why real numbers are "bad". Computations with rational numbers can be done **exactly** on a computer: for example, a computer algebra system (such as SageMath) can evaluate to $\frac{\frac{12}{13}-\frac{1}{15}}{\frac{2}{3}+3}$ to $\frac{167}{715}$ with a 100% accuracy and a 100% guarantee of correctness[28]. Similarly, it can do all kinds of computations with rational numbers with 100% accuracy (as long as it doesn't run out of memory). However, computations with real numbers are **doomed to be inexact**[29]. A real number can encode "an infinite amount of information"[30]; no computer can even store such a thing[31]. The consequence is that computers will work with approximations when you give them real numbers. Even real numbers like π that can be **defined** in finite time still cannot be **computed with** automatically without approximation. Typically, computers approximate real numbers by **floating-point numbers**, which store only a fixed number of significant digits.[32] When you use approximations, you need to be prepared to get wrong results; sometimes the error in the result will be

[23] because it is a root of the polynomial $X^2 - 2$

[24] since it is a root of $X^2 + 1$

[25] since it is a root of $X - \frac{5}{2}$ (this sounds stupid, but it is perfectly valid, since $-\frac{5}{2}$ is rational)

[26] since it is a root of $X^4 - 10X^2 + 1$ (check this!)

[27] ... which is the reason why the circle cannot be squared. The actual proof is rather difficult; see `http://mathoverflow.net/questions/34055/transcendence-of-pi` and `http://math.stackexchange.com/questions/31798/prove-that-pi-is-a-transcendental-number` for some references.

[28] barring bugs in the software

[29] Complex numbers suffer from the same problem.

[30] Think of a real number as an infinite decimal fraction, encoded by an infinite sequence of digits.

[31] This is precisely the above-mentioned problem of the ineffable real numbers.

[32] How many depends on the system and the application. For example, the Python 2.7.10 on my computer seems to treat 1.000000000000000000001 and 1 as identical, but it can tell 1.00000000000001 apart from 1. See `https://docs.python.org/2/tutorial/floatingpoint.html` for a more thorough explanation of how Python approximates real numbers.

completely out of proportion to the little inaccuracies in the approximation! Here is an example: The system of linear equations

$$\begin{cases} x + 2y = 2; \\ 3x + 6y = 6 \end{cases} \tag{31}$$

in two variables x and y has infinitely many solutions (namely, $(x, y) = (2t, 1 - t)$ is a solution for each number t). But if we allow ourself one little inaccuracy (the kind that computers necessarily do when they work with real numbers) and replace the 3 by 3.000000000001, then we obtain the system

$$\begin{cases} x + 2y = 2; \\ 3.000000000001x + 6y = 6 \end{cases}, \tag{32}$$

which has only one solution (namely, $(x, y) = (0, 1)$). On the other hand, if we instead change the second 6 in the second equation of (31) to a 6.00000000001 (again, a typical little imprecision), then the resulting system

$$\begin{cases} x + 2y = 2; \\ 3x + 6y = 6.00000000001 \end{cases} \tag{33}$$

will have no solutions at all. Thus, the minuscule differences between the three systems (31), (32) and (33) have led to three wildly different results (infinitely many, one or no solutions). The consequence is that **if you want a computer to reliably solve the system (31), you must make sure that it treats the coefficients (**$1, 2, 2, 3, 6, 6$**) as rational numbers (not as real numbers)** and avoids any approximations.

This, of course, only works well if the coefficients **are** rational numbers. You cannot solve a system like

$$\begin{cases} \sqrt{3}x + \sqrt{2}y = \pi; \\ \left(1 - \sqrt{3}\right) x + \pi y = 0 \end{cases} \tag{34}$$

this way. More annoyingly, you cannot solve a system like

$$\begin{cases} \sqrt{3}x + \sqrt{2}y = 1; \\ \left(1 - \sqrt{3}\right) x + 2y = 0 \end{cases} \tag{35}$$

this way, although you probably could solve it by hand! Systems like (34) are (in a sense) hopeless: Computers cannot reliably work with real numbers without approximating them at some point. But (35) can be salvaged: All the coefficients in (35) are algebraic numbers, and modern computer algebra

systems (e.g., SageMath) can work with algebraic numbers with 100% precision.[33] Thus, even if you don't care much about algebraic numbers, you will need to tell your computer that your numbers are algebraic in order to have it solve systems like (35).

Of course, this all doesn't mean that linear algebra with real numbers is useless in practice. The system (31) is a rather ill-behaved system; many systems allow for a pretty good approximation of their solutions even in spite of imprecisions, and even when a system is ill-behaved like (31), there are methods that compute the "likeliest solution" (such as the least-squares method). We shall (hopefully) see some of these methods in these notes. A less alarmist slogan would thus be: You can do linear algebra with real numbers, but you should be aware of its limitations and keep track of the errors ("numerical stability").

Note that $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ (at least if you identify a real number a with the complex number $(a, 0) = a + 0i$) and $\mathbb{Q} \subseteq \overline{\mathbb{Q}} \subseteq \mathbb{C}$.

So what is a number? Integers, rational numbers, real numbers, complex numbers and algebraic numbers all have good claims to the name "number". And even though all of these numbers can be viewed as complex numbers, it can be useful to **not** treat them as complex numbers by default, for example when you are using a computer and want 100% precision.

Here is a better question: What are the things we can fill a matrix with? We have so far used numbers, but we don't have to; we could also (for example) use polynomials. The matrix $\begin{pmatrix} 2X^2+1 & -1 \\ X & X+7 \end{pmatrix}$ is a 2×2-matrix whose entries are not numbers, but polynomials in the variable X (with rational coefficients). Such matrices can be highly useful (and, in fact, will be used when we come to eigenvalues). Such matrices can be added and multiplied (since polynomials can be added and multiplied). Of course, we could also fill a matrix with all kinds of things (words, names, smilies, scribbles) that **cannot** be added or multiplied (after all, matrices are just tables); but then we won't be able to add and multiply the resulting matrices, so we don't gain anything by calling them "matrices" (we don't want just a fancy synonym for "tables"). So it sounds most reasonable to expect that a matrix should be filled with things that can be added and multiplied. Moreover, addition and multiplication of these things should obey certain laws (such as $(a+b)\,c = ac + bc$ and $ab = ba$) in order to ensure that addition, scaling and multiplication of matrices will obey the usual laws (e.g., Proposition 2.20) as well. Formalizing this idea, we arrive at the notion of a commutative ring:

[33] This is **not** easy! For example, a classical puzzle asks you to prove that $\sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}} = 1$. This is not obvious; there are no straightforward "simplifications" that take you from $\sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}}$ to 1. Yet, this is one of the things that the computer must be taught to do, since otherwise it could not decide whether the linear equation $\left(\sqrt[3]{2+\sqrt{5}} + \sqrt[3]{2-\sqrt{5}} - 1\right) x = 0$ has one or infinitely many solutions.

Definition 2.50. A *commutative ring* means a set $\mathbb{K}$ equipped with the following additional data:

- a binary operation called "$+$" (that is, a function that takes two elements $a \in \mathbb{K}$ and $b \in \mathbb{K}$ as inputs, and outputs a new element of $\mathbb{K}$ which is denoted by $a+b$);
- a binary operation called "$\cdot$" (that is, a function that takes two elements $a \in \mathbb{K}$ and $b \in \mathbb{K}$ as inputs, and outputs a new element of $\mathbb{K}$ which is denoted by $a \cdot b$);
- an element of $\mathbb{K}$ called "0";
- an element of $\mathbb{K}$ called "1"

satisfying the following conditions ("axioms"):

- *Commutativity of addition:* We have $a+b=b+a$ for all $a \in \mathbb{K}$ and $b \in \mathbb{K}$.
- *Commutativity of multiplication:* We have $ab=ba$ for all $a \in \mathbb{K}$ and $b \in \mathbb{K}$. Here and in the following, ab is shorthand for $a \cdot b$ (as is usual for products of numbers).
- *Associativity of addition:* We have $a+(b+c)=(a+b)+c$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$.
- *Associativity of multiplication:* We have $a(bc)=(ab)c$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$.
- *Neutrality of 0:* We have $a+0=0+a=a$ for all $a \in \mathbb{K}$.
- *Existence of additive inverses:* For every $a \in \mathbb{K}$, there exists an element $a' \in \mathbb{K}$ such that $a+a'=a'+a=0$. This a' is commonly denoted by $-a$ and called the *additive inverse* of a. (It is easy to check that it is unique.)
- *Unitality (a.k.a. neutrality of 1):* We have $1a=a1=a$ for all $a \in \mathbb{K}$.
- *Annihilation:* We have $0a=a0=0$ for all $a \in \mathbb{K}$.
- *Distributivity:* We have $a(b+c)=ab+ac$ and $(a+b)c=ac+bc$ for all $a \in \mathbb{K}$, $b \in \mathbb{K}$ and $c \in \mathbb{K}$.

This definition was a mouthful, but its intention is rather simple: It defines a commutative ring as a set equipped with two operations which behave like addition and multiplication of numbers, and two elements which behave like the number 0 and the number 1. As a consequence, if we have a commutative ring $\mathbb{K}$, then matrices filled with elements of $\mathbb{K}$ will behave (at least with regard to their basic

properties, such as Proposition 2.20) like matrices filled with numbers.

Here are some examples of commutative rings:

- Each of the sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\overline{\mathbb{Q}}$ (endowed with the usual addition, the usual multiplication, the usual 0 and the usual 1) is a commutative ring.
- The set $\mathbb{Q}[x]$ of all polynomials (in the variable x) with rational coefficients (equipped with addition of polynomials, multiplication of polynomials, the polynomial 0 and the polynomial 1) is a ring.
- The set of all functions $\mathbb{R} \to \mathbb{R}$ (equipped with pointwise addition, pointwise multiplication, the constant-0 function and the constant-1 function) is a ring.
- If you know what "integers modulo n" are (for a given positive integer n): The integers modulo n (for a given n) also form a commutative ring.
- Here is a weirder example:

  For any two sets A and B, we let $A \triangle B$ denote the *symmetric difference* of A and B. This is the set of all elements which lie in exactly one of the two sets A and B. Thus, $A \triangle B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

  Fix some set S, and let $\mathcal{P}(S)$ denote the set of all subsets of S. The set $\mathcal{P}(S)$ equipped with the operation $\triangle$ (playing the role of "$+$"), the operation $\cap$ (playing the role of "$\cdot$"), the element[34] $\varnothing$ (playing the role of "0") and the element S (playing the role of "1") is a commutative ring. This is an example of a *Boolean ring* (and also an example of the fact that the operations "$+$" and "$\cdot$" don't always have anything to do with addition and multiplication of numbers!).

Many more examples of commutative rings can be found in textbooks on abstract algebra (e.g., [Artin10, Chapter 11]).

Matrices filled with elements of a commutative ring $\mathbb{K}$ are called *matrices over* $\mathbb{K}$. More precisely:

Definition 2.51. Let $\mathbb{K}$ be a commutative ring. If $n \in \mathbb{N}$ and $m \in \mathbb{N}$, then an $n \times m$*-matrix over* $\mathbb{K}$ simply means a rectangular table with n rows and m columns, such that each cell is filled with an element of $\mathbb{K}$.

Thus, matrices over $\mathbb{Q}$ are matrices with rational entries; matrices over $\mathbb{R}$ are matrices with real entries; matrices over $\mathbb{Z}$ are matrices with integer entries.

As we have already explained, matrices over commutative rings behave like matrices filled with numbers, at least as far as simple laws of computation are concerned. To wit, Proposition 2.20, Proposition 2.34 and Proposition 2.39 (as well as some other results that will be proven later) still hold if the matrices are filled with elements of a commutative ring $\mathbb{K}$ instead of numbers. Some other (deeper) results

[34] The notation $\varnothing$ stands for the empty set, i.e., the set $\{\}$.

might not hold for every commutative ring $\mathbb{K}$: For example, Gaussian elimination, which we will meet in the next chapter, requires that we can **divide** by nonzero elements of $\mathbb{K}$, but this is not always possible when $\mathbb{K}$ is a commutative ring. There is a word for that, too:

Definition 2.52. A commutative ring $\mathbb{K}$ is called a *field* if it satisfies the following two axioms:

- *Nontriviality:* We have $0 \neq 1$. (The "0" and "1" here, of course, are the two specific elements of $\mathbb{K}$ that we have chosen to call "0" and "1". They aren't always the same as the **numbers** 0 and 1. In particular, in a commutative ring $\mathbb{K}$ they can be equal; but for a field we want to disallow this.)
- *Existence of multiplicative inverses:* For every $a \in \mathbb{K}$, we have **either** $a = 0$, **or** there is an element $b \in \mathbb{K}$ satisfying $ab = ba = 1$.

The element b in the "existence of multiplicative inverses" axiom is called the *inverse* of a and is denoted by a^{-1}; it can be proven that this element is unique. If u and v are two elements of a field $\mathbb{K}$ such that $v \neq 0$, then the product uv^{-1} is denoted by $\dfrac{u}{v}$. Thus, any two elements of a field $\mathbb{K}$ can be divided by each other as long as the denominator is $\neq 0$.

For example, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\overline{\mathbb{Q}}$ are fields, but $\mathbb{Z}$ is not a field (because the integer 2 is nonzero, yet does not have an **integer** inverse). Some results about matrices are based on the possibility of dividing by nonzero numbers; these results cannot be directly generalized to matrices over a commutative ring $\mathbb{K}$. But most of them can be generalized to matrices over a field $\mathbb{K}$.

Remark 2.53. (a) The notion of a "commutative ring" is not standard, unfortunately. Some authors (e.g., Dummit and Foote in [DumFoo04, Part II], or Goodman in [Goodma15, Chapters 1 and 6]) omit the element "1" (and the unitality axiom), while some (older) authors even omit the associativity of multiplication. If you read any text on abstract algebra, it is prudent to check whether the author's concept of a commutative ring agrees with yours.

(b) As you might have guessed, there is also a notion of a "noncommutative ring". It is defined precisely as a "commutative ring", except that we omit the "commutativity of multiplication" axiom. ("Commutativity of addition" is left in!) It turns out that we already know a neat example of a noncommutative ring: For any commutative ring $\mathbb{K}$ and any $n \in \mathbb{N}$, the set of all $n \times n$-matrices over $\mathbb{K}$ is a noncommutative ring!

(c) The word "ring" (without the adjectives "commutative" and "noncommutative") usually means either "commutative ring" or "noncommutative ring", depending on the author's preferences.

3. Gaussian elimination

In this chapter, we shall take aim at understanding Gaussian elimination in terms of matrices. However, we will not head straight to this aim; instead, we will first introduce various classes of matrices (triangular matrices, matrix units, elementary matrices, permutation matrices), which will allow us to view certain pieces of the Gaussian elimination algorithm in isolation. Once we are done with that, we will finally explain Gaussian elimination in the general setting.

3.1. Linear equations and matrices

First of all, let us see what solving linear equations has to do with matrices.

Example 3.1. Consider the following system of equations in three unknowns x, y, z:

$$\begin{cases} 3x + 6y - z = 2; \\ 7x + 4y - 3z = 3; \\ -y + 8z = 1 \end{cases} . \tag{36}$$

I claim that this system of equations is equivalent to the single equation

$$\begin{pmatrix} 3 & 6 & -1 \\ 7 & 4 & -3 \\ 0 & -1 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \tag{37}$$

(this is an equation between two column vectors of size 3, so no wonder that it encodes a whole system of linear equations).

Why are (36) and (37) equivalent? Well, the left hand side of (37) is

$$\begin{pmatrix} 3 & 6 & -1 \\ 7 & 4 & -3 \\ 0 & -1 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3x + 6y + (-1) z \\ 7x + 4y + (-3) z \\ 0x + (-1) y + 8z \end{pmatrix} = \begin{pmatrix} 3x + 6y - z \\ 7x + 4y - 3z \\ -y + 8z \end{pmatrix}.$$

Thus, (37) is equivalent to

$$\begin{pmatrix} 3x + 6y - z \\ 7x + 4y - 3z \\ -y + 8z \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

But this is clearly equivalent to (36).

More generally:

Proposition 3.2. The system of m linear equations

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1; \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2; \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{cases}$$

in n unknowns $x_1, x_2, \ldots, x_n$ is equivalent to the vector equation

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \tag{38}$$

In other words, it is equivalent to the vector equation

$$Ax = b, \tag{39}$$

where

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

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

(Some authors write the A, the x and the b in (39) in boldface in order to stress that these are matrices, not numbers. We shall not.) The matrix A and the vector b are known; the vector x is what we want to find.

Thus, matrices give us a way to rewrite systems of linear equations as single equations between vectors. Moreover, as we will see, they give us a way to manipulate these equations easily.

To solve a vector equation like (39) means (in some sense) to "undo" a matrix multiplication. In fact, if we could divide by a matrix, then we could immediately solve $Ax = b$ by "dividing by A". Unfortunately, we cannot divide by a matrix in general. But the idea is fruitful: In fact, some matrices A are invertible (i.e., have an inverse A^{-1}), and for those matrices, we can transform $Ax = b$ into $x = A^{-1}b$, which gives us an explicit and unique solution for the system (38). This doesn't work for all A (since not all A are invertible), and is not a very practical way of

solving systems of linear equations; but the notion of invertible matrices is rather important, so we begin by studying them.

3.2. Inverse matrices

Definition 3.3. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix, and B be an $m \times n$-matrix.

(a) We say that B is a *right inverse* of A if $AB = I_n$.

(b) We say that B is a *left inverse* of A if $BA = I_m$.

(c) We say that B is an *inverse* of A if both $AB = I_n$ and $BA = I_m$.

Notice that we are saying "**a** right inverse" (not "**the** right inverse") in Definition 3.3, because a given matrix A can have several right inverses (but it can also have no right inverses at all). For the same reason, we are saying "**a** left inverse" (not "**the** left inverse"). However, when we are saying "**an** inverse" (not "**the** inverse"), we are just being cautious: We will later (in Corollary 3.7) see that A can never have several different inverses; thus, it would be legitimate to say "**the** inverse" as well. But as long as we have not proven this, we shall speak of "**an** inverse".

Example 3.4. (a) Let $A = (1, 4)$. (Recall that this means the 1×2-matrix $\begin{pmatrix} 1 & 4 \end{pmatrix}$.) When is a matrix B a right inverse of A ?

First, if B is a right inverse of A, then B must be a 2×1-matrix (since any right inverse of an $n \times m$-matrix has to be an $m \times n$-matrix). So let us assume that B is a 2×1-matrix. Thus, B must have the form $B = \begin{pmatrix} u \\ v \end{pmatrix}$ for some numbers u and v. Then, $AB = \begin{pmatrix} 1 & 4 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1u + 4v \end{pmatrix}$. In order for B to be a right inverse of A, it is necessary and sufficient that $AB = I_1$ (because this is how we defined "right inverse"). In other words, we must have $\begin{pmatrix} 1u + 4v \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix}$ (since $AB = \begin{pmatrix} 1u + 4v \end{pmatrix}$ and $I_1 = \begin{pmatrix} 1 \end{pmatrix}$). In other words, we must have $1u + 4v = 1$.

Hence, a matrix B is a right inverse of A if and only if it has the form $B = \begin{pmatrix} u \\ v \end{pmatrix}$ for some numbers u and v satisfying $1u + 4v = 1$. How do we find two such numbers u and v ? Well, we can view $1u + 4v = 1$ as a system of 1 linear equation in 2 variables, but actually we can just read off the solution: v can be chosen arbitrarily, and u then has to be $1 - 4v$. Hence, a matrix B is a right inverse of A if and only if it has the form $B = \begin{pmatrix} 1 - 4v \\ v \end{pmatrix}$ for some number v. In particular, there are **infinitely many** matrices B that are right inverses of A (because we have full freedom in choosing v).

(b) Let $A = (1, 4)$ again. When is a matrix B a left inverse of A ?

Again, B must be a 2×1-matrix in order for this to have any chance of being true. So let us assume that B is a 2×1-matrix, and write B in the form $B = \begin{pmatrix} u \\ v \end{pmatrix}$

for some numbers u and v. Then, $BA = \begin{pmatrix} u \\ v \end{pmatrix} (1,4) = \begin{pmatrix} u \cdot 1 & u \cdot 4 \\ v \cdot 1 & v \cdot 4 \end{pmatrix}$. In order for B to be a left inverse of A, it is necessary and sufficient that $BA = I_2$ (because this is how we defined "left inverse"). In other words, we must have $\begin{pmatrix} u \cdot 1 & u \cdot 4 \\ v \cdot 1 & v \cdot 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ (since $BA = \begin{pmatrix} u \cdot 1 & u \cdot 4 \\ v \cdot 1 & v \cdot 4 \end{pmatrix}$ and $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$). In other words, we must have

$$\begin{cases} u \cdot 1 = 1; \\ u \cdot 4 = 0; \\ v \cdot 1 = 0; \\ v \cdot 4 = 1 \end{cases}.$$

But this cannot happen! Indeed, the equations $u \cdot 1 = 1$ and $u \cdot 4 = 0$ contradict each other (because if $u \cdot 1 = 1$, then $u \cdot 4$ must be 4). Hence, B can never be a left inverse of A. In other words, the matrix A **has no** left inverse.

(c) Let $A = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. When is a matrix B a left inverse of A ? When is a matrix B a right inverse of A ? I will let you figure this out (see Exercise 3.5 below).

(d) Let $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. When is a matrix B a left inverse of A ? When is a matrix B a right inverse of A ?

A left inverse of A would have to be a 2×2-matrix. A 2×2-matrix $B = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ is a left inverse of A if and only if it satisfies $BA = I_2$, that is, $\begin{pmatrix} x & y \\ z & w \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, or, equivalently, $\begin{pmatrix} x+y & -x+y \\ w+z & w-z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, or, equivalently,

$$\begin{cases} x+y = 1; \\ -x+y = 0; \\ w+z = 0; \\ w-z = 1 \end{cases}.$$

This is a system of four linear equations in the four unknowns x, y, z, w; it has the unique solution

$$(x, y, z, w) = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right).$$

Thus, a 2×2-matrix $B = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$ is a left inverse of A if and only if $(x, y, z, w) = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right)$. Hence, there exists exactly one left inverse of A, and this left inverse is $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

A similar computation reveals that there exists exactly one right inverse of A, and this right inverse is $\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$. So the unique left inverse of A and the unique right inverse of A are actually equal (and thus are an inverse of A). This might not be clear from the definitions, but as we shall soon see, this is not a coincidence.

(e) Generalizing Example 3.4 **(d)**, we might wonder when a 2×2-matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has a left inverse, a right inverse or an inverse. For any given four values of a, b, c, d, we can answer this question similarly to how we answered it for the matrix $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ in Example 3.4 **(d)** (by solving a system of linear equations). The procedure will depend on whether some numbers are zero or not. (For example, we might want to divide the equation $bx + dy = 0$ by b, which requires $b \neq 0$; the case $b = 0$ will then have to be treated separately.) But the final result will be the following:

- If $ad - bc = 0$, then the matrix A has no left inverses and no right inverses.
- If $ad - bc \neq 0$, then the matrix A has a unique inverse, which is also the unique left inverse and the unique right inverse. This inverse is $\begin{pmatrix} \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$.

Again, this phenomenon of the left inverse equalling the right inverse appears. Notably, the number $ad - bc$ plays an important role here; we will later see more of it (it is an example of a *determinant*).

Exercise 3.5. Let $A = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. When is a matrix B a left inverse of A ? When is a matrix B a right inverse of A ?

As we know from Example 3.4 **(a)**, a matrix may have infinitely many right inverses. Similarly, a matrix may have infinitely many left inverses. But can a matrix have both infinitely many right inverses and infinitely many left inverses at the same time? The answer is "no", and in fact, something stronger is true:

Proposition 3.6. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let L be a left inverse of A. Let R be a right inverse of A. Then:
(a) We have $L = R$.
(b) The matrix L is the only left inverse of A.

(c) The matrix R is the only right inverse of A.
(d) The matrix $L = R$ is the only inverse of A.

Proof of Proposition 3.6. We know that $LA = I_m$ (since L is a left inverse of A) and that $AR = I_n$ (since R is a right inverse of A).

(a) Consider the product LAR. (Recall that this product is well-defined, because Proposition 2.20 **(g)** yields $L(AR) = (LA)R$.)

One way to rewrite LAR is as follows:

$$L\underbrace{AR}_{=I_n} = LI_n = L \qquad \text{(by Proposition 2.39 \textbf{(b)})}. \tag{40}$$

Another way is

$$\underbrace{LA}_{=I_m} R = I_m R = R \qquad \text{(by Proposition 2.39 \textbf{(a)})}. \tag{41}$$

Comparing (40) with (41), we obtain $L = R$. This proves Proposition 3.6 **(a)**.

(b) Let L' be any left inverse of A. Then, we can apply Proposition 3.6 **(a)** to L' instead of L (because all that was needed from L in Proposition 3.6 **(a)** was that it be a left inverse of A). As a result, we obtain $L' = R$.

Now, forget that we fixed L'. We thus have shown that if L' is any left inverse of A, then $L' = R$. In other words, any left inverse of A equals R. Thus, there exists at most one left inverse of A. Therefore, the matrix L is the only left inverse of A (since we already know that L is a left inverse of A). This proves Proposition 3.6 **(b)**.

(c) The proof of Proposition 3.6 **(c)** is analogous to the proof of Proposition 3.6 **(b)**. (We again need to apply Proposition 3.6 **(a)**, but this time, instead of a left inverse L', we have to introduce a right inverse R'. The details are left to the reader.)

(d) Proposition 3.6 **(a)** yields $L = R$. Hence, $A\underbrace{L}_{=R} = AR = I_n$.

Now, the matrix L is an inverse of A (since $LA = I_m$ and $AL = I_n$). In other words, the matrix $L = R$ is an inverse of A (since $L = R$). It remains to show that it is the only inverse of A. But this is easy: Let L' be any inverse of A. Then, $L'A = I_m$, so that L' is a left inverse of A. Proposition 3.6 **(a)** (applied to L' instead of L) therefore yields $L' = R$.

Now, forget that we fixed L'. We thus have shown that if L' is any inverse of A, then $L' = R$. In other words, any inverse of A equals R. Thus, there exists at most one inverse of A. Therefore, the matrix L is the only inverse of A (since we already know that L is an inverse of A). This proves Proposition 3.6 **(d)**. □

Corollary 3.7. Let A be a matrix. Then, A has at most one inverse.

Proof of Corollary 3.7. We need to show that any two inverses of A are equal. So let L and R be two inverses of A. We must show that $L = R$.

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that A is an $n \times m$-matrix. The matrix L is an inverse of A, thus satisfies $LA = I_m$. Hence, L is a left inverse of A. Also, the matrix R is an inverse of A, thus satisfies $AR = I_n$. Hence, R is a right inverse of A. Thus, Proposition 3.6 **(a)** shows that $L = R$. This proves Corollary 3.7. □

Definition 3.8. (a) A matrix A is said to be *invertible* if it has an inverse. (Similarly, we can define the words "left-invertible" and "right-invertible".)

(b) Let A be an invertible matrix. Then, A has an inverse. Due to Corollary 3.7, we furthermore know that A has at most one inverse. Thus, A has exactly one inverse. We can thus refer to this inverse as "**the** inverse of A" (not just "**an** inverse of A"), and denote it by A^{-1}. If A is an $n \times m$-matrix, then this inverse satisfies $A^{-1}A = I_m$ and $AA^{-1} = I_n$ (by its definition).

Notice that the equalities $A^{-1}A = I_m$ and $AA^{-1} = I_n$ show that a matrix A and its inverse A^{-1} cancel each other when they stand adjacent in a product: for example, $BA^{-1}AC$ simplifies to BC. However, they do not (generally) cancel each other when they appear apart from one another: for example, $BA^{-1}CA$ does **not** simplify to BC.

So what matrices are invertible? The following theorem significantly narrows the search down; we shall not prove it until later:

Theorem 3.9. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.

(a) If A has a right inverse, then $n \leq m$ (that is, the matrix A has at least as many columns as it has rows).

(b) If A has a left inverse, then $n \geq m$ (that is, the matrix A has at most as many columns as it has rows).

(c) If A is invertible (i.e., has an inverse), then $n = m$ (that is, the matrix A is square).

(d) If A is square (that is, $n = m$) and has a left inverse **or** a right inverse, then A is actually invertible (and so this left or right inverse is the inverse of A). Notice that this is false for rectangular matrices!

Let us now check some simpler facts about inverses:

Proposition 3.10. Let $n \in \mathbb{N}$. Then, the matrix I_n is invertible, and its inverse is $(I_n)^{-1} = I_n$.

Proof of Proposition 3.10. We have $I_n I_n = I_n$ and $I_n I_n = I_n$. Hence, the matrix I_n is an inverse of I_n (by the definition of "inverse"). This proves Proposition 3.10. □

Proposition 3.11. Let A and B be two invertible matrices such that the product AB is well-defined (i.e., such that A has as many columns as B has rows). Then, the matrix AB is also invertible, and its inverse is

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{42}$$

Proof of Proposition 3.11. Let n, m and p be nonnegative integers such that A is an $n \times m$-matrix and B is an $m \times p$-matrix[35]. (Actually, Theorem 3.9 **(c)** reveals that the matrices A and B are square and therefore $n = m = p$; but I do not want to use Theorem 3.9 **(c)** here, since I have not yet proven it.)

Recall once again that (by general associativity) products of matrices can be written without parentheses. Thus, for example, the products $B^{-1}A^{-1}AB$ and $ABB^{-1}A^{-1}$ make sense. Let us simplify these products:

$$B^{-1}\underbrace{A^{-1}A}_{=I_m}B = B^{-1}I_mB = B^{-1}B = I_p$$

and

$$A\underbrace{BB^{-1}}_{=I_m}A^{-1} = AI_mA^{-1} = AA^{-1} = I_n.$$

But these two equalities say precisely that $B^{-1}A^{-1}$ is an inverse of AB. (If you don't believe me, rewrite them with parentheses: $\left(B^{-1}A^{-1}\right)(AB) = I_p$ and $(AB)\left(B^{-1}A^{-1}\right) = I_n$.) In particular, this shows that AB is invertible. This proves Proposition 3.11. $\square$

In words, (42) says that the inverse of a product of two matrices is the product of their inverses, but **in opposite order**. This takes some getting used to, but is really a natural thing; the same rule holds for inverting the composition of functions[36].

Proposition 3.12. Let $A_1, A_2, \ldots, A_k$ be k invertible matrices (where k is a positive integer) such that the product $A_1A_2\cdots A_k$ is well-defined (i.e., such that A_i has as many columns as A_{i+1} has rows, for each $i < k$). Then, the matrix $A_1A_2\cdots A_k$ is invertible, and its inverse is

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}.$$

Proposition 3.12 is a natural extension of Proposition 3.11 to products of more than 2 matrices. The proof of Proposition 3.12 is straightforward, and I am only showing it as an example of proof by induction:

Proof of Proposition 3.12. We prove Proposition 3.12 by induction on k:

Induction base: If $k = 1$, then Proposition 3.12 says that $A_1^{-1} = A_1^{-1}$; this is obviously true. Hence, Proposition 3.12 holds for $k = 1$. This completes the induction base.

[35] We can indeed find such n, m and p because A has as many columns as B has rows.

[36] Namely: If X, Y and Z are three sets, and if $b : X \to Y$ and $a : Y \to Z$ are two invertible functions (i.e., bijections), then $a \circ b : X \to Z$ is an invertible function as well, and its inverse is $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$. (Some authors liken this to the fact that if you want to undo the process of putting on socks and then putting on shoes, you have to first take off your shoes and then take off your socks. See `https://proofwiki.org/wiki/Inverse_of_Product` .)

Induction step: Let ℓ be a positive integer. Assume (as our *induction hypothesis*) that Proposition 3.12 holds for $k = \ell$. In other words, for any ℓ invertible matrices $A_1, A_2, \ldots, A_\ell$ for which the product $A_1 A_2 \cdots A_\ell$ is well-defined, the matrix $A_1 A_2 \cdots A_\ell$ is invertible, and its inverse is

$$(A_1 A_2 \cdots A_\ell)^{-1} = A_\ell^{-1} A_{\ell-1}^{-1} \cdots A_1^{-1}.$$

We must now show that Proposition 3.12 also holds for $k = \ell + 1$. So let us fix $\ell + 1$ invertible matrices $A_1, A_2, \ldots, A_{\ell+1}$ for which the product $A_1 A_2 \cdots A_{\ell+1}$ is well-defined. We must then show that the matrix $A_1 A_2 \cdots A_{\ell+1}$ is invertible, and that its inverse is

$$(A_1 A_2 \cdots A_{\ell+1})^{-1} = A_{\ell+1}^{-1} A_\ell^{-1} \cdots A_1^{-1}.$$

The product $A_1 A_2 \cdots A_\ell$ is well-defined (since the product $A_1 A_2 \cdots A_{\ell+1}$ is well-defined). Hence, we can apply our induction hypothesis, and conclude that the matrix $A_1 A_2 \cdots A_\ell$ is invertible, and its inverse is

$$(A_1 A_2 \cdots A_\ell)^{-1} = A_\ell^{-1} A_{\ell-1}^{-1} \cdots A_1^{-1}.$$

Now, the matrices $A_1 A_2 \cdots A_\ell$ and $A_{\ell+1}$ are invertible, and their product $(A_1 A_2 \cdots A_\ell) A_{\ell+1} = A_1 A_2 \cdots A_{\ell+1}$ is well-defined (by assumption). Hence, Proposition 3.11 (applied to $A = A_1 A_2 \cdots A_\ell$ and $B = A_{\ell+1}$) shows that the matrix $(A_1 A_2 \cdots A_\ell) A_{\ell+1}$ is also invertible, and its inverse is

$$((A_1 A_2 \cdots A_\ell) A_{\ell+1})^{-1} = A_{\ell+1}^{-1} (A_1 A_2 \cdots A_\ell)^{-1}.$$

Since $(A_1 A_2 \cdots A_\ell) A_{\ell+1} = A_1 A_2 \cdots A_{\ell+1}$ and $A_{\ell+1}^{-1} \underbrace{(A_1 A_2 \cdots A_\ell)^{-1}}_{= A_\ell^{-1} A_{\ell-1}^{-1} \cdots A_1^{-1}} = A_{\ell+1}^{-1} \left(A_\ell^{-1} A_{\ell-1}^{-1} \cdots A_1^{-1} \right) = A_{\ell+1}^{-1} A_\ell^{-1} \cdots A_1^{-1}$, this rewrites as follows: The matrix $A_1 A_2 \cdots A_{\ell+1}$ is invertible, and its inverse is

$$(A_1 A_2 \cdots A_{\ell+1})^{-1} = A_{\ell+1}^{-1} A_\ell^{-1} \cdots A_1^{-1}.$$

This is precisely what we wanted to show! Thus, Proposition 3.12 holds for $k = \ell + 1$. This completes the induction step. Thus, Proposition 3.12 is proven by induction.

(I have written up this proof with a lot of detail. You do not have to! If you are used to mathematical induction, then you can easily afford omitting many of the incantations I made above, and taking certain shortcuts – for example, instead of introducing a new variable ℓ in the induction step, you could reuse k, thus stepping "from k to $k+1$" instead of "from $k = \ell$ to $k = \ell + 1$". You also don't need to formally state the induction hypothesis, because it is just a copy of the claim (with k replaced by ℓ in our case). Finally, what we did in our proof was obvious enough that you could just say that "Proposition 3.12 follows by a straightforward induction on k, where Proposition 3.11 is being applied in the induction step", and declare the proof finished.) $\square$

Remark 3.13. It is common to define the product of 0 square matrices of size $n \times n$ (an "empty product of $n \times n$-matrices") as the identity matrix I_n (similarly to how a product of 0 numbers is defined to be 1). With this convention, Proposition 3.12 holds for $k = 0$ too (it then states that $(I_n)^{-1} = I_n$), as long as we agree what size our non-existing matrices are considered to have (it has to be $n \times n$ for some $n \in \mathbb{N}$). With this convention, we could have started our induction (in the above proof of Proposition 3.12) at $k = 0$ instead of $k = 1$.

Corollary 3.14. Let $n \in \mathbb{N}$. Let $k \in \mathbb{N}$. Let A be an invertible $n \times n$-matrix. Then, A^k is also invertible, and its inverse is $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$.

Note that Corollary 3.14 is not obvious! You cannot argue that $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$ because both sides simplify to A^{-k}; this argument makes no sense unless you have defined A^{-k} (and we have not defined A^{-k}) and proved that standard rules of exponentiation (such as $(A^u)^v = A^{uv}$) apply to matrices.

Proof of Corollary 3.14. Recall that $A^0 = I_n$ (by the definition of A^0). Hence, in the case when $k = 0$, Corollary 3.14 says that I_n is invertible, and its inverse is $(I_n)^{-1} = I_n$. This follows from Proposition 3.10. Thus, Corollary 3.14 is proven in the case when $k = 0$. Therefore, we can WLOG[37] assume that $k \neq 0$. Assume this. Thus, k is a positive integer. Hence, $A^k = \underbrace{AA\cdots A}_{k \text{ times}}$ and $\left(A^{-1}\right)^k = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{k \text{ times}}$.

Proposition 3.12 (applied to $A, A, \ldots, A$ instead of $A_1, A_2, \ldots, A_k$) shows that the matrix $\underbrace{AA\cdots A}_{k \text{ times}}$ is invertible, and its inverse is

$$\left(\underbrace{AA\cdots A}_{k \text{ times}}\right)^{-1} = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{k \text{ times}}.$$

In other words, the matrix A^k is invertible, and its inverse is $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k$. This proves Corollary 3.14. □

Proposition 3.15. Let $n \in \mathbb{N}$. Let A be an invertible $n \times n$-matrix. Then, its inverse A^{-1} is also invertible, and has the inverse $\left(A^{-1}\right)^{-1} = A$.

[37] "WLOG" is shorthand for "without loss of generality". See, for example, the Wikipedia article for "WLOG" (or any book on mathematical proofs) for the meaning of this phrase.

(As far as the proof of Corollary 3.14 is concerned, the meaning of "we can WLOG assume that $k \neq 0$" is the following: "If we can prove Corollary 3.14 for $k \neq 0$, then we know how to obtain a proof of Corollary 3.14 for all k (because Corollary 3.14 is already proven in the case when $k = 0$). Thus, it will suffice to prove Corollary 3.14 for $k \neq 0$; hence, let us assume that $k \neq 0$.")

Proof of Proposition 3.15. Since A^{-1} is an inverse of A, we have the two equalities $A^{-1}A = I_n$ and $AA^{-1} = I_n$. But these very same equalities show that A is an inverse of A^{-1} (if you do not trust me, just check with the definition of "inverse"). Thus, the matrix A^{-1} is invertible, and its inverse is $\left(A^{-1}\right)^{-1} = A$. Proposition 3.15 is proven. □

Proposition 3.16. Let $n \in \mathbb{N}$. Let λ be a nonzero number. Let A be an invertible $n \times n$-matrix. Then, the matrix λA is also invertible, and its inverse is $(\lambda A)^{-1} = \lambda^{-1} A^{-1} = \dfrac{1}{\lambda} A^{-1}$.

Exercise 3.17. Prove Proposition 3.16.

3.3. More on transposes

How do the matrix operations we have seen above (addition, multiplication, inversion, etc.) behave with respect to transposes? The answer is "fairly nicely":

Proposition 3.18. **(a)** Let $n \in \mathbb{N}$. Then, $(I_n)^T = I_n$.

(b) Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, $(0_{n\times m})^T = 0_{m\times n}$.

(c) Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let λ be a number. Then, $(\lambda A)^T = \lambda A^T$.

(d) Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A and B be two $n \times m$-matrices. Then, $(A+B)^T = A^T + B^T$.

(e) Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let B be an $m \times p$-matrix. Then, $(AB)^T = B^T A^T$.

(f) Let $n \in \mathbb{N}$. Let A be an invertible $n \times n$-matrix. Then, A^T is invertible, and its inverse is $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Notice that the right hand side in Proposition 3.18 **(e)** is $B^T A^T$, not $A^T B^T$ (in fact, $A^T B^T$ does not always make sense, since the number of columns of A^T is not necessarily the number of rows of B^T). This is similar to the $B^{-1}A^{-1}$ in Proposition 3.11.

Proposition 3.18 is fairly easy to show; let us only give a proof of part **(e)**:

Proof of Proposition 3.18 (e). The definition of A^T shows that $A^T = \left(A_{j,i}\right)_{1\le i\le m,\ 1\le j\le n}$. Thus,

$$\left(A^T\right)_{i,j} = A_{j,i} \qquad \text{for all } i \in \{1,2,\ldots,m\} \text{ and } j \in \{1,2,\ldots,n\}. \tag{43}$$

Similarly, the definition of B^T shows that $B^T = \left(B_{j,i}\right)_{1\le i\le p,\ 1\le j\le m}$. Hence,

$$\left(B^T\right)_{i,j} = B_{j,i} \qquad \text{for all } i \in \{1,2,\ldots,p\} \text{ and } j \in \{1,2,\ldots,m\}. \tag{44}$$

Now, B^T is a $p \times m$-matrix, while A^T is an $m \times n$-matrix. Hence, $B^T A^T$ is a $p \times n$-matrix. Also, $(AB)^T$ is a $p \times n$-matrix (since AB is an $n \times p$-matrix). The definition of $(AB)^T$ shows that $(AB)^T = \left((AB)_{j,i}\right)_{1 \leq i \leq p,\ 1 \leq j \leq n}$. Hence,

$$\begin{aligned}\left((AB)^T\right)_{i,j} &= (AB)_{j,i} \\ &= A_{j,1}B_{1,i} + A_{j,2}B_{2,i} + \cdots + A_{j,m}B_{m,i} \qquad (45) \\ &\qquad \text{(by Proposition 2.19 \textbf{(a)}, applied to } j \text{ and } i \text{ instead of } i \text{ and } j)\end{aligned}$$

for all $i \in \{1, 2, \ldots, p\}$ and $j \in \{1, 2, \ldots, n\}$.

On the other hand, we can apply Proposition 2.19 **(a)** to p, m, n, B^T and A^T instead of n, m, p, A and B. We thus conclude that

$$\left(B^T A^T\right)_{i,j} = \left(B^T\right)_{i,1}\left(A^T\right)_{1,j} + \left(B^T\right)_{i,2}\left(A^T\right)_{2,j} + \cdots + \left(B^T\right)_{i,m}\left(A^T\right)_{m,j} \qquad (46)$$

for all $i \in \{1, 2, \ldots, p\}$ and $j \in \{1, 2, \ldots, n\}$.

But for every $k \in \{1, 2, \ldots, m\}$, we have

$$\underbrace{\left(B^T\right)_{i,k}}_{\substack{=B_{k,i} \\ \text{(by (44), applied} \\ \text{to } k \text{ instead of } j)}} \underbrace{\left(A^T\right)_{k,j}}_{\substack{=A_{j,k} \\ \text{(by (43), applied} \\ \text{to } k \text{ instead of } i)}} = B_{k,i}A_{j,k} = A_{j,k}B_{k,i}. \qquad (47)$$

Hence, for all $i \in \{1, 2, \ldots, p\}$ and $j \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned}\left(B^T A^T\right)_{i,j} &= \underbrace{\left(B^T\right)_{i,1}\left(A^T\right)_{1,j}}_{\substack{=A_{j,1}B_{1,i} \\ \text{(by (47), applied} \\ \text{to } k=1)}} + \underbrace{\left(B^T\right)_{i,2}\left(A^T\right)_{2,j}}_{\substack{=A_{j,2}B_{2,i} \\ \text{(by (47), applied} \\ \text{to } k=2)}} + \cdots + \underbrace{\left(B^T\right)_{i,m}\left(A^T\right)_{m,j}}_{\substack{=A_{j,m}B_{m,i} \\ \text{(by (47), applied} \\ \text{to } k=m)}} \qquad \text{(by (46))} \\ &= A_{j,1}B_{1,i} + A_{j,2}B_{2,i} + \cdots + A_{j,m}B_{m,i}.\end{aligned}$$

Comparing this with (45), we conclude that $\left((AB)^T\right)_{i,j} = \left(B^T A^T\right)_{i,j}$ for all $i \in \{1, 2, \ldots, p\}$ and $j \in \{1, 2, \ldots, n\}$. In other words, each entry of the matrix $(AB)^T$ equals the corresponding entry of the matrix $B^T A^T$. Thus, $(AB)^T = B^T A^T$. This proves Proposition 3.18 **(e)**. □

Exercise 3.19. Prove Proposition 3.18 **(f)**.

3.4. Triangular matrices

We next discuss some particular classes of matrices: the so-called *triangular matrices* and some of their variations.

Definition 3.20. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix.

(a) We say that the matrix A is *upper-triangular* if and only if we have

$$A_{i,j} = 0 \qquad \text{whenever } i > j.$$

(Of course, "whenever $i > j$" is shorthand for "for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ satisfying $i > j$".)

(b) We say that the matrix A is *lower-triangular* if and only if we have

$$A_{i,j} = 0 \qquad \text{whenever } i < j.$$

(c) We say that the matrix A is *diagonal* if and only if we have

$$A_{i,j} = 0 \qquad \text{whenever } i \neq j.$$

Notice that only square matrices can be upper-triangular or lower-triangular or diagonal (by definition). Why the name "triangular"? Because visually speaking, a matrix is upper-triangular if and only if all its entries (strictly) below the diagonal are 0 (which means that its nonzero entries are concentrated in the **triangle** bordered by the diagonal, the upper rim and the right rim). I hope Example 3.21 will clarify this if it is unclear. Similarly, a matrix is lower-triangular if and only if all its entries (strictly) above the diagonal are 0 (which again means that its nonzero entries are concentrated in a triangle, this time to the southwest of the diagonal). Finally, a matrix is diagonal if and only if all its entries except for the diagonal entries are 0.

I think the following example should explain this:

Example 3.21. (a) A 4×4-matrix is upper-triangular if and only if it has the form $\begin{pmatrix} a & b & c & d \\ 0 & b' & c' & d' \\ 0 & 0 & c'' & d'' \\ 0 & 0 & 0 & d''' \end{pmatrix}$ for some numbers $a, b, c, d, b', c', d', c'', d'', d'''$. Notice that we are making no requirements on these numbers; in particular, they **can** be 0. Upper-triangularity means that $A_{i,j} = 0$ whenever $i > j$; it does **not** require that $A_{i,j} \neq 0$ in all other cases.

(b) A 4×4-matrix is lower-triangular if and only if it has the form $\begin{pmatrix} a & 0 & 0 & 0 \\ a' & b' & 0 & 0 \\ a'' & b'' & c'' & 0 \\ a''' & b''' & c''' & d''' \end{pmatrix}$ for some numbers $a, a', b', a'', b'', c'', a''', b''', c''', d'''$.

(c) A 4×4-matrix is diagonal if and only if it has the form $\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b' & 0 & 0 \\ 0 & 0 & c'' & 0 \\ 0 & 0 & 0 & d''' \end{pmatrix}$

for some numbers a, b', c'', d'''.

Here is something obvious:

Proposition 3.22. Let $n \in \mathbb{N}$.
(a) An $n \times n$-matrix A is diagonal if and only if A is both upper-triangular and lower-triangular.
(b) The zero matrix $0_{n \times n}$ and the identity matrix I_n are upper-triangular, lower-triangular and diagonal.

A less trivial fact is that the product of two upper-triangular matrices is upper-triangular again. We shall show this, and a little bit more, in the following theorem:

Theorem 3.23. Let $n \in \mathbb{N}$. Let A and B be two upper-triangular $n \times n$-matrices.
(a) Then, AB is an upper-triangular $n \times n$-matrix.
(b) The diagonal entries of AB are

$$(AB)_{i,i} = A_{i,i} B_{i,i} \qquad \text{for all } i \in \{1, 2, \ldots, n\}.$$

(c) Also, $A + B$ is an upper-triangular $n \times n$-matrix. Furthermore, λA is an upper-triangular matrix whenever λ is a number.

Note that Theorem 3.23 **(b)** says that each diagonal entry of AB is the product of the corresponding diagonal entries of A and of B. Thus, **in this specific case**, the product AB does behave as if matrices were multiplied entry by entry (but only for its diagonal entries). Before I prove Theorem 3.23, let me give an example:

Example 3.24. Let $A = \begin{pmatrix} a & b & c \\ 0 & b' & c' \\ 0 & 0 & c'' \end{pmatrix}$ and $B = \begin{pmatrix} x & y & z \\ 0 & y' & z' \\ 0 & 0 & z'' \end{pmatrix}$ be two upper-triangular 3×3-matrices. Then,

$$AB = \begin{pmatrix} a & b & c \\ 0 & b' & c' \\ 0 & 0 & c'' \end{pmatrix} \begin{pmatrix} x & y & z \\ 0 & y' & z' \\ 0 & 0 & z'' \end{pmatrix} = \begin{pmatrix} ax & ay + by' & az + bz' + cz'' \\ 0 & b'y' & b'z' + c'z'' \\ 0 & 0 & c''z'' \end{pmatrix}.$$

Thus, AB is again upper-triangular (as Theorem 3.23 **(a)** predicts), and the diagonal entries $ax, b'y', c''z''$ of AB are the products of the respective entries of A and of B (as Theorem 3.23 **(b)** predicts).

Proof of Theorem 3.23. The matrix A is upper-triangular. In other words,

$$A_{i,j} = 0 \qquad \text{whenever } i > j \tag{48}$$

(because this is what it means for A to be upper-triangular). Similarly,

$$B_{i,j} = 0 \qquad \text{whenever } i > j \tag{49}$$

(because B, too, is upper-triangular).

Now, fix two elements i and j of $\{1,2,\ldots,n\}$ satisfying $i > j$. We shall prove that for every $k \in \{1,2,\ldots,n\}$, we have

$$A_{i,k}B_{k,j} = 0. \tag{50}$$

[*Proof of (50):* Let $k \in \{1,2,\ldots,n\}$. Then, we are in one of the following two cases:

Case 1: We have $i \leq k$.

Case 2: We have $i > k$.

We shall prove (50) in each of these two cases separately:

1. Let us first consider Case 1. In this case, we have $i \leq k$. Thus, $k \geq i$, so that $k \geq i > j$. Hence, we can apply (49) to k instead of i. As a result, we obtain $B_{k,j} = 0$. Hence, $A_{i,k}\underbrace{B_{k,j}}_{=0} = A_{i,k}0 = 0$. Thus, (50) is proven in Case 1.

2. Let us now consider Case 2. In this case, we have $i > k$. Hence, we can apply (48) to k instead of j. As a result, we obtain $A_{i,k} = 0$. Hence, $\underbrace{A_{i,k}}_{=0}B_{k,j} = 0B_{k,j} = 0$. Thus, (50) is proven in Case 2.

We have now proven (50) in both Cases 1 and 2. Thus, (50) is proven.]

Now, Proposition 2.19 **(a)** shows that

$$\begin{aligned}(AB)_{i,j} = &\underbrace{A_{i,1}B_{1,j}}_{\substack{=0\\ \text{(by (50),}\\ \text{applied to } k=1)}} + \underbrace{A_{i,2}B_{2,j}}_{\substack{=0\\ \text{(by (50),}\\ \text{applied to } k=2)}} + \cdots + \underbrace{A_{i,m}B_{m,j}}_{\substack{=0\\ \text{(by (50),}\\ \text{applied to } k=m)}}\\ &= 0+0+\cdots+0 = 0.\end{aligned}$$

Now, forget that we fixed i and j. We thus have shown that

$$(AB)_{i,j} = 0 \qquad \text{whenever } i > j. \tag{51}$$

But this says precisely that AB is upper-triangular. Thus, Theorem 3.23 **(a)** is proven.

(b) Let $i \in \{1,2,\ldots,n\}$. We must prove that $(AB)_{i,i} = A_{i,i}B_{i,i}$.

In a sense, this is similar to how we proved (51), but a little bit more complicated.

We first observe that, for every $k \in \{1,2,\ldots,n\}$ satisfying $k \neq i$, we have

$$A_{i,k}B_{k,i} = 0. \tag{52}$$

The proof of this will be very similar to the proof of (50), with j replaced by i:

[*Proof of (52):* Let $k \in \{1,2,\ldots,n\}$ be such that $k \neq i$. Then, we are in one of the following two cases:

Case 1: We have $i \leq k$.

Case 2: We have $i > k$.

We shall prove (52) in each of these two cases separately:

1. Let us first consider Case 1. In this case, we have $i \leq k$. Thus, $k \geq i$, so that $k > i$ (because $k \neq i$). Hence, we can apply (49) to k and i instead of i and j. As a result, we obtain $B_{k,i} = 0$. Hence, $A_{i,k} \underbrace{B_{k,i}}_{=0} = A_{i,k}0 = 0$. Thus, (52) is proven in Case 1.

2. Let us now consider Case 2. In this case, we have $i > k$. Hence, we can apply (48) to k instead of j. As a result, we obtain $A_{i,k} = 0$. Hence, $\underbrace{A_{i,k}}_{=0} B_{k,i} = 0B_{k,i} = 0$. Thus, (52) is proven in Case 2.

We have now proven (52) in both Cases 1 and 2. Thus, (52) is proven.]
Now, Proposition 2.19 **(a)** (applied to $j = i$) shows that

$$\begin{aligned}
&(AB)_{i,i} \\
&= A_{i,1}B_{1,i} + A_{i,2}B_{2,i} + \cdots + A_{i,m}B_{m,i} \\
&= (\text{the sum of the terms } A_{i,k}B_{k,i} \text{ for all } k \in \{1,2,\ldots,n\}) \\
&= A_{i,i}B_{i,i} + \underbrace{(\text{the sum of the terms } A_{i,k}B_{k,i} \text{ for all } k \in \{1,2,\ldots,n\} \text{ satisfying } k \neq i)}_{\substack{=0 \\ \text{(because (52) shows that all of these terms are 0)}}} \\
&\qquad (\text{here, we have taken the term } A_{i,i}B_{i,i} \text{ out of the sum}) \\
&= A_{i,i}B_{i,i} + 0 = A_{i,i}B_{i,i}.
\end{aligned}$$

This proves Theorem 3.23 **(b)**.

(c) Theorem 3.23 **(c)** is straightforward to check (due to the simple definitions of $A + B$ and λA); the details are left to the reader. □

The natural analogue of Theorem 3.23 for lower-triangular matrices also holds:

Theorem 3.25. Let $n \in \mathbb{N}$. Let A and B be two lower-triangular $n \times n$-matrices.
(a) Then, AB is a lower-triangular $n \times n$-matrix.
(b) The diagonal entries of AB are

$$(AB)_{i,i} = A_{i,i}B_{i,i} \qquad \text{for all } i \in \{1,2,\ldots,n\}.$$

(c) Also, $A + B$ is a lower-triangular $n \times n$-matrix. Furthermore, λA is a lower-triangular matrix whenever λ is a number.

The proof of Theorem 3.25 is analogous to that of Theorem 3.23, and the changes required are fairly straightforward (change some inequality signs). Let me pose this as an exercise:

Exercise 3.26. Prove Theorem 3.25 **(a)**. (Feel free to repeat my proof of Theorem 3.23 **(a)**, changing only what little needs to be changed. This is not plagiarism for the purpose of this exercise!)

(Similarly, you can prove Theorem 3.25 **(b)** and **(c)**, but you don't need to write it up.)

The following result is an analogue of Theorem 3.23 and Theorem 3.25 for diagonal matrices:

Theorem 3.27. Let $n \in \mathbb{N}$. Let A and B be two diagonal $n \times n$-matrices.

(a) Then, AB is a diagonal $n \times n$-matrix.

(b) The diagonal entries of AB are

$$(AB)_{i,i} = A_{i,i}B_{i,i} \qquad \text{for all } i \in \{1,2,\ldots,n\}.$$

Thus, diagonal matrices actually **are** multiplied entry by entry!

(c) Also, $A + B$ is a diagonal $n \times n$-matrix. Furthermore, λA is a diagonal matrix whenever λ is a number.

Exercise 3.28. Prove Theorem 3.27.

Lower-triangular and upper-triangular matrices are not only analogues of each other; they are also closely related:

Proposition 3.29. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then, A is upper-triangular if and only if A^T is lower-triangular.

Proof of Proposition 3.29. The definition of A^T shows that $A^T = \left(A_{j,i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Thus,

$$\left(A^T\right)_{i,j} = A_{j,i} \qquad \text{for all } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,n\}. \tag{53}$$

Now, consider the following chain of equivalent statements[38]:

$$
\begin{aligned}
&(A \text{ is upper-triangular}) \\
\iff\ &\left(A_{i,j} = 0 \text{ whenever } i > j\right) \\
&\quad (\text{because this is how "upper-triangular" is defined}) \\
\iff\ &\left(A_{j,i} = 0 \text{ whenever } j > i\right) \\
&\quad (\text{here, we have just renamed } i \text{ and } j \text{ as } j \text{ and } i) \\
\iff\ &\left(A_{j,i} = 0 \text{ whenever } i < j\right) \\
&\quad (\text{because } j > i \text{ is equivalent to } i < j) \\
\iff\ &\left(\left(A^T\right)_{i,j} = 0 \text{ whenever } i < j\right) \\
&\quad \left(\text{here, we have replaced } A_{j,i} \text{ by } \left(A^T\right)_{i,j}, \text{ because of (53)}\right) \\
\iff\ &\left(A^T \text{ is lower-triangular}\right) \\
&\quad (\text{because this is how "lower-triangular" is defined}) .
\end{aligned}
$$

Thus, Proposition 3.29 holds. □

There are a few special classes of triangular matrices worth giving names:

Definition 3.30. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix.

(a) The matrix A is said to be *upper-unitriangular* if and only if it is upper-triangular and all its diagonal entries are 1 (that is, $A_{i,i} = 1$ for all i).

(b) The matrix A is said to be *invertibly upper-triangular* if and only if it is upper-triangular and all its diagonal entries are nonzero (that is, $A_{i,i} \neq 0$ for all i).

(c) The matrix A is said to be *strictly upper-triangular* if and only if it is upper-triangular and all its diagonal entries are 0 (that is, $A_{i,i} = 0$ for all i).

Similar notions can be defined with the word "lower" instead of "upper":

(d) The matrix A is said to be *lower-unitriangular* if and only if it is lower-triangular and all its diagonal entries are 1 (that is, $A_{i,i} = 1$ for all i).

(e) The matrix A is said to be *invertibly lower-triangular* if and only if it is lower-triangular and all its diagonal entries are nonzero (that is, $A_{i,i} \neq 0$ for all i).

(f) The matrix A is said to be *strictly lower-triangular* if and only if it is lower-triangular and all its diagonal entries are 0 (that is, $A_{i,i} = 0$ for all i).

The words we have just defined are not as important as the word "upper-triangular" (you certainly don't need to learn them by heart); but these notions appear from time to time in mathematics, and it helps if you know how to recognize them.

[38] After each equivalence, we give a justification for why it is an equivalence.

Example 3.31. **(a)** A 3×3-matrix is upper-unitriangular if and only if it has the form $\begin{pmatrix} 1 & b & c \\ 0 & 1 & c' \\ 0 & 0 & 1 \end{pmatrix}$ for some b, c, c'.

(b) A 3×3-matrix is invertibly upper-triangular if and only if it has the form $\begin{pmatrix} a & b & c \\ 0 & b' & c' \\ 0 & 0 & c'' \end{pmatrix}$ for some a, b, c, b', c', c'' with $a \neq 0$, $b' \neq 0$ and $c'' \neq 0$.

(c) A 3×3-matrix is strictly upper-triangular if and only if it has the form $\begin{pmatrix} 0 & b & c \\ 0 & 0 & c' \\ 0 & 0 & 0 \end{pmatrix}$ for some b, c, c'.

Olver and Shakiban (in [OlvSha06]) use the word "special upper triangular" instead of "upper-unitriangular". But I prefer "upper-unitriangular", since the word "uni" hints directly to the definition (namely, the 1's on the diagonal), whereas the word "special" can mean pretty much anything.

The word "invertibly upper-triangular" is my invention. I have chosen it because an upper-triangular matrix is invertible if and only if it is invertibly upper-triangular. (This is not obvious. In Theorem 3.99, we will prove the "if" direction. The "only if" direction is also true.)

Here is a simple fact to connect the above definitions:

Proposition 3.32. **(a)** Each upper-unitriangular matrix is invertibly upper-triangular.

(b) Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then, A is upper-unitriangular if and only if $I_n - A$ is strictly upper-triangular.

We shall give a proof of this proposition in Section 3.5 (but mainly as an example of how to write a proof; the mathematics itself is trivial).

Strictly upper-triangular $n \times n$-matrices can also be characterized as the $n \times n$-matrices A which satisfy

$$A_{i,j} = 0 \qquad \text{whenever } i \geq j.$$

Notice the weak inequality "$i \geq j$" (as opposed to the strict inequality "$i > j$" in the definition of upper-triangular matrices).

Exercise 3.33. Let $A = \begin{pmatrix} a & b & c \\ 0 & b' & c' \\ 0 & 0 & c'' \end{pmatrix}$ be an invertibly upper-triangular 3×3-matrix. Show that A is invertible by explicitly computing the inverse of A (in terms of a, b, c, b', c', c'').

[**Hint:** In order to find a right inverse of A, it is enough to find three column vectors u, v, w (each of size 3) satisfying the equations

$$Au = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad Av = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad Aw = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

In fact, once these vectors are found, assembling them into a matrix yields a right inverse of A (why?). Find these u, v, w. Then, check that the resulting right inverse of A is also a left inverse.]

Corollary 3.34. Let $n \in \mathbb{N}$. Let A and B be two upper-unitriangular $n \times n$-matrices. Then, AB is also an upper-unitriangular $n \times n$-matrix.

Proof of Corollary 3.34. The matrices A and B are upper-triangular (since they are upper-unitriangular). Hence, Theorem 3.23 **(a)** shows that AB is an upper-triangular $n \times n$-matrix. Moreover, Theorem 3.23 **(b)** shows that

$$(AB)_{i,i} = \underbrace{A_{i,i}}_{\substack{=1 \\ \text{(since } A \text{ is} \\ \text{unitriangular)}}} \underbrace{B_{i,i}}_{\substack{=1 \\ \text{(since } B \text{ is} \\ \text{unitriangular)}}} = 1 \cdot 1 = 1$$

for all $i \in \{1, 2, \ldots, n\}$. Thus, the matrix AB is upper-unitriangular (since we already know that AB is upper-triangular). Corollary 3.34 is thus proven. □

Corollary 3.35. Let $n \in \mathbb{N}$. If $A_1, A_2, \ldots, A_k$ (for some $k \in \mathbb{N}$) are upper-unitriangular $n \times n$-matrices, then $A_1 A_2 \cdots A_k$ is also an upper-unitriangular $n \times n$-matrix.

Proof of Corollary 3.35. This can be proven in a straightforward way by induction over k, similarly to how we proved Proposition 3.12. The important differences are:

- We should now use $k = 0$ as an induction base (because we have not required k to be positive). For $k = 0$, the product $A_1 A_2 \cdots A_k$ is an empty product (i.e., a product with no factors), and thus equals I_n (because we have **defined** an empty product of $n \times n$-matrices to equal I_n). We thus need to prove that I_n is an upper-unitriangular matrix. But this is clear by inspection.
- In the induction step, instead of using Proposition 3.11, we need to use Corollary 3.34.

□

Analogues of Corollary 3.34 and Corollary 3.35 hold for invertibly upper-triangular matrices:

Corollary 3.36. Let $n \in \mathbb{N}$. Let A and B be two invertibly upper-triangular $n \times n$-matrices. Then, AB is also an invertibly upper-triangular $n \times n$-matrix.

Corollary 3.37. Let $n \in \mathbb{N}$. If $A_1, A_2, \ldots, A_k$ (for some $k \in \mathbb{N}$) are invertibly upper-triangular $n \times n$-matrices, then $A_1 A_2 \cdots A_k$ is also an invertibly upper-triangular $n \times n$-matrix.

Exercise 3.38. Prove Corollary 3.36 and Corollary 3.37.

Similarly, analogues of the above-mentioned results hold for lower-triangular matrices. For example, the following analogues of Corollary 3.35 and of Corollary 3.37 hold:

Corollary 3.39. Let $n \in \mathbb{N}$. If $A_1, A_2, \ldots, A_k$ (for some $k \in \mathbb{N}$) are lower-unitriangular $n \times n$-matrices, then $A_1 A_2 \cdots A_k$ is also a lower-unitriangular $n \times n$-matrix.

Corollary 3.40. Let $n \in \mathbb{N}$. If $A_1, A_2, \ldots, A_k$ (for some $k \in \mathbb{N}$) are invertibly lower-triangular $n \times n$-matrices, then $A_1 A_2 \cdots A_k$ is also an invertibly lower-triangular $n \times n$-matrix.

As usual, the proofs of these analogues can be obtained from the proofs of the original versions through minor (and straightforward) modifications.

The reader can easily check that the following analogue of Proposition 3.29 holds:

Proposition 3.41. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then:

(a) The matrix A is upper-unitriangular if and only if A^T is lower-unitriangular.

(b) The matrix A is invertibly upper-triangular if and only if A^T is invertibly lower-triangular.

(c) The matrix A is strictly upper-triangular if and only if A^T is strictly lower-triangular.

3.5. (*) Proof of Proposition 3.32

Here is a detailed proof of Proposition 3.32:

Proof of Proposition 3.32. **(a)** Let A be an upper-unitriangular $n \times n$-matrix. We must then prove that A is invertibly upper-triangular.

We have assumed that A is upper-unitriangular. In other words, A is upper-triangular and all its diagonal entries are 1 (because this is what "upper-unitriangular" means). Now, all diagonal entries of A are 1, and thus are nonzero (since 1 is nonzero). So we know that A is upper-triangular and all its diagonal entries are nonzero. In other words, A is invertibly upper-triangular (by the definition of "invertibly upper-triangular"). This proves Proposition 3.32 **(a)**.

(b) The statement of Proposition 3.32 **(b)** is an "if and only if" statement. Thus, it splits into the following two claims:

Claim 1: If A is upper-unitriangular, then $I_n - A$ is strictly upper-triangular.

Claim 2: If $I_n - A$ is strictly upper-triangular, then A is upper-unitriangular.

We are going to prove both of these claims. But first, let us make a simple observation: Recall that $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, $\left(I_n\right)_{i,j} = \delta_{i,j}$ for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$. In particular, for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ satisfying $i > j$, we have

$$\left(I_n\right)_{i,j} = \delta_{i,j} = 0 \tag{54}$$

(since $i \neq j$ (because $i > j$)). Also, the diagonal entries of the matrix I_n are 1; in other words, every $i \in \{1,2,\ldots,n\}$ satisfies

$$\left(I_n\right)_{i,i} = 1. \tag{55}$$

Proof of Claim 1: Assume that A is upper-unitriangular. We must show that $I_n - A$ is strictly upper-triangular.

We have assumed that A is upper-unitriangular. In other words, A is upper-triangular and all its diagonal entries are 1 (because this is what "upper-unitriangular" means). Since A is upper-triangular, we have

$$A_{i,j} = 0 \qquad \text{whenever } i > j \tag{56}$$

(by the definition of "upper-triangular"). Since all diagonal entries of A are 1, we have

$$A_{i,i} = 1 \qquad \text{for each } i \in \{1,2,\ldots,n\}. \tag{57}$$

Now, recall that matrices are subtracted entry by entry. Hence, for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ satisfying $i > j$, we have

$$\left(I_n - A\right)_{i,j} = \underbrace{\left(I_n\right)_{i,j}}_{\substack{=0 \\ \text{(by (54))}}} - \underbrace{A_{i,j}}_{\substack{=0 \\ \text{(by (56))}}} = 0 - 0 = 0.$$

In other words, $\left(I_n - A\right)_{i,j} = 0$ whenever $i > j$. This means that $I_n - A$ is upper-triangular (by the definition of "upper-triangular").

Recall again that matrices are subtracted entry by entry. Hence, every $i \in \{1,2,\ldots,n\}$ satisfies

$$\left(I_n - A\right)_{i,i} = \underbrace{\left(I_n\right)_{i,i}}_{\substack{=1 \\ \text{(by (55))}}} - \underbrace{A_{i,i}}_{\substack{=1 \\ \text{(by (57))}}} = 1 - 1 = 0.$$

In other words, all diagonal entries of the matrix $I_n - A$ are 0.

So we have shown that the matrix $I_n - A$ is upper-triangular, and that all its diagonal entries are 0. In other words, the matrix $I_n - A$ is strictly upper-triangular (by the definition of "strictly upper-triangular"). This proves Claim 1.

Proof of Claim 2: Assume that $I_n - A$ is strictly upper-triangular. We must show that A is upper-unitriangular.

We have assumed that $I_n - A$ is strictly upper-triangular. In other words, $I_n - A$ is upper-triangular and all its diagonal entries are 0 (because this is what "strictly upper-triangular" means). Since $I_n - A$ is upper-triangular, we have

$$(I_n - A)_{i,j} = 0 \qquad \text{whenever } i > j \tag{58}$$

(by the definition of "upper-triangular"). Since all diagonal entries of $I_n - A$ are 0, we have

$$(I_n - A)_{i,i} = 0 \qquad \text{for each } i \in \{1, 2, \ldots, n\}. \tag{59}$$

Now, recall that matrices are subtracted entry by entry. Hence, for all $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfying $i > j$, we have

$$(I_n - A)_{i,j} = \underbrace{(I_n)_{i,j}}_{\substack{=0 \\ \text{(by (54))}}} - A_{i,j} = 0 - A_{i,j} = -A_{i,j}$$

and therefore

$$A_{i,j} = -\underbrace{(I_n - A)_{i,j}}_{\substack{=0 \\ \text{(by (58))}}} = -0 = 0.$$

In other words, $A_{i,j} = 0$ whenever $i > j$. This means that A is upper-triangular (by the definition of "upper-triangular").

Recall again that matrices are subtracted entry by entry. Hence, every $i \in \{1, 2, \ldots, n\}$ satisfies

$$(I_n - A)_{i,i} = \underbrace{(I_n)_{i,i}}_{\substack{=1 \\ \text{(by (55))}}} - A_{i,i} = 1 - A_{i,i}$$

and therefore

$$A_{i,i} = 1 - \underbrace{(I_n - A)_{i,i}}_{\substack{=0 \\ \text{(by (59))}}} = 1 - 0 = 1.$$

In other words, all diagonal entries of the matrix A are 1.

So we have shown that the matrix A is upper-triangular, and that all its diagonal entries are 1. In other words, the matrix A is upper-unitriangular. This proves Claim 2.

Now, both Claim 1 and Claim 2 are proven, so the proof of Proposition 3.32 **(b)** is complete. □

As you might have noticed, I have written down the above proof at an unusually high level of detail (whereas most textbooks would have only sketched it, or even left it to the reader to fill in). The reason for that is that I wanted to demonstrate the structure of such proofs. An experienced writer (writing for experienced readers) would have been able to shorten the above proof considerably in the following way:

- Our proof of Proposition 3.32 **(a)** was really obvious; most of it was boilerplate (writing down the assumptions, writing down the claims, etc.).
- The way we proved Proposition 3.32 **(b)** is a typical way how "if and only if" statements are proven.[39] What we called Claim 1 in this proof would normally be called the "$\Longrightarrow$ direction"[40] of Proposition 3.32 **(b)** (because, rewritten in logical symbols, it says that "(A is upper-unitriangular) $\Longrightarrow$ ($I_n - A$ is strictly upper-triangular)"), while our Claim 2 would be called the "$\Longleftarrow$ direction"[41] of Proposition 3.32 **(b)** (for similar reasons). In general, if you have an assertion of the form "X holds if and only if Y holds", then the "$\Longrightarrow$ direction" of this assertion says "if X holds, then Y holds", whereas the "$\Longleftarrow$ direction" of this assertion says "if Y holds, then X holds". In order to prove the assertion, it suffices to prove both its $\Longrightarrow$ direction and its $\Longleftarrow$ direction; these two directions can often be proven separately. It is customary to mark the proof of the $\Longrightarrow$ direction by a single "$\Longrightarrow$:" at the beginning of this proof (instead of writing "Proof of Claim 1:" as we did), and to mark the proof of the $\Longleftarrow$ direction by a single "$\Longleftarrow$:" at the beginning of this proof (instead of writing "Proof of Claim 2:" as we did). Furthermore, explicitly stating Claim 1 and Claim 2 (like I did above) is not necessary: They are just the $\Longrightarrow$ direction and the $\Longleftarrow$ direction of Proposition 3.32 **(b)**, respectively; this is enough to fully characterize them.
- Our proof of the $\Longrightarrow$ direction (i.e., of Claim 1) was straightforward: it was following the most obvious route from the givens (i.e., from the assumption that A is upper-unitriangular) to the goal (i.e., to the claim that $I_n - A$ is strictly upper-triangular). In fact, once you have unraveled the definitions of "upper-unitriangular" and of "triangular", the assumption translates into "$A_{i,j} = 0$ whenever $i > j$, and $A_{i,i} = 1$ for all i". Similarly, once you have unraveled the definitions of "strictly upper-triangular" and "triangular", the claim translates into "$(I_n - A)_{i,j} = 0$ whenever $i > j$, and $(I_n - A)_{i,i} = 0$ for all i". Thus, in order to get from the assumption to the goal, you need to find a way to process knowledge about entries of A into knowledge about entries of $I_n - A$. But there is one obvious way to do that: Observe that $(I_n - A)_{i,j} = (I_n)_{i,j} - A_{i,j}$, and recall the formula for $(I_n)_{i,j}$ (which is clear from the definition of I_n). The rest is simple arithmetic.

[39] But not the only way: our proof of Proposition 3.29 was organized differently.
[40] also known as the "only if direction" or the "$\Longrightarrow$ part"
[41] also known as the "if direction" or the "$\Longleftarrow$ part"

- Our proof of the $\Longleftarrow$ direction (i.e., of Claim 2) was essentially the proof of the $\Longrightarrow$ direction (i.e., of Claim 1) "read backwards" (the assumption and the claim have switched places, so we are taking the same argument but in reverse). To read an argument backwards means, whenever necessary, to switch reasons with consequences (for example, instead of deriving $(I_n - A)_{i,j} = 0$ from $A_{i,j} = 0$, we now derive $A_{i,j} = 0$ from $(I_n - A)_{i,j} = 0$). This is not possible (after all, not every valid statement remains valid if we switch the assumption and the claim!), but when it is, it is not a difficult task, so it can safely be left to the reader. (And it does work in our case.)

Altogether, we can thus shrink down our above proof of Proposition 3.32 to the following short form:

Proof of Proposition 3.32 (sketched). **(a)** This follows from the definitions, since 1 is nonzero.

(b) $\Longrightarrow$: Each entry of the matrix $I_n - A$ equals the corresponding entry of I_n minus the corresponding entry of A. Thus, any statement about entries of $I_n - A$ can be reduced to a statement about corresponding entries of A. Using this observation, the $\Longrightarrow$ direction of Proposition 3.32 **(b)** becomes straightforward.

$\Longleftarrow$: To obtain a proof of the $\Longleftarrow$ direction, read the proof of the $\Longrightarrow$ direction backwards. $\square$

(Or we could have left the whole proof to the reader, seeing that pretty much all of it was straightforward walking from the assumptions to the goals.)

3.6. The standard matrix units $E_{u,v}$

We shall now define a certain family of matrices which are (in a sense) the building blocks of all matrices:

Definition 3.42. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, m\}$. Then, $E_{u,v,n,m}$ shall denote the $n \times m$-matrix whose (u, v)-th entry is 1 and whose all other entries are 0. We shall abbreviate $E_{u,v,n,m}$ as $E_{u,v}$ when the values of n and m are clear from the context (for example, when we say "the 2×5-matrix $E_{1,4}$", it is clear that $n = 2$ and $m = 5$).

The matrices $E_{u,v}$ (for varying u and v) are called the *standard matrix units*.

Example 3.43. The 2×3-matrix $E_{1,3}$ (also known as $E_{1,3,2,3}$) is $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. As per its definition, its $(1, 3)$-th entry is 1 and its all other entries are 0.

The 3×2-matrix $E_{2,2}$ (also known as $E_{2,2,3,2}$) is $\begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

What happens when a matrix is multiplied by $E_{u,v}$? There are two cases: either the $E_{u,v}$ is on the left (so we are talking of a product $E_{u,v}C$) or the $E_{u,v}$ is on the right (so we are talking of a product $CE_{u,v}$). Let us see what each of these looks like:

Proposition 3.44. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let $u \in \{1,2,\ldots,n\}$ and $v \in \{1,2,\ldots,m\}$. Let C be an $m \times p$-matrix. Then, $E_{u,v}C$ is the $n \times p$-matrix whose u-th row is the v-th row of C, and whose all other rows are filled with zeroes. (Here, again, $E_{u,v}$ means $E_{u,v,n,m}$.)

Example 3.45. Let $n = 2$, $m = 3$ and $p = 3$. Let $C = \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix}$ be a 3×3-matrix. Proposition 3.44 (applied to $u = 1$ and $v = 2$) claims that $E_{1,2}C$ is the 2×3-matrix whose 1-st row is the 2-nd row of C, and whose all other rows are filled with zeroes. In other words, it claims that

$$E_{1,2}C = \begin{pmatrix} a' & b' & c' \\ 0 & 0 & 0 \end{pmatrix}.$$

We can verify this by actually doing the multiplication:

$$\begin{aligned} E_{1,2}C &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{pmatrix} \\ &= \begin{pmatrix} 0a+1a'+0a'' & 0b+1b'+0b'' & 0c+1c'+0c'' \\ 0a+0a'+0a'' & 0b+0b'+0b'' & 0c+0c'+0c'' \end{pmatrix} \\ &= \begin{pmatrix} a' & b' & c' \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

The matrix unit $E_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ acts as a sort of mask which, when multiplied by C, moves the v-th row into the u-th row of the product, while destroying all other rows.

We shall give a formal proof of Proposition 3.44 in Section 3.7; but the example above should have given you a good intuition for it.

So much for products of the form $E_{u,v}C$. What about $CE_{u,v}$?

Proposition 3.46. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let $u \in \{1,2,\ldots,n\}$ and $v \in \{1,2,\ldots,m\}$. Let C be an $p \times n$-matrix. Then, $CE_{u,v}$ is the $p \times m$-matrix whose v-th column is the u-th column of C, and whose all other columns are filled with zeroes. (Here, again, $E_{u,v}$ means $E_{u,v,n,m}$.)

Notice that the numbers u and v play different parts in Proposition 3.46 as in Proposition 3.44.

Example 3.47. Let us demonstrate Proposition 3.46 on a more spartanic example: Let $n = 2$, $m = 2$ and $p = 1$. Let $C = \begin{pmatrix} a & b \end{pmatrix}$ be a 1×2-matrix. Proposition 3.46 (applied to $u = 1$ and $v = 2$) claims that $CE_{1,2}$ is the 1×2-matrix whose 2-nd column is the 1-st column of C, and whose all other columns are filled with zeroes. In other words, it claims that

$$CE_{1,2} = \begin{pmatrix} 0 & a \end{pmatrix}.$$

Again, we can verify this by actually doing the multiplication:

$$CE_{1,2} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a \cdot 0 + b \cdot 0 & a \cdot 1 + b \cdot 0 \end{pmatrix} = \begin{pmatrix} 0 & a \end{pmatrix}.$$

3.7. (*) A bit more on the standard matrix units

Let us get some practice by rewriting the definition of the matrices $E_{u,v}$:

Proposition 3.48. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$ and $v \in \{1, 2, \ldots, m\}$. Let $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$. Then,

$$(E_{u,v})_{i,j} = \delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v} \tag{60}$$

(where $E_{u,v}$ is short for $E_{u,v,n,m}$). (Recall that the meaning of the symbols $\delta_{(i,j),(u,v)}$, $\delta_{i,u}$ and $\delta_{j,v}$ is defined as in (20).)

Proof of Proposition 3.48. We have defined $E_{u,v}$ as the $n \times m$-matrix whose (u, v)-th entry is 1 and whose all other entries are 0. Hence, the (i, j)-th entry of the matrix $E_{u,v}$ is 1 if $(i, j) = (u, v)$ and 0 otherwise. In formulas, this says that

$$(E_{u,v})_{i,j} = \begin{cases} 1, & \text{if } (i,j) = (u,v); \\ 0, & \text{if } (i,j) \neq (u,v) \end{cases}. \tag{61}$$

On the other hand, the definition of $\delta_{(i,j),(u,v)}$ yields

$$\delta_{(i,j),(u,v)} = \begin{cases} 1, & \text{if } (i,j) = (u,v); \\ 0, & \text{if } (i,j) \neq (u,v) \end{cases}. \tag{62}$$

Comparing this with (61), we immediately obtain $(E_{u,v})_{i,j} = \delta_{(i,j),(u,v)}$.

Let us now show that

$$\delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}. \tag{63}$$

[*Proof of (63):* We are in one of the following three cases:

Case 1: We have $i \neq u$.

Case 2: We have $j \neq v$.

Case 3: We have neither $i \neq u$ nor $j \neq v$.

(In fact, it is possible that we are in Case 1 and Case 2 simultaneously. But this does not invalidate our proof; it is perfectly fine if the cases "overlap", as long as every possible situation is covered by at least one case.)

We shall prove (63) in each of the three cases:

1. Let us first consider Case 1. In this case, we have $i \neq u$. Hence, $(i,j) \neq (u,v)$. Thus, $\delta_{(i,j),(u,v)} = 0$. Comparing this with $\underbrace{\delta_{i,u}}_{\substack{=0 \\ (\text{since } i \neq u)}} \delta_{j,v} = 0\delta_{j,v} = 0$, we find $\delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}$. Hence, (63) is proven in Case 1.

2. Let us next consider Case 2. In this case, we have $j \neq v$. Hence, $(i,j) \neq (u,v)$. Thus, $\delta_{(i,j),(u,v)} = 0$. Comparing this with $\delta_{i,u} \underbrace{\delta_{j,v}}_{\substack{=0 \\ (\text{since } j \neq v)}} = \delta_{i,u}0 = 0$, we find $\delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}$. Hence, (63) is proven in Case 2.

3. Let us finally consider Case 3. In this case, we have neither $i \neq u$ and $j \neq v$. Hence, $i = u$ (since not $i \neq u$) and $j = v$ (since not $j \neq v$). As a consequence, $(i,j) = (u,v)$, so that $\delta_{(i,j),(u,v)} = 1$. Comparing this with $\underbrace{\delta_{i,u}}_{\substack{=1 \\ (\text{since } i = u)}} \underbrace{\delta_{j,v}}_{\substack{=1 \\ (\text{since } j = v)}} = 1 \cdot 1 = 1$, we find $\delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}$. Hence, (63) is proven in Case 3.

We have now proven (63) in all three Cases; this finishes our proof of (63).]

Now, we have shown both $(E_{u,v})_{i,j} = \delta_{(i,j),(u,v)}$ and $\delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}$. Combining these two equalities gives us (60), and so Proposition 3.48 is proven. □

We can now easily prove Proposition 3.44:

Proof of Proposition 3.44. We need to prove the following two claims:

Claim 1: The u-th row of the $n \times p$-matrix $E_{u,v}C$ is the v-th row of C.

Claim 2: For each $i \in \{1,2,\ldots,n\}$ satisfying $i \neq u$, the i-th row of the $n \times p$-matrix $E_{u,v}C$ is filled with zeroes.

Before we do so, let us derive a few formulas. First of all, we have

$$(E_{u,v})_{i,j} = \delta_{i,u}\delta_{j,v} \tag{64}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$ (according to Proposition 3.48). Furthermore, for any $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$, we have

$$(E_{u,v}C)_{i,j} = \sum_{k=1}^{m} \underbrace{(E_{u,v})_{i,k}}_{\substack{=\delta_{i,u}\delta_{k,v} \\ \text{(by (64), applied to} \\ k \text{ instead of } j)}} C_{k,j} \qquad \left(\begin{array}{c} \text{by Proposition 2.28,} \\ \text{applied to } A = E_{u,v} \text{ and } B = C \end{array} \right)$$

$$= \sum_{k=1}^{m} \delta_{i,u}\delta_{k,v}C_{k,j} = \sum_{k=1}^{m} \delta_{i,u}C_{k,j}\delta_{k,v} = \delta_{i,u}C_{v,j} \tag{65}$$

(by Proposition 2.40, applied to 1, m, v and $\delta_{i,u}C_{k,j}$ instead of p, q, r and a_k). Now, we can prove both claims easily:

Proof of Claim 1: For every $j \in \{1,2,\ldots,p\}$, we have

$$(E_{u,v}C)_{u,j} = \underbrace{\delta_{u,u}}_{\substack{=1 \\ \text{(since } u=u\text{)}}} C_{v,j} \qquad \text{(by (65), applied to } i = u\text{)}$$
$$= C_{v,j}.$$

In other words, for every $j \in \{1,2,\ldots,p\}$, the (u,j)-th entry of the matrix $E_{u,v}C$ equals the (v,j)-th entry of the matrix C. In other words, each entry of the u-th row of $E_{u,v}C$ equals the corresponding entry of the v-th row of C. In other words, the u-th row of the $n \times p$-matrix $E_{u,v}C$ is the v-th row of C. This proves Claim 1.

Proof of Claim 2: Let $i \in \{1,2,\ldots,n\}$ be such that $i \neq u$. We must prove that the i-th row of the $n \times p$-matrix $E_{u,v}C$ is filled with zeroes.

For every $j \in \{1,2,\ldots,p\}$, we have

$$(E_{u,v}C)_{i,j} = \underbrace{\delta_{i,u}}_{\substack{=0 \\ \text{(since } i\neq u\text{)}}} C_{v,j} \qquad \text{(by (65))}$$
$$= 0C_{v,j} = 0.$$

In other words, for every $j \in \{1,2,\ldots,p\}$, the (i,j)-th entry of the matrix $E_{u,v}C$ is 0. In other words, each entry of the i-th row of $E_{u,v}C$ is 0. In other words, the i-th row of the $n \times p$-matrix $E_{u,v}C$ is filled with zeroes. This proves Claim 2.

Now, both Claim 1 and Claim 2 are proven, and so we are finished proving Proposition 3.44. □

The proof of Proposition 3.46 is similar, and is left to the reader.

Here is another simple property of matrix units:

Proposition 3.49. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $u \in \{1,2,\ldots,n\}$ and $v \in \{1,2,\ldots,m\}$. Then, $(E_{u,v})^T = E_{v,u}$. (More precisely: $(E_{u,v,n,m})^T = E_{v,u,m,n}$.)

The proof of Proposition 3.49 is easy enough that it could be left as an exercise at this point, but let me nevertheless give it for the sake of completeness:

Proof of Proposition 3.49. We first notice that both $\left(E_{u,v,n,m}\right)^T$ and $E_{v,u,m,n}$ are $m \times n$-matrices.

For every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, we have

$$\left(E_{u,v,n,m}\right)_{i,j} = \delta_{(i,j),(u,v)} = \delta_{i,u}\delta_{j,v}. \tag{66}$$

(Indeed, this is just a more precise way to state (60), because the "$E_{u,v}$" in (60) is a shorthand for "$E_{u,v,n,m}$".) Thus, in particular, every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$ satisfy

$$\left(E_{u,v,n,m}\right)_{i,j} = \delta_{i,u}\delta_{j,v}. \tag{67}$$

(This is an immediate consequence of (66).)

The definition of the transpose $\left(E_{u,v,n,m}\right)^T$ now yields $\left(E_{u,v,n,m}\right)^T = \left(\left(E_{u,v,n,m}\right)_{j,i}\right)_{1\leq i\leq m,\ 1\leq j\leq n}$. Hence, for every $i \in \{1,2,\ldots,m\}$ and $j \in \{1,2,\ldots,n\}$, we have

$$\left(\left(E_{u,v,n,m}\right)^T\right)_{i,j} = \left(E_{u,v,n,m}\right)_{j,i} = \delta_{j,u}\delta_{i,v} \tag{68}$$

(by (67), applied to j and i instead of i and j).

On the other hand, the same reasoning that we used to obtain (67) can be applied to m, n, v and u instead of n, m, u and v. As a result of this reasoning, we then find that every $i \in \{1,2,\ldots,m\}$ and $j \in \{1,2,\ldots,n\}$ satisfy

$$\left(E_{v,u,m,n}\right)_{i,j} = \delta_{i,v}\delta_{j,u}. \tag{69}$$

Hence, for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, we have

$$\begin{aligned}\left(\left(E_{u,v,n,m}\right)^T\right)_{i,j} &= \delta_{j,u}\delta_{i,v} \qquad \text{(by (68))} \\ &= \delta_{i,v}\delta_{j,u} = \left(E_{v,u,m,n}\right)_{i,j} \qquad \text{(by (69))}.\end{aligned}$$

In other words, for every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, the (i,j)-th entry of the matrix $\left(E_{u,v,n,m}\right)^T$ equals the (i,j)-th entry of the matrix $E_{v,u,m,n}$. In other words, each entry of the matrix $\left(E_{u,v,n,m}\right)^T$ equals the corresponding entry of the matrix $E_{v,u,m,n}$. Hence, $\left(E_{u,v,n,m}\right)^T = E_{v,u,m,n}$. Using our shorthand notations, this rewrites as $\left(E_{u,v}\right)^T = E_{v,u}$. Thus, Proposition 3.49 is proven. □

As I have said, the standard matrix units are building blocks for matrices: every matrix can be obtained from them by scaling and adding. More precisely:

Proposition 3.50. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Then,

$$A = \sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}E_{u,v}.$$

Here, we are using the $\sum$ symbol introduced in Section 2.14. (Of course, $E_{u,v}$ means $E_{u,v,n,m}$ here.)

Example 3.51. In the case when $n = 2$ and $m = 2$, Proposition 3.50 says that

$$A = A_{1,1}E_{1,1} + A_{1,2}E_{1,2} + A_{2,1}E_{2,1} + A_{2,2}E_{2,2}.$$

It is easy to check this directly:

$$\begin{aligned}
&A_{1,1}E_{1,1} + A_{1,2}E_{1,2} + A_{2,1}E_{2,1} + A_{2,2}E_{2,2} \\
&= A_{1,1}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + A_{1,2}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + A_{2,1}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + A_{2,2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} A_{1,1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & A_{1,2} \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ A_{2,1} & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & A_{2,2} \end{pmatrix} \\
&= \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = A.
\end{aligned}$$

This should make the truth of Proposition 3.50 obvious (even in the general case). The proof below is just for the fans of formal reasoning.

Proof of Proposition 3.50. For every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, we have

$$\begin{aligned}
&\left(\sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}E_{u,v}\right)_{i,j} \\
&= \sum_{u=1}^{n} \underbrace{\left(\sum_{v=1}^{m} A_{u,v}E_{u,v}\right)_{i,j}}_{\substack{=\sum_{v=1}^{m}(A_{u,v}E_{u,v})_{i,j} \\ \text{(by an application of Proposition 2.45)}}} \qquad \text{(by an application of Proposition 2.45)} \\
&= \sum_{u=1}^{n}\sum_{v=1}^{m} \underbrace{(A_{u,v}E_{u,v})_{i,j}}_{\substack{=A_{u,v}(E_{u,v})_{i,j} \\ \text{(since matrices are scaled entry by entry)}}} \\
&= \sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}\underbrace{(E_{u,v})_{i,j}}_{\substack{=\delta_{i,u}\delta_{j,v} \\ \text{(by (60))}}} = \sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}\underbrace{\delta_{i,u}}_{=\delta_{u,i}}\underbrace{\delta_{j,v}}_{=\delta_{v,j}} \\
&= \sum_{u=1}^{n} \underbrace{\sum_{v=1}^{m} A_{u,v}\delta_{u,i}\delta_{v,j}}_{\substack{=\sum_{k=1}^{m} A_{u,k}\delta_{u,i}\delta_{k,j} \\ =A_{u,j}\delta_{u,i} \\ \text{(by Proposition 2.40, applied to } p=1,\ q=m,\ r=j \text{ and } a_k=A_{u,k}\delta_{u,i}\text{)}}} = \sum_{u=1}^{n} A_{u,j}\delta_{u,i} = \sum_{k=1}^{n} A_{k,j}\delta_{k,i} \\
&= A_{i,j} \qquad \left(\text{by Proposition 2.40, applied to } p=1,\ q=n,\ r=i \text{ and } a_k=A_{k,j}\right).
\end{aligned}$$

In other words, each entry of the matrix $\sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}E_{u,v}$ equals the corresponding entry of the matrix A. Thus, $\sum_{u=1}^{n}\sum_{v=1}^{m} A_{u,v}E_{u,v} = A$. This proves Proposition 3.50. □

Continuing the flow of boring and straightforward little propositions, let us see how standard matrix units multiply:

Proposition 3.52. Let $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $p \in \mathbb{N}$. Let $u \in \{1,2,\ldots,n\}$, $v \in \{1,2,\ldots,m\}$, $x \in \{1,2,\ldots,m\}$ and $y \in \{1,2,\ldots,p\}$. Then,

$$E_{u,v,n,m}E_{x,y,m,p} = \delta_{v,x}E_{u,y,n,p}.$$

(We shall write this equality as $E_{u,v}E_{x,y} = \delta_{v,x}E_{u,y}$, since we hope that the sizes of the matrices will be clear from the context.)

It might be a good exercise to devise at least two examples for this proposition (one with $v = x$ and one with $v \neq x$), and to prove it. Nevertheless, let me give a proof, because noone ever seems to do so in writing:

Proof of Proposition 3.52. There are two ways to prove Proposition 3.52: One is to apply Proposition 3.44 to $C = E_{x,y}$. Another is to obstinately applying the definitions. We opt for the second way, because the first is too simple.

From (60), we obtain

$$\left(E_{u,v,n,m}\right)_{i,j} = \delta_{i,u}\delta_{j,v} \tag{70}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$. (If this doesn't look like (60) to you, do remember that $E_{u,v,n,m}$ was abbreviated as $E_{u,v}$ in (60).)

The same reasoning (applied to m, p, x and y instead of n, m, u and v) shows that

$$\left(E_{x,y,m,p}\right)_{i,j} = \delta_{i,x}\delta_{j,y} \tag{71}$$

for all $i \in \{1,2,\ldots,m\}$ and $j \in \{1,2,\ldots,p\}$.

The same reasoning that gave us (70) can also be applied to n, p, u and y instead of n, m, u and v. As a result, we find

$$\left(E_{u,y,n,p}\right)_{i,j} = \delta_{i,u}\delta_{j,y} \tag{72}$$

for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$.

Now, let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$. Then, Proposition 2.28 (applied to $A = E_{u,v,n,m}$ and $B = E_{x,y,m,p}$) shows that

$$\begin{aligned}
\left(E_{u,v,n,m}E_{x,y,m,p}\right)_{i,j} &= \sum_{k=1}^{m} \underbrace{\left(E_{u,v,n,m}\right)_{i,k}}_{\substack{=\delta_{i,u}\delta_{k,v} \\ \text{(by (70), applied to } k \text{ instead of } j\text{)}}} \underbrace{\left(E_{x,y,m,p}\right)_{k,j}}_{\substack{=\delta_{k,x}\delta_{j,y} \\ \text{(by (71), applied to } k \text{ instead of } i\text{)}}} \\
&= \sum_{k=1}^{m} \delta_{i,u}\delta_{k,v}\delta_{k,x}\delta_{j,y} = \sum_{k=1}^{m} \delta_{i,u}\delta_{k,x}\delta_{j,y}\delta_{k,v} \\
&= \delta_{i,u}\delta_{v,x}\delta_{j,y}
\end{aligned}$$

(by Proposition 2.40, applied to 1, m, v and $\delta_{i,u}\delta_{k,x}\delta_{j,y}$ instead of p, q, r and a_k). Comparing this with

$$\begin{aligned}
\left(\delta_{v,x}E_{u,y,n,p}\right)_{i,j} &= \delta_{v,x} \underbrace{\left(E_{u,y,n,p}\right)_{i,j}}_{\substack{=\delta_{i,u}\delta_{j,y} \\ \text{(by (72))}}} \\
&\qquad \left(\text{since matrices are scaled entry by entry}\right) \\
&= \delta_{v,x}\delta_{i,u}\delta_{j,y} = \delta_{i,u}\delta_{v,x}\delta_{j,y},
\end{aligned}$$

we find $\left(E_{u,v,n,m}E_{x,y,m,p}\right)_{i,j} = \left(\delta_{v,x}E_{u,y,n,p}\right)_{i,j}$.

Now, forget that we fixed i and j. We thus have shown that $\left(E_{u,v,n,m}E_{x,y,m,p}\right)_{i,j} = \left(\delta_{v,x}E_{u,y,n,p}\right)_{i,j}$ for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,p\}$. In other words, each entry of the matrix $E_{u,v,n,m}E_{x,y,m,p}$ equals the corresponding entry of the matrix $\delta_{v,x}E_{u,y,n,p}$. Thus, $E_{u,v,n,m}E_{x,y,m,p} = \delta_{v,x}E_{u,y,n,p}$. This proves Proposition 3.52. □

3.8. The λ-addition matrices $A^{\lambda}_{u,v}$

Now, we come to another important class of matrices (that can also be seen as building blocks of a kind): the *λ-addition matrices*. Those are square matrices, and are defined as follows:

Definition 3.53. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1,2,\ldots,n\}$. Let λ be a number. Then, $A^{\lambda}_{u,v}$ shall denote the $n \times n$-matrix $I_n + \lambda E_{u,v}$ (where $E_{u,v}$ means the $n \times n$-matrix $E_{u,v}$, that is, $E_{u,v,n,n}$).

A few remarks about the notation:

(a) The superscript λ in the notation "$A^{\lambda}_{u,v}$" is not an exponent; i.e., the matrix $A^{\lambda}_{u,v}$ is not the λ-th power of some matrix $A_{u,v}$. Instead, it is just an argument that we have chosen to write as a superscript instead of a subscript. So the role of the λ in "$A^{\lambda}_{u,v}$" is completely different from the role of the -1 in "A_1^{-1}" in Proposition 3.12.

(b) To be really precise, we ought to denote $A^{\lambda}_{u,v}$ by $A^{\lambda}_{u,v,n}$, because it depends on n. (This is similar to how we ought to denote $E_{u,v}$ by $E_{u,v,n,n}$.) But the n will be really clear from the context almost every time we deal with these matrices, so we shall keep it out of our notation.

(c) The notation $A^{\lambda}_{u,v}$ is not standard in the literature, but I will use this notation in the following.

Example 3.54. Let $n = 4$. Then,

$$A^{\lambda}_{1,3} = I_n + \lambda E_{1,3} = \begin{pmatrix} 1 & 0 & \lambda & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{and}$$

$$A^{\lambda}_{4,2} = I_n + \lambda E_{4,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \lambda & 0 & 1 \end{pmatrix}.$$

The pattern that you see on these examples is true in general:

Proposition 3.55. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1,2,\ldots,n\}$. Let λ be a number. Then, the matrix $A^{\lambda}_{u,v}$ has the following entries:

- All its diagonal entries are 1.
- Its (u, v)-th entry is λ.
- All its remaining entries are 0.

Proof of Proposition 3.55. Recall that $E_{u,v}$ is the $n \times n$-matrix whose (u, v)-th entry is 1 and whose all other entries are 0 (indeed, this is how $E_{u,v}$ was defined). Since matrices are scaled entry by entry, we can therefore conclude how $\lambda E_{u,v}$ looks like: Namely, $\lambda E_{u,v}$ is the $n \times n$-matrix whose (u, v)-th entry is $\lambda \cdot 1 = \lambda$ and whose all other entries are $\lambda \cdot 0 = 0$. Thus, we know the following:

- The matrix I_n is the $n \times n$-matrix whose diagonal entries are 1, and whose all other entries are 0.
- The matrix $\lambda E_{u,v}$ is the $n \times n$-matrix whose (u, v)-th entry is λ, and whose all other entries are 0.

Since matrices are added entry by entry, we can thus infer how $I_n + \lambda E_{u,v}$ looks like: Namely, the matrix $I_n + \lambda E_{u,v}$ is the $n \times n$-matrix whose diagonal entries[42] are $1 + 0 = 1$, whose (u, v)-th entry is $0 + \lambda = \lambda$, and whose all other entries are $0 + 0 = 0$. Since $I_n + \lambda E_{u,v} = A^{\lambda}_{u,v}$, this rewrites as follows: The matrix $A^{\lambda}_{u,v}$ is the $n \times n$-matrix whose diagonal entries are 1, whose (u, v)-th entry is λ, and whose all other entries are 0. This proves Proposition 3.55. □

We can next see what happens to a matrix when it is multiplied by $A^{\lambda}_{u,v}$:

Proposition 3.56. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let λ be a number. Let C be an $n \times m$-matrix. Then, $A^{\lambda}_{u,v} C$ is the $n \times m$-matrix obtained from C by adding $\lambda \operatorname{row}_v C$ to the u-th row.

(Recall that $\operatorname{row}_k C$ denotes the k-th row of C for each k. Recall also that the rows of C are row vectors, and thus are added and scaled entry by entry. Hence, adding $\lambda \operatorname{row}_v C$ to the u-th row means adding λ times each entry of the v-th row of C to the corresponding entry of the u-th row.)

Example 3.57. Let $n = 3$ and $m = 2$. Let C be the 3×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix}$. Let λ be a number. Then, Proposition 3.56 (applied to $u = 2$ and $v = 1$) claims that $A^{\lambda}_{2,1} C$ is the 3×2-matrix obtained from C by adding $\lambda \operatorname{row}_1 C$ to the 2-nd row. A computation confirms this claim:

$$A^{\lambda}_{2,1} C = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix} = \begin{pmatrix} a & b \\ a' + \lambda a & b' + \lambda b \\ a'' & b'' \end{pmatrix}.$$

[42] Here, we are using the fact that the (u, v)-th entry is not a diagonal entry; this is because u and v are distinct! If u and v were equal, then the (u, v)-th entry of $I_n + \lambda E_{u,v}$ would be $1 + \lambda$ instead.

Proposition 3.58. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let λ be a number. Let C be an $m \times n$-matrix. Then, $CA^{\lambda}_{u,v}$ is the $m \times n$-matrix obtained from C by adding $\lambda \operatorname{col}_u C$ to the v-th column.

Note how Proposition 3.58 differs from Proposition 3.56: not only have rows been replaced by columns, but also have u and v switched roles.

You might find Example 3.57 sufficient to convince you of the truth of Proposition 3.56. If not, a proof will be given in Section 3.9 below.

We shall refer to the matrix $A^{\lambda}_{u,v}$ defined in Definition 3.53 as a "λ-addition matrix"; it is one of three kinds of matrices that are called elementary matrices. We shall learn about the other two kinds below.

Here are a few more properties of λ-addition matrices:

Proposition 3.59. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let λ be a number. Then, $\left(A^{\lambda}_{u,v}\right)^T = A^{\lambda}_{v,u}$.

Proposition 3.60. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$.
(a) We have $A^{0}_{u,v} = I_n$.
(b) If λ and μ are two numbers, then $A^{\lambda}_{u,v} A^{\mu}_{u,v} = A^{\lambda+\mu}_{u,v}$.
(c) Let λ be a number. Then, the matrix $A^{\lambda}_{u,v}$ is invertible, and its inverse is $\left(A^{\lambda}_{u,v}\right)^{-1} = A^{-\lambda}_{u,v}$.

A proof of this proposition will also be given in Section 3.9.

3.9. (*) Some proofs about the λ-addition matrices

Proof of Proposition 3.56. Clearly, $A^{\lambda}_{u,v}C$ is an $n \times m$-matrix.

Proposition 3.44 (applied to n and m instead of m and p) shows that $E_{u,v}C$ is the $n \times m$-matrix whose u-th row is the v-th row of C, and whose all other rows are filled with zeroes. Thus,

$$\operatorname{row}_u \left(E_{u,v}C\right) = \operatorname{row}_v C \tag{73}$$

(since the u-the row of $E_{u,v}C$ is the v-th row of C) and

$$\operatorname{row}_i \left(E_{u,v}C\right) = 0_{1\times m} \qquad \text{for every } i \in \{1, 2, \ldots, n\} \text{ satisfying } i \neq u \tag{74}$$

(since all other rows of $E_{u,v}C$ are filled with zeroes).

But recall that matrices are added entry by entry. Thus, matrices are also added by row by row – i.e., if U and V are two $n \times m$-matrices, then any row of $U + V$ is the sum of the corresponding rows of U and of V. In other words, if U and V are two $n \times m$-matrices, then

$$\operatorname{row}_i \left(U + V\right) = \operatorname{row}_i U + \operatorname{row}_i V \qquad \text{for every } i \in \{1, 2, \ldots, n\}. \tag{75}$$

Also, if U is an $n \times m$-matrix, then

$$\operatorname{row}_i (\lambda U) = \lambda \operatorname{row}_i U \qquad \text{for every } i \in \{1, 2, \ldots, n\} \tag{76}$$

(since matrices are scaled entry by entry).

We have

$$\underbrace{A^{\lambda}_{u,v}}_{=I_n+\lambda E_{u,v}} C = (I_n + \lambda E_{u,v}) C = \underbrace{I_n C}_{=C} + \lambda E_{u,v} C = C + \lambda E_{u,v} C.$$

Hence, for each $i \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned}
\operatorname{row}_i \left(\underbrace{A^{\lambda}_{u,v} C}_{=C+\lambda E_{u,v} C} \right) &= \operatorname{row}_i (C + \lambda E_{u,v} C) = \operatorname{row}_i C + \underbrace{\operatorname{row}_i (\lambda E_{u,v} C)}_{\substack{=\lambda \operatorname{row}_i (E_{u,v} C) \\ \text{(by (76))}}} \\
&\qquad \text{(by (75), applied to } U = C \text{ and } V = \lambda E_{u,v} C) \\
&= \operatorname{row}_i C + \lambda \operatorname{row}_i (E_{u,v} C) .
\end{aligned} \tag{77}$$

Now, we must prove that $A^{\lambda}_{u,v} C$ is the $n \times m$-matrix obtained from C by adding $\lambda \operatorname{row}_v C$ to the u-th row. In other words, we must prove the following two claims:

> *Claim 1:* The u-th row of the $n \times m$-matrix $A^{\lambda}_{u,v} C$ is the sum of $\lambda \operatorname{row}_v C$ with the u-th row of C.
>
> *Claim 2:* For each $i \in \{1, 2, \ldots, n\}$ satisfying $i \neq u$, the i-th row of the $n \times m$-matrix $A^{\lambda}_{u,v} C$ equals the i-th row of C.

Proof of Claim 1: The u-th row of the $n \times m$-matrix $A^{\lambda}_{u,v} C$ is

$$\begin{aligned}
\operatorname{row}_u \left(A^{\lambda}_{u,v} C \right) &= \operatorname{row}_u C + \lambda \underbrace{\operatorname{row}_u (E_{u,v} C)}_{\substack{=\operatorname{row}_v C \\ \text{(by (73))}}} \qquad \text{(by (77), applied to } i = u) \\
&= \operatorname{row}_u C + \lambda \operatorname{row}_v C.
\end{aligned}$$

In other words, the u-th row of the $n \times m$-matrix $A^{\lambda}_{u,v} C$ is the sum of $\lambda \operatorname{row}_v C$ with the u-th row of C. This proves Claim 1.

Proof of Claim 2: Let $i \in \{1, 2, \ldots, n\}$ be such that $i \neq u$. Then, the i-th row of the $n \times m$-matrix $A^{\lambda}_{u,v} C$ is

$$\begin{aligned}
\operatorname{row}_i \left(A^{\lambda}_{u,v} C \right) &= \operatorname{row}_i C + \lambda \underbrace{\operatorname{row}_i (E_{u,v} C)}_{\substack{=0_{1 \times m} \\ \text{(by (74))}}} \qquad \text{(by (77))} \\
&= \operatorname{row}_i C + \underbrace{\lambda 0_{1 \times m}}_{=0_{1 \times m}} = \operatorname{row}_i C + 0_{1 \times m} = \operatorname{row}_i C.
\end{aligned}$$

In other words, the i-th row of the $n \times m$-matrix $A_{u,v}^{\lambda} C$ equals the i-th row of C. This proves Claim 2.

Now, we have proven both Claim 1 and Claim 2; this completes the proof of Proposition 3.56. □

The proof of Proposition 3.58 is analogous.

Proof of Proposition 3.59. Proposition 3.49 (applied to $m = n$) yields $\left(E_{u,v}\right)^T = E_{v,u}$.

We have $A_{u,v}^{\lambda} = I_n + \lambda E_{u,v}$ (by the definition of $A_{u,v}^{\lambda}$) and $A_{v,u}^{\lambda} = I_n + \lambda E_{v,u}$ (by the definition of $A_{v,u}^{\lambda}$). Now,

$$\left(\underbrace{A_{u,v}^{\lambda}}_{=I_n+\lambda E_{u,v}} \right)^T = (I_n + \lambda E_{u,v})^T = \underbrace{\left(I_n\right)^T}_{\substack{=I_n \\ \text{(by Proposition 3.18 \textbf{(a)})}}} + \underbrace{\left(\lambda E_{u,v}\right)^T}_{\substack{=\lambda \left(E_{u,v}\right)^T \\ \text{(by Proposition 3.18 \textbf{(c)},} \\ \text{applied to } m=n \text{ and } A=E_{u,v})}}$$

$$\left(\begin{array}{c} \text{by Proposition 3.18 \textbf{(d)}}, \\ \text{applied to } m = n,\ A = I_n \text{ and } B = \lambda E_{u,v} \end{array} \right)$$

$$= I_n + \lambda \underbrace{\left(E_{u,v}\right)^T}_{=E_{v,u}} = I_n + \lambda E_{v,u} = A_{v,u}^{\lambda}.$$

Thus, Proposition 3.59 is proven. □

Proof of Proposition 3.60. **(a)** The definition of $A_{u,v}^{0}$ yields $A_{u,v}^{0} = I_n + \underbrace{0 E_{u,v}}_{=0_{n\times n}} = I_n + 0_{n\times n} = I_n$. This proves Proposition 3.60 **(a)**.

(b) *First proof:* Let λ and μ be two numbers. Proposition 3.55 (applied to μ instead of λ) tells us how the matrix $A_{u,v}^{\mu}$ looks like: Its diagonal entries are 1; its (u,v)-th entry is μ; all its remaining entries are 0. In particular, its v-th row is

$$\operatorname{row}_v \left(A_{u,v}^{\mu}\right) = (0, 0, \ldots, 0, 1, 0, 0, \ldots, 0) \tag{78}$$

(where the lonely 1 stands in the v-th position).

Proposition 3.56 (applied to $m = n$ and $C = A_{u,v}^{\mu}$) shows that $A_{u,v}^{\lambda} A_{u,v}^{\mu}$ is the $n \times n$-matrix obtained from $A_{u,v}^{\mu}$ by adding $\lambda \operatorname{row}_v \left(A_{u,v}^{\mu}\right)$ to the u-th row. Since

$$\begin{aligned} \lambda \operatorname{row}_v \left(A_{u,v}^{\mu}\right) &= \lambda \left(0, 0, \ldots, 0, 1, 0, 0, \ldots, 0\right) \qquad \text{(by (78))} \\ &= (0, 0, \ldots, 0, \lambda, 0, 0, \ldots, 0) \end{aligned}$$

(where the lonely λ stands in the v-th position), this shows that $A_{u,v}^{\lambda} A_{u,v}^{\mu}$ is the $n \times n$-matrix obtained from $A_{u,v}^{\mu}$ by adding $(0, 0, \ldots, 0, \lambda, 0, 0, \ldots, 0)$ to the u-th row. This addition has the effect that the (u,v)-th entry increases by λ, whereas all the other entries remain unchanged. The resulting matrix $A_{u,v}^{\lambda} A_{u,v}^{\mu}$ therefore has its (u,v)-th entry equal to $\mu + \lambda = \lambda + \mu$, whereas all its other entries are the same as

in $A_{u,v}^{\mu}$ (that is, the diagonal entries are 1 and the remaining entries are 0). But this is precisely how the matrix $A_{u,v}^{\lambda+\mu}$ looks like (because of Proposition 3.55, applied to $\lambda+\mu$ instead of λ). Hence, $A_{u,v}^{\lambda}A_{u,v}^{\mu} = A_{u,v}^{\lambda+\mu}$. This proves Proposition 3.60 **(b)**.

Second proof: We can also prove Proposition 3.60 **(b)** easily using Proposition 3.52: Indeed, we have $v \neq u$ (since u and v are distinct), so that $\delta_{v,u} = 0$. But Proposition 3.52 (applied to $m = n$, $p = n$, $x = u$ and $y = v$) yields $E_{u,v,n,n}E_{u,v,n,n} = \delta_{v,u}E_{u,v,n,n}$. Since $E_{u,v,n,n} = E_{u,v}$, this rewrites as $E_{u,v}E_{u,v} = \underbrace{\delta_{v,u}}_{=0}E_{u,v} = 0E_{u,v} = 0_{n\times n}$. Now, the definitions of $A_{u,v}^{\lambda}$ and $A_{u,v}^{\mu}$ yield $A_{u,v}^{\lambda} = I_n + \lambda E_{u,v}$ and $A_{u,v}^{\mu} = I_n + \mu E_{u,v}$. Multiplying these two equalities, we find

$$\begin{aligned}
A_{u,v}^{\lambda}A_{u,v}^{\mu} &= (I_n + \lambda E_{u,v})(I_n + \mu E_{u,v}) \\
&= \underbrace{I_n(I_n + \mu E_{u,v})}_{=I_n+\mu E_{u,v}} + \underbrace{\lambda E_{u,v}(I_n + \mu E_{u,v})}_{=\lambda E_{u,v}I_n+\lambda E_{u,v}\mu E_{u,v}} \\
&= I_n + \mu E_{u,v} + \lambda \underbrace{E_{u,v}I_n}_{=E_{u,v}} + \underbrace{\lambda E_{u,v}\mu E_{u,v}}_{=\lambda\mu E_{u,v}E_{u,v}} \\
&= I_n + \underbrace{\mu E_{u,v} + \lambda E_{u,v}}_{=(\mu+\lambda)E_{u,v}} + \lambda\mu \underbrace{E_{u,v}E_{u,v}}_{=0_{n\times n}} \\
&= I_n + \underbrace{(\mu+\lambda)}_{=\lambda+\mu}E_{u,v} + \underbrace{\lambda\mu 0_{n\times n}}_{=0_{n\times n}} = I_n + (\lambda+\mu)E_{u,v} + 0_{n\times n} = I_n + (\lambda+\mu)E_{u,v}.
\end{aligned}$$

Comparing this with

$$A_{u,v}^{\lambda+\mu} = I_n + (\lambda+\mu)E_{u,v} \qquad \left(\text{by the definition of } A_{u,v}^{\lambda+\mu}\right),$$

we obtain $A_{u,v}^{\lambda}A_{u,v}^{\mu} = A_{u,v}^{\lambda+\mu}$. This proves Proposition 3.60 **(b)** again.

(c) Proposition 3.60 **(b)** (applied to $\mu = -\lambda$) yields $A_{u,v}^{\lambda}A_{u,v}^{-\lambda} = A_{u,v}^{\lambda+(-\lambda)} = A_{u,v}^{0} = I_n$ (by Proposition 3.60 **(a)**).

But we can also apply Proposition 3.60 **(b)** to $-\lambda$ and λ instead of λ and μ. We thus obtain $A_{u,v}^{-\lambda}A_{u,v}^{\lambda} = A_{u,v}^{(-\lambda)+\lambda} = A_{u,v}^{0} = I_n$ (by Proposition 3.60 **(a)**).

The two equalities $A_{u,v}^{\lambda}A_{u,v}^{-\lambda} = I_n$ and $A_{u,v}^{-\lambda}A_{u,v}^{\lambda} = I_n$ show that $A_{u,v}^{-\lambda}$ is an inverse of $A_{u,v}^{\lambda}$. This proves Proposition 3.60 **(c)**. □

3.10. Unitriangular matrices are products of $A_{u,v}^{\lambda}$'s

We have already said that the matrices $A_{u,v}^{\lambda}$ are building blocks (of a sort). Before we explain what this means in detail, let us define a convenient word:

Definition 3.61. Let $n \in \mathbb{N}$. A *lower addition $n \times n$-matrix* means a matrix of the form $A_{u,v}^{\lambda}$, where λ is a number, and where u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u > v$. When n is clear from the context, we shall omit the "$n \times n$-" and simply say "lower addition matrix".

The name "lower addition matrix" is, again, not standard, but it will be useful for me in this chapter.

Example 3.62. If $n = 3$, then the lower addition 3×3-matrices are the matrices of the form

$$A_{2,1}^{\lambda} = \begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad A_{3,1}^{\lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda & 0 & 1 \end{pmatrix},$$

$$A_{3,2}^{\lambda} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda & 1 \end{pmatrix}$$

for all numbers λ. If $n = 2$, then the lower addition 2×2-matrices are the matrices of the form $A_{2,1}^{\lambda} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$ for all numbers λ. If $n = 1$ or $n = 0$, then there are no lower addition $n \times n$-matrices (because there is nowhere to put the λ, speaking visually).

It is clear that each lower addition matrix is lower-unitriangular[43] (because Proposition 3.55 shows that its entries above the diagonal are 0, and its diagonal entries are 1). Thus, every product of lower addition matrices is a product of lower-unitriangular matrices, and thus itself must be lower-unitriangular[44]. It turns out that the converse is also true: Every lower-unitriangular matrix is a product of lower addition matrices! This is a first, simple particular case of Gaussian elimination; let me state it as a theorem:

Theorem 3.63. Let $n \in \mathbb{N}$. An $n \times n$-matrix C is lower-unitriangular if and only if C is a product of lower addition matrices.

Keep in mind that "a product of lower addition matrices" (and any other product, unless declared otherwise) may contain zero factors (in which case it is empty, and thus equals the identity matrix) or one factor (in which case it equals that factor). Usually, of course, it will contain more than one factor.

Example 3.64. **(a)** The lower-unitriangular 2×2-matrix $\begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix}$ is a product of lower addition matrices: Namely, it equals $A_{2,1}^{5}$.

(b) The lower-unitriangular 1×1-matrix $\begin{pmatrix} 1 \end{pmatrix}$ is a product of lower addition matrices: Namely, it is the empty product. (Recall that the empty product of 1×1-matrices is defined to be I_1, and this is precisely our matrix $\begin{pmatrix} 1 \end{pmatrix}$.)

[43] See Definition 3.30 for the meaning of "lower-unitriangular".

[44] because Corollary 3.39 shows that any product of lower-unitriangular matrices is lower-unitriangular

(c) Let C be the lower-unitriangular 3×3-matrix $\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}$. Then, C is a product of lower addition matrices: Namely, it equals $A^a_{2,1} A^b_{3,1} A^c_{3,2}$.

Let us actually see how this representation of C can be found. We shall proceed by writing C as a product of one lower addition matrix with a second matrix C', which is still lower-triangular but has one less nonzero entry than C. We then will do the same with C', obtaining a third matrix C''; then, do the same with C'', and so on. At the end, we will be left with an identity matrix. In more detail:

Step 1: Let us get rid of the $(2,1)$-th entry of C (that is, turn this entry into a 0) by subtracting $a \operatorname{row}_1 C$ from row 2 of C. Denote the resulting matrix by C'. Thus, $C' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & c & 1 \end{pmatrix}$. (No entries other than the $(2,1)$-th one have been changed, because $a \operatorname{row}_1 C = (a, 0, 0)$.) Since C' was obtained from C by subtracting $a \operatorname{row}_1 C$ from row 2 of C, we can conversely obtain C from C' by adding $a \operatorname{row}_1 (C')$ to row 2 of C'. According to Proposition 3.56 (applied to n, 2, 1, a and C' instead of m, u, v, λ and C), this means that $C = A^a_{2,1} C'$.

Step 2: Let us get rid of the $(3,1)$-th entry of C' by subtracting $b \operatorname{row}_1 (C')$ from row 3 of C'. Denote the resulting matrix by C''. Thus, $C'' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix}$. (Again, no entries other than the $(3,1)$-th one have been changed.) Similarly to how we found that $C = A^a_{2,1} C'$ in Step 1, we now obtain $C' = A^b_{3,1} C''$.

Step 3: Let us get rid of the $(3,2)$-th entry of C'' by subtracting $c \operatorname{row}_2 (C')$ from row 3 of C''. Denote the resulting matrix by C'''. Thus, $C''' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. (Again, no entries other than the $(3,2)$-th one have been changed.) Similarly to how we found that $C = A^a_{2,1} C'$ in Step 1, we now obtain $C'' = A^c_{3,2} C'''$.

We have now removed all nonzero entries below the diagonal. Our final matrix C''' is simply the identity matrix: $C''' = I_3$. Combining the three equalities we have found, we obtain

$$\begin{aligned} C &= A^a_{2,1} \underbrace{C'}_{= A^b_{3,1} C''} = A^a_{2,1} A^b_{3,1} \underbrace{C''}_{= A^c_{3,2} C'''} = A^a_{2,1} A^b_{3,1} A^c_{3,2} \underbrace{C'''}_{= I_3} = A^a_{2,1} A^b_{3,1} A^c_{3,2} I_3 \\ &= A^a_{2,1} A^b_{3,1} A^c_{3,2}. \end{aligned} \tag{79}$$

Thus we have represented C as a product of lower addition matrices.

Notice that we need to be careful about the order in which we perform the above steps. We have first gotten rid of the $(2,1)$-st entry, then gotten rid of the $(3,1)$-th entry, and then gotten rid of the $(3,2)$-th entry. If we had tried to clean out the entries in a different order, we might have run into trouble. Namely, if we first get rid of the $(3,1)$-st entry, and then get rid of the $(3,2)$-th entry, then the $(3,1)$-th entry can become "polluted" again (i.e., the resulting matrix might again have a nonzero $(3,1)$-th entry). One way to avoid this kind of trouble is to clear out entries column by column: first clear out all entries in the 1-st column (except for the 1 on the diagonal); then clear out all entries in the 2-nd column (except for the 1 on the diagonal); and so on, moving left to right. (This is what we have done in our three steps above.)

The general proof of Theorem 3.63 follows the idea outlined in Example 3.64 **(c)**:

Proof of Theorem 3.63. $\Longleftarrow$:[45] We have already proven that every product of lower addition matrices is lower-unitriangular. Hence, if C is a product of lower addition matrices, then C is lower-unitriangular. This proves the $\Longleftarrow$ direction of Theorem 3.63.

$\Longrightarrow$:[46] We need to prove that if C is lower-unitriangular, then C is a product of lower addition matrices.

So let us assume that C is lower-unitriangular. Our goal is to prove that C is a product of lower addition matrices.

Let me introduce a notation first: A *downward row addition* shall mean a transformation that changes an $n \times m$-matrix (for some $m \in \mathbb{N}$) by adding a scalar multiple[47] of one of its rows to another row further down. In more formal terms: A *downward row addition* means a transformation of the form "add λ times the v-th row to the u-th row", for some fixed number λ and some fixed integers u and v satisfying $1 \leq v < u \leq n$. As we know from Proposition 3.56, this transformation amounts to multiplying a matrix by $A_{u,v}^{\lambda}$ from the left (i.e., this transformation sends any $n \times m$-matrix B to $A_{u,v}^{\lambda} B$); we shall therefore denote this transformation itself by $A_{u,v}^{\lambda}$ as well (hoping that the reader will not confuse the transformation with the matrix).

Here is an example (for $n = 4$): The downward row addition $A_{3,1}^{\lambda}$ is the transformation that changes a $4 \times m$-matrix by adding λ times the 1-st row to the 3-rd row.

[45] The symbol "$\Longleftarrow$" means that we are now going to prove the "$\Longleftarrow$ direction" of Theorem 3.63 (that is, we are going to prove that if C is a product of lower addition matrices, then C is lower-triangular). See Section 3.5 for more about this notation.

[46] The symbol "$\Longrightarrow$" means that we are now going to prove the "$\Longrightarrow$ direction" of Theorem 3.63 (that is, we are going to prove that if C is lower-triangular, then C is a product of lower addition matrices). See Section 3.5 for more about this notation.

[47] Recall that a *scalar multiple* of a matrix A means a matrix of the form λA with λ being a number. Row vectors are matrices (by definition), so we can talk about a scalar multiple of a row.

For example, it transforms the 4×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \\ a''' & b''' \end{pmatrix}$ into $\begin{pmatrix} a & b \\ a' & b' \\ a'' + \lambda a & b'' + \lambda b \\ a''' & b''' \end{pmatrix}$.

Notice that any downward row addition $A_{u,v}^{\lambda}$ is invertible: Namely, it can be undone by the downward row addition $A_{u,v}^{-\lambda}$. [48]

Notice that, for any downward row addition $A_{u,v}^{\lambda}$, the matrix $A_{u,v}^{\lambda}$ is a lower addition matrix. This is because $u > v$ (by the definition of a downward row addition).

I claim that we can transform the lower-triangular $n \times n$-matrix C into the identity matrix I_n by performing a sequence of downward row additions. Namely, we should proceed by the following method:[49]

- At first, our matrix is

$$C = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_{2,1} & 1 & 0 & \cdots & 0 \\ C_{3,1} & C_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & 1 \end{pmatrix}.$$

  In particular, its 1-st row is $(1, 0, 0, \ldots, 0)$ (same as the 1-st row of I_n).

- Now, we perform the downward row addition $A_{2,1}^{\lambda}$ (for an appropriate choice of λ, namely for $\lambda = -C_{2,1}$) to clear out the $(2,1)$-th entry of the matrix (i.e., to put a 0 where this entry stood). This does not affect any of the other entries, because the 1-st row of the matrix is $(1, 0, 0, \ldots, 0)$ (thus has only one nonzero entry). Similarly, we then perform the downward row addition $A_{3,1}^{\lambda}$ (with $\lambda = -C_{3,1}$) to clear out the $(3,1)$-th entry; then, we perform the downward row addition $A_{4,1}^{\lambda}$ (with $\lambda = -C_{4,1}$) to clear out the $(4,1)$-th entry; and so on,

[48] Here are two ways to prove this:

First proof: The downward row addition $A_{u,v}^{\lambda}$ transforms a matrix by adding λ times the v-th row to the u-th row. The downward row addition $A_{u,v}^{-\lambda}$ transforms a matrix by adding $-\lambda$ times the v-th row to the u-th row, i.e., by subtracting λ times the v-th row from the u-th row. Hence, these two row additions undo each other (i.e., if we perform one and then the other, then we arrive back at the matrix we have started with), because (for example) adding λ times the v-th row to the u-th row and then subtracting it back recovers the original u-th row. (Here, we have tacitly used the fact that $u \neq v$. If we had $u = v$, then adding λ times the v-th row to the u-th row would have changed the v-th row, and then subtracting it back would mean subtracting λ times a **changed** v-th row from the u-th row.) So we have shown that the downward row addition $A_{u,v}^{\lambda}$ can be undone by the downward row addition $A_{u,v}^{-\lambda}$. Qed.

Second proof: Proposition 3.60 **(c)** shows that the matrix $A_{u,v}^{-\lambda}$ is the inverse of the matrix $A_{u,v}^{\lambda}$. Hence, multiplying a matrix by $A_{u,v}^{-\lambda}$ undoes multiplying a matrix by $A_{u,v}^{\lambda}$. In other words, the downward row addition $A_{u,v}^{-\lambda}$ undoes the downward row addition $A_{u,v}^{\lambda}$. Qed.

[49] See the three-step procedure in Example 3.64 **(c)** for an illustration of this method.

ending with the downward row addition $A_{n,1}^{\lambda}$ (with $\lambda = -C_{n,1}$) to clear out the $(n,1)$-th entry. As the result, we have cleared out all entries in the 1-st column of our matrix, except for the 1 on the diagonal. In other words, our matrix now looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & C_{3,2} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & C_{n,2} & C_{n,3} & \cdots & 1 \end{pmatrix}.$$

In particular, its 2-nd row is $(0,1,0,0,\ldots,0)$ (same as the 2-nd row of I_n).

- Next, we similarly clear out all entries in the 2-nd column of the matrix, by performing the downward row additions $A_{3,2}^{\lambda}, A_{4,2}^{\lambda}, \ldots, A_{n,2}^{\lambda}$ (each time picking an appropriate value of λ). Again, this does not affect any of the other entries, because the 2-nd row of the matrix is $(0,1,0,0,\ldots,0)$. As the result, we have cleared out all entries in the 2-nd column of our matrix, except for the 1 on the diagonal. In other words, our matrix now looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & C_{n,3} & \cdots & 1 \end{pmatrix}.$$

In particular, its 3-rd row is $(0,0,1,0,0,\ldots,0)$ (same as the 3-rd row of I_n).

- Next, we similarly clear out all entries in the 3-rd column of the matrix, by performing the downward row additions $A_{4,3}^{\lambda}, A_{5,3}^{\lambda}, \ldots, A_{n,3}^{\lambda}$. As the result, we have cleared out all entries in the 3-rd column of our matrix, except for the 1 on the diagonal. In other words, our matrix now looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Don't mistake this for the identity matrix I_n: There might still be nonzero entries outside of the diagonal. They are just hidden by the $\cdots$ notation.

- We continue this process, clearing out column after column, until the n-th column is cleared out. By then, our matrix has become the identity matrix I_n.

Thus, we have found an algorithm to transform our matrix C into the identity matrix I_n by a sequence of downward row additions. Therefore, we can conversely transform the identity matrix I_n into C by a sequence of downward row additions[50]. Let us denote these downward row additions (used to transform I_n into C) by $A^{\lambda_1}_{u_1,v_1}, A^{\lambda_2}_{u_2,v_2}, \cdots, A^{\lambda_k}_{u_k,v_k}$, **numbered backwards** (i.e., starting from the one used last). Since each downward row addition $A^{\lambda}_{u,v}$ amounts to multiplying a matrix by the matrix $A^{\lambda}_{u,v}$ (that is, it sends any $n \times m$-matrix B to $A^{\lambda}_{u,v}B$), we thus conclude that

$$C = A^{\lambda_1}_{u_1,v_1} A^{\lambda_2}_{u_2,v_2} \cdots A^{\lambda_k}_{u_k,v_k} I_n = A^{\lambda_1}_{u_1,v_1} A^{\lambda_2}_{u_2,v_2} \cdots A^{\lambda_k}_{u_k,v_k}.$$

Thus, C is a product of lower addition matrices (because each of the matrices $A^{\lambda_1}_{u_1,v_1}, A^{\lambda_2}_{u_2,v_2}, \cdots, A^{\lambda_k}_{u_k,v_k}$ is a lower addition matrix[51]). This is precisely what we had to prove. This proves the $\Longrightarrow$ direction of Theorem 3.63. Hence, the proof of Theorem 3.63 is complete. □

Remark 3.65. Our proof of Theorem 3.63 (specifically, of its $\Longrightarrow$ direction) actually gives an explicit representation of a lower-unitriangular $n \times n$-matrix C as a product of lower addition matrices:

$$\begin{aligned} C = &\left(A_{2,1}^{C_{2,1}} A_{3,1}^{C_{3,1}} \cdots A_{n,1}^{C_{n,1}}\right) \left(A_{3,2}^{C_{3,2}} A_{4,2}^{C_{4,2}} \cdots A_{n,2}^{C_{n,2}}\right) \left(A_{4,3}^{C_{4,3}} A_{5,3}^{C_{5,3}} \cdots A_{n,3}^{C_{n,3}}\right) \\ &\cdots \left(A_{n-2,n-3}^{C_{n-2,n-3}} A_{n-1,n-3}^{C_{n-1,n-3}} A_{n,n-3}^{C_{n,n-3}}\right) \left(A_{n-1,n-2}^{C_{n-1,n-2}} A_{n,n-2}^{C_{n,n-2}}\right) \left(A_{n,n-1}^{C_{n,n-1}}\right) (I_n). \end{aligned}$$

This complicated product consists of n factors, each of which is itself a product. More precisely, the k-th factor is the product $A_{k+1,k}^{C_{k+1,k}} A_{k+2,k}^{C_{k+2,k}} \cdots A_{n,k}^{C_{n,k}}$ of all matrices $A_{i,k}^{C_{i,k}}$ with $i > k$; this corresponds to the $n-k$ downward row additions that clear out the non-diagonal entries in the k-th column of the matrix. In particular, the n-th factor is thus an empty product, hence equals I_n; we could thus leave it out (but we prefer to keep it in, for reasons of clarity).

The reason why the above explicit representation works is that in our process of clearing out zero entries, we have never modified any entry other than the one we were clearing out. Thus, for each (i,j), the (i,j)-th entry of our matrix remained equal to its original value $C_{i,j}$ up until the moment when we cleared it out.

Exercise 3.66. In Example 3.64 **(c)**, we have cleared out the three sub-diagonal (= below the diagonal) entries of C in a specific order: first the $(2,1)$-th entry, then the $(3,1)$-th entry, then the $(3,2)$-th entry. We thus obtained the formula $C = A^{a}_{2,1} A^{b}_{3,1} A^{c}_{3,2}$.

[50] because (as we have shown) any downward row addition $A^{\lambda}_{u,v}$ is invertible, and can be undone by another downward row addition

[51] Here we are using the fact that, for any downward row addition $A^{\lambda}_{u,v}$, the matrix $A^{\lambda}_{u,v}$ is a lower addition matrix.

We could also have proceeded differently: for instance, we could have cleared out the $(3,2)$-th entry first, then the $(3,1)$-th entry, then the $(2,1)$-th entry. This would have resulted in the formula $C = A_{3,2}^{c} A_{3,1}^{b-ac} A_{2,1}^{a}$. (Be aware of the $b-ac$ in the $A_{3,1}^{b-ac}$; this is because our first clearing operation has changed the $(3,1)$-th entry to $b-ac$.)

There is a total of 6 different orders in which we can try clearing out the three sub-diagonal entries of our 3×3-matrix C. The two of them just shown work; on the other hand, in Example 3.64 **(c)**, we have seen one order which does not (namely, starting with the $(3,1)$-th entry, then doing the $(3,2)$-th and then the $(2,1)$-th one). Of the remaining three, which ones work? And what formulas do they result in?

(By "work", I mean "work for every lower-unitriangular matrix C".)

3.11. The inverse of a lower-unitriangular matrix

Theorem 3.63 has the following neat consequence:

Theorem 3.67. Let $n \in \mathbb{N}$. Let A be a lower-unitriangular $n \times n$-matrix. Then, A is invertible, and its inverse A^{-1} is again lower-unitriangular.

Example 3.68. Let A be the lower-unitriangular 3×3-matrix $\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 6 & 1 \end{pmatrix}$. Theorem 3.67 (applied to $n = 3$) then claims that A is invertible, and that its inverse A^{-1} is again lower-unitriangular. This can easily be checked: We have $A^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -13 & -6 & 1 \end{pmatrix}$.

Proof of Theorem 3.67. Theorem 3.63 (applied to $C = A$) shows that A is lower-unitriangular if and only if A is a product of lower addition matrices. Hence, A is a product of lower addition matrices (since A is lower-unitriangular). In other words, A has the form $A = A_1 A_2 \cdots A_k$ for some $k \in \mathbb{N}$ and some k lower addition matrices $A_1, A_2, \ldots, A_k$. Consider these k and $A_1, A_2, \ldots, A_k$.

We are in one of the following two cases:

Case 1: We have $k \neq 0$.

Case 2: We have $k = 0$.

Let us deal with Case 1. In this case, we have $k \neq 0$; thus, k is a positive integer.

For each $i \in \{1, 2, \ldots, k\}$, the matrix A_i is invertible[52], and its inverse A_i^{-1} is again a lower addition matrix[53]. In other words, the matrices $A_1, A_2, \ldots, A_k$ are invertible, and their inverses $A_1^{-1}, A_2^{-1}, \ldots, A_k^{-1}$ are again lower addition matrices. In other words, $A_k^{-1}, A_{k-1}^{-1}, \ldots, A_1^{-1}$ are lower addition matrices.

Now, Proposition 3.12 shows that the matrix $A_1 A_2 \cdots A_k$ is invertible, and its inverse is $(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_1^{-1}$. Since $A = A_1 A_2 \cdots A_k$, this rewrites as follows: The matrix A is invertible, and its inverse is $A^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_1^{-1}$.

The equality $A^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_1^{-1}$ shows that A^{-1} is a product of lower addition matrices (since $A_k^{-1}, A_{k-1}^{-1}, \ldots, A_1^{-1}$ are lower addition matrices). But Theorem 3.63 (applied to $C = A^{-1}$) shows that A^{-1} is lower-unitriangular if and only if A^{-1} is a product of lower addition matrices. Hence, A^{-1} is lower-unitriangular (since A^{-1} is a product of lower addition matrices). This completes the proof of Theorem 3.67 in Case 1.

Case 2 is trivial (indeed, $A = I_n$ in this case) and is left to the reader.[54] Thus, Theorem 3.67 is proven in both Cases 1 and 2; this shows that Theorem 3.67 is always valid. □

A similar result holds for upper-triangular matrices:

Theorem 3.69. Let $n \in \mathbb{N}$. Let A be an upper-unitriangular $n \times n$-matrix. Then, A is invertible, and its inverse A^{-1} is again upper-unitriangular.

We could prove Theorem 3.69 by modifying our above proof of Theorem 3.67 (replacing "lower" by "upper" everywhere); of course, this would necessitate an analogue for Theorem 3.63 concerning upper-unitriangular instead of lower-unitriangular

[52] *Proof.* Let $i \in \{1, 2, \ldots, k\}$. We must show that the matrix A_i is invertible.

We know that A_i is a lower addition matrix (since $A_1, A_2, \ldots, A_k$ are lower addition matrices). In other words, A_i has the form $A_i = A_{u,v}^{\lambda}$, where λ is a number, and where u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u > v$ (by the definition of a "lower addition matrix"). Consider these λ, u and v. Proposition 3.60 **(c)** shows that the matrix $A_{u,v}^{\lambda}$ is invertible. In other words, the matrix A_i is invertible (since $A_i = A_{u,v}^{\lambda}$). This completes our proof.

[53] *Proof.* Let $i \in \{1, 2, \ldots, k\}$. We must show that A_i^{-1} is a lower addition matrix.

We know that A_i is a lower addition matrix (since $A_1, A_2, \ldots, A_k$ are lower addition matrices). In other words, A_i has the form $A_i = A_{u,v}^{\lambda}$, where λ is a number, and where u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u > v$ (by the definition of a "lower addition matrix"). Consider these λ, u and v. Proposition 3.60 **(c)** shows that the matrix $A_{u,v}^{\lambda}$ is invertible, and that its inverse is $\left(A_{u,v}^{\lambda}\right)^{-1} = A_{u,v}^{-\lambda}$.

But $A_{u,v}^{-\lambda}$ is a lower addition matrix (since $-\lambda$ is a number, and since $u > v$). In other words, A_i^{-1} is a lower addition matrix (since $\left(\underbrace{A_i}_{=A_{u,v}^{\lambda}}\right)^{-1} = \left(A_{u,v}^{\lambda}\right)^{-1} = A_{u,v}^{-\lambda}$). This completes our proof.

[54] Alternatively, our proof for Case 1 can be made to work in Case 2 as well, because Proposition 3.12 holds for $k = 0$ (as long as we define the empty product to be I_n). See Remark 3.13 for the details.

matrices (and upper addition matrices instead of lower addition matrices[55]). The proof of this analogue would then proceed similarly to our proof of Theorem 3.63, but again with some changes (e.g., instead of clearing out the columns from the first to the last, we would have to clear out the columns from the last to the first, and we would achieve this using "upward row additions").

However, we can also quickly derive Theorem 3.69 from Theorem 3.67 using transposes:

Proof of Theorem 3.69. We know that A is upper-unitriangular. Hence, by Proposition 3.41 **(a)**, we can conclude that A^T is lower-unitriangular. Therefore, we can apply Theorem 3.67 to A^T instead of A. As a result, we see that A^T is invertible, and its inverse $\left(A^T\right)^{-1}$ is again lower-unitriangular.

Hence, we can apply Proposition 3.18 **(f)** to A^T instead of A. Thus, we conclude that the matrix $\left(A^T\right)^T$ is invertible, and its inverse is $\left(\left(A^T\right)^T\right)^{-1} = \left(\left(A^T\right)^{-1}\right)^T$. Since $\left(A^T\right)^T = A$ (by Proposition 2.9, applied to $m = n$), this rewrites as follows: The matrix A is invertible, and its inverse is $A^{-1} = \left(\left(A^T\right)^{-1}\right)^T$.

It remains to prove that A^{-1} is upper-unitriangular. This is again quite easy: We have

$$\left(\underbrace{A^{-1}}_{=\left(\left(A^T\right)^{-1}\right)^T}\right)^T = \left(\left(\left(A^T\right)^{-1}\right)^T\right)^T = \left(A^T\right)^{-1}$$

(by Proposition 2.9, applied to n and $\left(A^T\right)^{-1}$ instead of m and A). Hence, $\left(A^{-1}\right)^T$ is lower-unitriangular (because we already know that $\left(A^T\right)^{-1}$ is lower-unitriangular). But Proposition 3.41 **(a)** (applied to A^{-1} instead of A) shows that A^{-1} is upper-unitriangular if and only if $\left(A^{-1}\right)^T$ is lower-unitriangular. Since $\left(A^{-1}\right)^T$ is lower-unitriangular, we thus conclude that A^{-1} is upper-unitriangular. This completes our proof of Theorem 3.69. □

3.12. (*) Products of strictly upper-triangular matrices

In this section, I shall give a different proof of Theorem 3.67, which is somewhat of a digression from our road towards Gaussian elimination, but has the advantage of showing off some other ideas. The proof is motivated by the following example:

[55] Of course, these upper addition matrices are defined exactly as you would expect: They are the matrices of the form $A^{\lambda}_{u,v}$, where λ is a number, and where u and v are two elements of $\{1, 2, \ldots, n\}$ satisfying $u < v$.

Example 3.70. Let A be a strictly lower-triangular 4×4-matrix. Thus, A has the form $\begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \end{pmatrix}$ for some numbers a, b, c, d, e, f. How do the powers of A look like?

Well, we can just compute the first few of them and see what happens:

$$A^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \end{pmatrix}, \qquad A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ ac & 0 & 0 & 0 \\ ae+bf & cf & 0 & 0 \end{pmatrix},$$

$$A^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ acf & 0 & 0 & 0 \end{pmatrix}, \qquad A^4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Two things strike the eye:

- We have $A^4 = 0_{4\times 4}$. As a consequence, every $n \geq 4$ satisfies $A^n = \underbrace{A^4}_{=0_{4\times 4}} A^{n-4} = 0_{4\times 4} A^{n-4} = 0_{4\times 4}$. So we actually know all the powers of A.
- Every time we pass from one power of A to the next (for example, from A^2 to A^3), the nonzero entries recede one level further towards the bottom-left corner. At the step from A^3 to A^4, they finally recede beyond that corner, so that only zeroes are left in the matrix.

The pattern seen in Example 3.70 actually generalizes: The powers of a strictly lower-triangular $n \times n$-matrix behave similarly, except that this time we have $A^n = 0_{n\times n}$ instead of $A^4 = 0_{4\times 4}$. Moreover, the same rule holds more generally if we multiply several strictly lower-triangular $n \times n$-matrices (instead of taking powers of one such matrix). In order to explore this more rigorously, I shall introduce a (nonstandard) notion:

Definition 3.71. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Let $k \in \mathbb{Z}$. We say that the matrix A is *k-lower-triangular* if and only if we have

$$A_{i,j} = 0 \qquad \text{whenever } i < j + k.$$

Example 3.72. Visually speaking, a square matrix A is k-lower-triangular if and only if its nonzero entries begin no earlier than k levels below the main diagonal. For example:

- Any 4×4-matrix is k-lower-triangular for $k \leq -3$.
- A 4×4-matrix is (-1)-lower-triangular if and only if it has the form $\begin{pmatrix} a & b & 0 & 0 \\ c & d & e & 0 \\ f & g & h & i \\ j & k & l & m \end{pmatrix}$ for some numbers $a, b, c, d, e, f, g, h, i, j, k, l, m$.
- A 4×4-matrix is 0-lower-triangular if and only if it has the form $\begin{pmatrix} a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \\ g & h & i & j \end{pmatrix}$ for some numbers $a, b, c, d, e, f, g, h, i, j$. This is, of course, the same as being lower-triangular.
- A 4×4-matrix is 1-lower-triangular if and only if it has the form $\begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ d & e & f & 0 \end{pmatrix}$ for some numbers a, b, c, d, e, f. This is the same as being strictly lower-triangular.
- A 4×4-matrix is 2-lower-triangular if and only if it has the form $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ b & c & 0 & 0 \end{pmatrix}$ for some numbers a, b, c.
- A 4×4-matrix is 3-lower-triangular if and only if it has the form $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \end{pmatrix}$ for some number a.
- A 4×4-matrix is 4-lower-triangular if and only if it has the form $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ (i.e., if and only if it is the zero matrix $0_{4\times 4}$). The same holds for 5-lower-triangular matrices and higher on.

We state some simple facts (which should have been clear from the example already):

Proposition 3.73. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix.

(a) The matrix A is k-lower-triangular for every integer k satisfying $k \leq -n+1$.

(b) The matrix A is 0-lower-triangular if and only if A is lower-triangular.

(c) The matrix A is 1-lower-triangular if and only if A is strictly lower-triangular.

(d) Let k be an integer such that $k \geq n$. The matrix A is k-lower-triangular if and only if $A = 0_{n \times n}$.

Proof of Proposition 3.73. **(a)** Let k be an integer satisfying $k \leq -n+1$. We shall show that the matrix A is k-lower-triangular. In order to do so, we must prove that

$$A_{i,j} = 0 \qquad \text{whenever } i < j + k \tag{80}$$

(because this is how "k-lower-triangular" was defined).

This is an example of a "vacuously true" statement. Let me explain the concept: A logical statement of the form "if $\mathcal{A}$, then $\mathcal{B}$" is said to be *vacuously true* if $\mathcal{A}$ never holds. For example, the statement "if a positive integer n is negative, then $n = 15$" is vacuously true, since a positive integer n will never be negative in the first place. Similarly, the statement "we have $n = m$ for any two integers n and m satisfying $n = m + \frac{1}{2}$" is also vacuously true, since two integers n and m will never satisfy $n = m + \frac{1}{2}$. (I have not worded this statement as an "if $\mathcal{A}$, then $\mathcal{B}$" statement, but I could easily have done so, by rewriting it as "if two integers n and m satisfy $n = m + \frac{1}{2}$, then $n = m$". The wording doesn't matter much.) Finally, the statement "every element of the empty set is a prime number" is also vacuously true, since there is no element of the empty set. (Again, you can rewrite this statement as "if a is an element of the empty set, then a is a prime number" in order to bring it into the "if $\mathcal{A}$, then $\mathcal{B}$" form.)

As the name suggests, mathematicians consider vacuously true statements to be true. The reasoning here is that, as long as you say nothing (and vacuously true statements say nothing, in a sense), you remain truthful.

We are now going to prove that the statement (80) is vacuously true. In other words, we are going to prove that two elements i and j of $\{1, 2, \ldots, n\}$ never satisfy $i < j + k$ in the first place.

In fact, let i and j be two elements of $\{1, 2, \ldots, n\}$ that satisfy $i < j + k$. Then, $i \geq 1$ (since $i \in \{1, 2, \ldots, n\}$) and $j \leq n$ (since $j \in \{1, 2, \ldots, n\}$), so that $\underbrace{j}_{\leq n} + \underbrace{k}_{\leq -n+1} \leq n + (-n+1) = 1 \leq i$ (since $i \geq 1$). This contradicts $i < j + k$.

Now, forget that we fixed i and j. We thus have found a contradiction **for any two elements** i and j of $\{1, 2, \ldots, n\}$ that satisfy $i < j + k$. This shows that there exist no two elements i and j of $\{1, 2, \ldots, n\}$ that satisfy $i < j + k$. In other words,

two elements i and j of $\{1,2,\ldots,n\}$ never satisfy $i < j+k$. Thus, the statement (80) is vacuously true, and therefore true. In other words, the matrix A is k-lower-triangular (by the definition of "k-lower-triangular"). This proves Proposition 3.73 **(a)**.

(Again, we shall be a lot briefer in proofs like this in the future.)

(b) Behold the following chain of equivalent statements[56]:

$$\begin{aligned}
&(A \text{ is 0-lower-triangular}) \\
\iff\ &\left(A_{i,j} = 0 \text{ whenever } i < j+0\right) \\
&\qquad (\text{because this is how "0-lower-triangular" is defined}) \\
\iff\ &\left(A_{i,j} = 0 \text{ whenever } i < j\right) \\
&\qquad (\text{here, we have replaced } j+0 \text{ by } j, \text{ since every } j \text{ satisfies } j+0=j) \\
\iff\ &(A \text{ is lower-triangular}) \\
&\qquad (\text{because this is how "lower-triangular" is defined}).
\end{aligned}$$

This proves Proposition 3.73 **(b)**.

(c) $\Longrightarrow$: We must prove that if A is 1-lower-triangular, then A is strictly lower-triangular.

Indeed, assume that A is 1-lower-triangular. In other words,

$$A_{i,j} = 0 \qquad \text{whenever } i < j+1 \tag{81}$$

(because this is how "1-lower-triangular" is defined).

Now, we have $A_{i,j} = 0$ whenever $i < j$ [57]. In other words, A is lower-triangular (by the definition of "lower-triangular"). Moreover, every $i \in \{1,2,\ldots,n\}$ satisfies $i < i+1$ and thus $A_{i,i} = 0$ (by (81), applied to $j = i$). In other words, all diagonal entries of A are 0.

So we have shown that the matrix A is lower-triangular, and that all its diagonal entries are 0. In other words, A is strictly lower-triangular (by the definition of "strictly lower-triangular"). This proves the $\Longrightarrow$ direction of Proposition 3.73 **(c)**.

$\Longleftarrow$: We must prove that if A is strictly lower-triangular, then A is 1-lower-triangular.

Indeed, assume that A is strictly lower-triangular. According to the definition of "strictly lower-triangular", this means that the matrix A is lower-triangular, and that all its diagonal entries are 0.

The matrix A is lower-triangular; in other words,

$$A_{i,j} = 0 \qquad \text{whenever } i < j \tag{82}$$

(according to the definition of "lower-triangular"). Also, all diagonal entries of A are 0; in other words,

$$A_{i,i} = 0 \qquad \text{for each } i \in \{1,2,\ldots,n\}. \tag{83}$$

[56] After each equivalence, we give a justification for why it is an equivalence.

[57] *Proof.* Let i and j be two elements of $\{1,2,\ldots,n\}$ such that $i < j$. Then, $i < j < j+1$. Hence, (81) shows that $A_{i,j} = 0$. Qed.

Now, we can easily see that

$$A_{i,j} = 0 \qquad \text{whenever } i < j+1$$

[58]. But this means precisely that A is 1-lower-triangular (because this is how "1-lower-triangular" is defined). Thus, we have shown that A is 1-lower-triangular. This proves the $\Longleftarrow$ direction of Proposition 3.73 **(c)**.

Now, the proof of Proposition 3.73 **(c)** is complete (since both its $\Longrightarrow$ and its $\Longleftarrow$ directions are proven).

(d) $\Longrightarrow$: We must prove that if A is k-lower-triangular, then $A = 0_{n\times n}$.

Indeed, assume that the matrix A is k-lower-triangular. In other words,

$$A_{i,j} = 0 \qquad \text{whenever } i < j+k \tag{84}$$

(by the definition of "k-lower-triangular").

Now, let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ be arbitrary. Then, $j \geq 1$ (since $j \in \{1,2,\ldots,n\}$) and $i \leq n$ (since $i \in \{1,2,\ldots,n\}$). Now, $\underbrace{j}_{\geq 1} + \underbrace{k}_{\geq n} \geq 1+n > n$, so that $n < j+k$ and thus $i \leq n < j+k$. Hence, $A_{i,j} = 0$ (by (84)). Comparing this with $(0_{n\times n})_{i,j} = 0$ (since each entry of the matrix $0_{n\times n}$ is 0), we obtain $A_{i,j} = (0_{n\times n})_{i,j}$.

Now, let us forget that we fixed i and j. We thus have shown that $A_{i,j} = (0_{n\times n})_{i,j}$ for each $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$. In other words, each entry of the matrix A equals the corresponding entry of the matrix $0_{n\times n}$. In other words, $A = 0_{n\times n}$. This proves the $\Longrightarrow$ direction of Proposition 3.73 **(d)**.

$\Longleftarrow$: We must prove that if $A = 0_{n\times n}$, then A is k-lower-triangular.

If i and j are two elements of $\{1,2,\ldots,n\}$ satisfying $i < j+k$, then

$$\begin{aligned} A_{i,j} &= (0_{n\times n})_{i,j} \qquad (\text{since } A = 0_{n\times n}) \\ &= 0 \qquad (\text{since each entry of the matrix } 0_{n\times n} \text{ is } 0)\,. \end{aligned}$$

In other words, $A_{i,j} = 0$ whenever $i < j+k$. But this means precisely that the matrix A is k-lower-triangular (by the definition of "k-lower-triangular"). Hence, we have shown that A is k-lower-triangular. This proves the $\Longleftarrow$ direction of Proposition 3.73 **(d)**.

[58] *Proof.* Let i and j be two elements of $\{1,2,\ldots,n\}$ such that $i < j+1$. We must prove that $A_{i,j} = 0$.

We are in one of the following two cases:

Case 1: We have $i = j$.

Case 2: We have $i \neq j$.

Let us first consider Case 1. In this case, we have $i = j$. Hence, $j = i$, so that $A_{i,j} = A_{i,i} = 0$ (by (83)). Hence, $A_{i,j} = 0$ is proven in Case 1.

Let us now consider Case 2. In this case, we have $i \neq j$. On the other hand, $i < j+1$, so that $i \leq (j+1)-1$ (since i and $j+1$ are integers). Thus, $i \leq (j+1)-1 = j$. Combining this with $i \neq j$, we obtain $i < j$. Hence, (82) shows that $A_{i,j} = 0$. Thus, $A_{i,j} = 0$ is proven in Case 2.

We now have proven $A_{i,j} = 0$ in each of the two Cases 1 and 2. Thus, $A_{i,j} = 0$ always holds, qed.

Now, Proposition 3.73 **(d)** is proven (since both its $\Longrightarrow$ and its $\Longleftarrow$ directions are proven). $\square$

Next, we state a fact which is crucial for our argument:

Proposition 3.74. Let $n \in \mathbb{N}$. Let p and q be two integers. Let A be a p-lower-triangular $n \times n$-matrix. Let B be a q-lower-triangular $n \times n$-matrix. Then, AB is a $(p+q)$-lower-triangular $n \times n$-matrix.

Remark 3.75. Proposition 3.74 generalizes Theorem 3.25 **(a)**. In fact, recall that an $n \times n$-matrix is 0-lower-triangular if and only if it is lower-triangular (by Proposition 3.73 **(b)**). Hence, applying Proposition 3.74 to $p = 0$ and $q = 0$, we obtain precisely Theorem 3.25 **(a)**.

Proof of Proposition 3.74. We shall imitate our above proof of Theorem 3.25 **(a)** as well as we can.

The matrix A is p-lower-triangular. In other words,

$$A_{i,j} = 0 \qquad \text{whenever } i < j + p \tag{85}$$

(because this is what it means for A to be p-lower-triangular).

The matrix B is q-lower-triangular. In other words,

$$B_{i,j} = 0 \qquad \text{whenever } i < j + q \tag{86}$$

(because this is what it means for B to be q-lower-triangular).

Now, fix two elements i and j of $\{1, 2, \ldots, n\}$ satisfying $i < j + (p+q)$. We shall prove that for every $k \in \{1, 2, \ldots, n\}$, we have

$$A_{i,k} B_{k,j} = 0. \tag{87}$$

[*Proof of (87):* Let $k \in \{1, 2, \ldots, n\}$. Then, we are in one of the following two cases:

Case 1: We have $i \geq k + p$.

Case 2: We have $i < k + p$.

We shall prove (87) in each of these two cases separately:

1. Let us first consider Case 1. In this case, we have $i \geq k + p$. Thus, $k + p \leq i < j + (p+q) = (j+q) + p$. Subtracting p from both sides of this inequality, we obtain $k < j + q$. Hence, we can apply (86) to k instead of i. As a result, we obtain $B_{k,j} = 0$. Hence, $A_{i,k} \underbrace{B_{k,j}}_{=0} = A_{i,k} 0 = 0$. Thus, (87) is proven in Case 1.

2. Let us now consider Case 2. In this case, we have $i < k + p$. Hence, we can apply (85) to k instead of j. As a result, we obtain $A_{i,k} = 0$. Hence, $\underbrace{A_{i,k}}_{=0} B_{k,j} = 0 B_{k,j} = 0$. Thus, (87) is proven in Case 2.

We have now proven (87) in both Cases 1 and 2. Thus, (87) is proven.]

Now, Proposition 2.19 **(a)** shows that

$$(AB)_{i,j} = \underbrace{A_{i,1}B_{1,j}}_{\substack{=0 \\ \text{(by (87),} \\ \text{applied to } k=1)}} + \underbrace{A_{i,2}B_{2,j}}_{\substack{=0 \\ \text{(by (87),} \\ \text{applied to } k=2)}} + \cdots + \underbrace{A_{i,m}B_{m,j}}_{\substack{=0 \\ \text{(by (87),} \\ \text{applied to } k=m)}}$$
$$= 0 + 0 + \cdots + 0 = 0.$$

Now, forget that we fixed i and j. We thus have shown that

$$(AB)_{i,j} = 0 \qquad \text{whenever } i < j + (p+q).$$

But this says precisely that the matrix AB is $(p+q)$-lower-triangular (by the definition of "$(p+q)$-lower-triangular"). Thus, Proposition 3.74 is proven. □

Using Proposition 3.74, we can show the following fact:

Corollary 3.76. Let $n \in \mathbb{N}$. Let $A_1, A_2, \ldots, A_k$ be k strictly lower-triangular $n \times n$-matrices (where $k \in \mathbb{N}$). Then, the $n \times n$-matrix $A_1 A_2 \cdots A_k$ is k-lower-triangular.

This is proven by induction similarly to how we proved Proposition 3.12 (we are actually copying the structure of that proof):

Proof of Corollary 3.76. We prove Corollary 3.76 by induction on k:

Induction base: If $k = 0$, then Corollary 3.76 says that the $n \times n$-matrix $A_1 A_2 \cdots A_0$ is 0-lower-triangular. But this is indeed true[59]. Hence, Corollary 3.76 holds for $k = 0$. This completes the induction base.

Induction step: Let ℓ be a positive integer. Assume (as our *induction hypothesis*) that Corollary 3.76 holds for $k = \ell$. In other words, for any ℓ strictly lower-triangular $n \times n$-matrices $A_1, A_2, \ldots, A_\ell$, the $n \times n$-matrix $A_1 A_2 \cdots A_\ell$ is ℓ-lower-triangular.

We must now show that Corollary 3.76 also holds for $k = \ell + 1$. So let us fix $\ell + 1$ strictly lower-triangular $n \times n$-matrices $A_1, A_2, \ldots, A_{\ell+1}$. We must then show that the $n \times n$-matrix $A_1 A_2 \cdots A_{\ell+1}$ is $(\ell+1)$-lower-triangular.

Clearly, $A_1, A_2, \ldots, A_\ell$ are ℓ strictly lower-triangular $n \times n$-matrices (since $A_1, A_2, \ldots, A_{\ell+1}$ are $\ell + 1$ strictly lower-triangular $n \times n$-matrices). Thus, we can apply our induction hypothesis, and conclude that the $n \times n$-matrix $A_1 A_2 \cdots A_\ell$ is ℓ-lower-triangular.

But the $n \times n$-matrix $A_{\ell+1}$ is 1-lower-triangular[60]. Thus, we can apply Proposition 3.74 to $p = \ell$, $q = 1$, $A = A_1 A_2 \cdots A_\ell$ and $B = A_{\ell+1}$. As a result, we

[59] *Proof.* The product $A_1 A_2 \cdots A_0$ is an empty product of $n \times n$-matrices, and thus equals I_n (by definition).

But Proposition 3.73 **(b)** (applied to $A = I_n$) shows that the matrix I_n is 0-lower-triangular if and only if I_n is lower-triangular. Hence, the matrix I_n is 0-lower-triangular (since I_n is lower-triangular). In other words, the $n \times n$-matrix $A_1 A_2 \cdots A_0$ is 0-lower-triangular (since $A_1 A_2 \cdots A_0 = I_n$). Qed.

[60] *Proof.* We know that $A_1, A_2, \ldots, A_{\ell+1}$ are $\ell + 1$ strictly lower-triangular $n \times n$-matrices. In particular, $A_{\ell+1}$ is a strictly lower-triangular $n \times n$-matrix.

But Proposition 3.73 **(c)** (applied to $A = A_{\ell+1}$) shows that the matrix $A_{\ell+1}$ is 1-lower-triangular if and only if $A_{\ell+1}$ is strictly lower-triangular. Hence, the matrix $A_{\ell+1}$ is 1-lower-triangular (since $A_{\ell+1}$ is strictly lower-triangular). Qed.

conclude that $(A_1 A_2 \cdots A_\ell) A_{\ell+1}$ is an $(\ell+1)$-lower-triangular $n \times n$-matrix. Since $(A_1 A_2 \cdots A_\ell) A_{\ell+1} = A_1 A_2 \cdots A_{\ell+1}$, this rewrites as follows: $A_1 A_2 \cdots A_{\ell+1}$ is an $(\ell+1)$-lower-triangular $n \times n$-matrix. In other words, the $n \times n$-matrix $A_1 A_2 \cdots A_{\ell+1}$ is $(\ell+1)$-lower-triangular. This is precisely what we wanted to show! Thus, Corollary 3.76 holds for $k = \ell + 1$. This completes the induction step. Thus, Corollary 3.76 is proven by induction. □

As a consequence of Corollary 3.76, we obtain the following fact, which we experimentally observed right after Example 3.70:

Corollary 3.77. Let $n \in \mathbb{N}$. Let $A_1, A_2, \ldots, A_n$ be n strictly lower-triangular $n \times n$-matrices. Then, $A_1 A_2 \cdots A_n = 0_{n \times n}$.

Proof of Corollary 3.77. Corollary 3.76 (applied to $k = n$) shows that the $n \times n$-matrix $A_1 A_2 \cdots A_n$ is n-lower-triangular. But we have $n \geq n$. Hence, Proposition 3.73 **(d)** (applied to $k = n$ and $A = A_1 A_2 \cdots A_n$) shows that the matrix $A_1 A_2 \cdots A_n$ is n-lower-triangular if and only if $A_1 A_2 \cdots A_n = 0_{n \times n}$. Hence, $A_1 A_2 \cdots A_n = 0_{n \times n}$ (since we know that the matrix $A_1 A_2 \cdots A_n$ is n-lower-triangular). Corollary 3.77 is thus proven. □

The following corollary is obtained as a particular case of Corollary 3.77 when we set all the n matrices $A_1, A_2, \ldots, A_n$ equal to one and the same matrix A:

Corollary 3.78. Let $n \in \mathbb{N}$. Let A be a strictly lower-triangular $n \times n$-matrix. Then, $A^n = 0_{n \times n}$.

Proof of Corollary 3.78. Clearly, $\underbrace{A, A, \ldots, A}_{n \text{ times}}$ are n strictly lower-triangular $n \times n$-matrices. Thus, Corollary 3.77 (applied to $A_i = A$) shows that $\underbrace{AA \cdots A}_{n \text{ times}} = 0_{n \times n}$. Now, $A^n = \underbrace{AA \cdots A}_{n \text{ times}} = 0_{n \times n}$. Corollary 3.78 is proven. □

Another corollary of the preceding results is the following:

Corollary 3.79. Let $n \in \mathbb{N}$. Let k be a positive integer. Let A be a strictly lower-triangular $n \times n$-matrix. Then, the $n \times n$-matrix A^k is strictly lower-triangular.

Proof of Corollary 3.79. Clearly, $\underbrace{A, A, \ldots, A}_{k \text{ times}}$ are k strictly lower-triangular $n \times n$-matrices. Thus, Corollary 3.76 (applied to $A_i = A$) shows that the $n \times n$-matrix $\underbrace{AA \cdots A}_{k \text{ times}}$ is k-lower-triangular. Since $\underbrace{AA \cdots A}_{k \text{ times}} = A^k$, this rewrites as follows: The $n \times n$-matrix A^k is k-lower-triangular. In other words,

$$\left(A^k\right)_{i,j} = 0 \qquad \text{whenever } i < j + k \tag{88}$$

(according to the definition of "k-lower-triangular").

But k is a positive integer. Thus, $1 \leq k$, so that $j + \underbrace{1}_{\leq k} \leq j + k$ for every $j \in \{1, 2, \ldots, n\}$. Hence, every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfying $i < j + 1$ must also satisfy $i < j + k$ (since $i < j + 1 \leq j + k$) and consequently $\left(A^k\right)_{i,j} = 0$ (by (88)). In other words,

$$\left(A^k\right)_{i,j} = 0 \qquad \text{whenever } i < j + 1.$$

In other words, the matrix A^k is 1-lower-triangular (by the definition of "1-lower-triangular").

But Proposition 3.73 **(c)** (applied to A^k instead of A) shows that the matrix A^k is 1-lower-triangular if and only if A^k is strictly lower-triangular. Hence, A^k is strictly lower-triangular (since A^k is 1-lower-triangular). This proves Corollary 3.79. □

So much for products of strictly lower-triangular matrices. What about their sums?

Proposition 3.80. Let $n \in \mathbb{N}$. Let $A_1, A_2, \ldots, A_k$ be k strictly lower-triangular $n \times n$-matrices (where $k \in \mathbb{N}$). Then, $A_1 + A_2 + \cdots + A_k$ is a strictly lower-triangular $n \times n$-matrix.

Proof of Proposition 3.80. This is left to the reader. (What makes this proof easy is that matrices are added entry by entry.) □

Just as trivial is the following fact:

Proposition 3.81. Let $n \in \mathbb{N}$. Let A be a strictly lower-triangular $n \times n$-matrix. Then, $-A$ is a strictly lower-triangular $n \times n$-matrix.

Proof of Proposition 3.81. Left to the reader. □

Let us also state an analogue of Proposition 3.32 for lower-triangular matrices:

Proposition 3.82. (a) Each lower-unitriangular matrix is invertibly lower-triangular.

(b) Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Then, A is lower-unitriangular if and only if $I_n - A$ is strictly lower-triangular.

Proof of Proposition 3.82. This is proven in the same way as we proved Proposition 3.32 (once the obvious modifications are made). □

Next, we show another simple fact:

Proposition 3.83. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix such that $A^n = 0_{n \times n}$. Then, the matrix $I_n - A$ is invertible, and its inverse is $(I_n - A)^{-1} = A^0 + A^1 + \cdots + A^{n-1}$.

Proof of Proposition 3.83. Let B be the $n \times n$-matrix $A^0 + A^1 + \cdots + A^{n-1}$.

Multiplying the equalities $A = A$ and $B = A^0 + A^1 + \cdots + A^{n-1}$, we obtain

$$\begin{aligned} AB &= A\left(A^0 + A^1 + \cdots + A^{n-1}\right) = \underbrace{AA^0}_{=A^1} + \underbrace{AA^1}_{=A^2} + \cdots + \underbrace{AA^{n-1}}_{=A^n} \\ &= A^1 + A^2 + \cdots + A^n = \left(A^1 + A^2 + \cdots + A^{n-1}\right) + \underbrace{A^n}_{=0_{n\times n}} = A^1 + A^2 + \cdots + A^{n-1}. \end{aligned}$$

Now,

$$\begin{aligned} (I_n - A)\, B &= \underbrace{I_n B}_{=B=A^0+A^1+\cdots+A^{n-1}} - \underbrace{AB}_{=A^1+A^2+\cdots+A^{n-1}} \\ &= \left(A^0 + A^1 + \cdots + A^{n-1}\right) - \left(A^1 + A^2 + \cdots + A^{n-1}\right) = A^0 = I_n. \end{aligned}$$

A similar argument shows that $B\,(I_n - A) = I_n$. (To be more precise: This is proven by multiplying the equalities $B = A^0 + A^1 + \cdots + A^{n-1}$ and $A = A$, as opposed to $A = A$ and $B = A^0 + A^1 + \cdots + A^{n-1}$.)

The two equalities $(I_n - A)\, B = I_n$ and $B\,(I_n - A) = I_n$ show that the matrix B is an inverse of $I_n - A$. Thus, the matrix $I_n - A$ is invertible, and its inverse is $(I_n - A)^{-1} = B$. Hence, $(I_n - A)^{-1} = B = A^0 + A^1 + \cdots + A^{n-1}$. The proof of Proposition 3.83 is thus complete. □

Remark 3.84. **(a)** The above proof of Proposition 3.83 might look like a slick and artful trick. However, it is actually an incarnation of a well-known idea: the same idea that enters in the proof of the infinite-sum formula

$$\frac{1}{1-a} = a^0 + a^1 + a^2 + \cdots \qquad \text{for any real number } a \text{ with } -1 < a < 1.$$

The main difference here is that we are working with a matrix A instead of a real number a, and that the infinite sum $a^0 + a^1 + a^2 + \cdots$ is replaced by a **finite** sum $A^0 + A^1 + \cdots + A^{n-1}$ (because all the powers $A^n, A^{n+1}, A^{n+2}, \ldots$ equal the zero matrix).

(b) Proposition 3.83 requires an $n \times n$ matrix A satisfying $A^n = 0_{n \times n}$. How do we find such matrices?

Corollary 3.78 shows that every strictly lower-triangular $n \times n$-matrix A has this property; this gives us an infinite supply of such matrices (at least for $n \geq 2$). Similarly, every strictly upper-triangular $n \times n$-matrix A has this property. But there are also others: For example, if A is the 2×2-matrix $\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$, then A also satisfies $A^2 = 0_{2\times 2}$, despite not being triangular.

We can now prove Theorem 3.67 again:

Second proof of Theorem 3.67. Proposition 3.82 **(b)** shows that A is lower-unitriangular if and only if $I_n - A$ is strictly lower-triangular. Hence, $I_n - A$ is strictly lower-triangular (because A is lower-unitriangular).

Let $C = I_n - A$. Thus, C is strictly lower-triangular (since $I_n - A$ is strictly lower-triangular). Hence, Corollary 3.78 (applied to C instead of A) shows that $C^n = 0_{n \times n}$. Proposition 3.83 (applied to C instead of A) thus shows that the matrix $I_n - C$ is invertible, and that its inverse is $(I_n - C)^{-1} = C^0 + C^1 + \cdots + C^{n-1}$.

We have $I_n - \underbrace{C}_{= I_n - A} = I_n - (I_n - A) = A$. Now, we know that the matrix $I_n - C$ is invertible. In other words, the matrix A is invertible (since $I_n - C = A$). It remains to check that its inverse A^{-1} is again lower-unitriangular.

We assume WLOG that n is positive (since otherwise, there is nothing to check). We have

$$\begin{aligned}(I_n - C)^{-1} &= C^0 + C^1 + \cdots + C^{n-1} = \underbrace{C^0}_{= I_n} + \left(C^1 + C^2 + \cdots + C^{n-1}\right) \\ &= I_n + \left(C^1 + C^2 + \cdots + C^{n-1}\right).\end{aligned}$$

Since $I_n - C = A$, this rewrites as

$$A^{-1} = I_n + \left(C^1 + C^2 + \cdots + C^{n-1}\right).$$

Subtracting this equality from $I_n = I_n$, we obtain

$$\begin{aligned}I_n - A^{-1} &= I_n - \left(I_n + \left(C^1 + C^2 + \cdots + C^{n-1}\right)\right) \\ &= -\left(C^1 + C^2 + \cdots + C^{n-1}\right). \qquad (89)\end{aligned}$$

But recall that the matrix C is strictly lower-triangular. Hence, for every positive integer k, the $n \times n$-matrix C^k is strictly lower-triangular (by Corollary 3.79, applied to C instead of A). Thus, the matrices $C^1, C^2, \ldots, C^{n-1}$ are $n-1$ strictly lower-triangular $n \times n$-matrices. Proposition 3.80 (applied to $k = n-1$ and $A_k = C^k$) thus shows that $C^1 + C^2 + \cdots + C^{n-1}$ is a strictly lower-triangular $n \times n$-matrix. Hence, Proposition 3.81 (applied to $C^1 + C^2 + \cdots + C^{n-1}$ instead of A) yields that $-\left(C^1 + C^2 + \cdots + C^{n-1}\right)$ is a strictly lower-triangular $n \times n$-matrix. In light of (89), this rewrites as follows: $I_n - A^{-1}$ is a strictly lower-triangular $n \times n$-matrix.

But Proposition 3.82 **(b)** (applied to A^{-1} instead of A) shows that A^{-1} is lower-unitriangular if and only if $I_n - A^{-1}$ is strictly lower-triangular. Since we have just seen that $I_n - A^{-1}$ is strictly lower-triangular, we can therefore conclude that A^{-1} is lower-unitriangular. Thus, the second proof of Theorem 3.67 is complete. □

Of course, an analogous argument (with some inequality signs turned around, and some "lower"s replaced by "upper"s) can be used to prove Theorem 3.69.

3.13. The λ-scaling matrices S_u^λ

Now we shall explore another kind of square matrices not unlike the matrices $A_{u,v}^\lambda$ from Definition 3.53[61]:

Definition 3.85. Let $n \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a number. Then, S_u^λ shall denote the $n \times n$-matrix $I_n + (\lambda - 1) E_{u,u}$ (where $E_{u,u}$ means the $n \times n$-matrix $E_{u,u}$, that is, $E_{u,u,n,n}$).

A few remarks about the notation:

(a) The superscript λ in the notation "S_u^λ" is not an exponent; i.e., the matrix S_u^λ is not the λ-th power of some matrix S_u. Instead, it is just an argument that we have chosen to write as a superscript instead of a subscript. So the role of the λ in "S_u^λ" is completely different from the role of the -1 in "A_1^{-1}" in Proposition 3.12.

(b) To be really precise, we ought to denote S_u^λ by $S_{u,n}^\lambda$, because it depends on n. (This is similar to how we ought to denote $E_{u,v}$ by $E_{u,v,n,n}$.) But the n will be really clear from the context almost every time we deal with these matrices, so we shall keep it out of our notation.

(c) The notation S_u^λ is not standard in the literature, but I will use this notation in the following.

Example 3.86. Let $n = 4$. Then,

$$S_3^\lambda = I_n + (\lambda - 1) E_{3,3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{and}$$

$$S_4^\lambda = I_n + (\lambda - 1) E_{4,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

The pattern that you see on these examples is true in general:

Proposition 3.87. Let $n \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a number. Then, the matrix S_u^λ has the following entries:

- Its (u, u)-th entry is λ.
- All its other diagonal entries are 1.
- All its remaining entries are 0.

[61] The present section imitates the structure of Section 3.8; this is, of course, fully intentional.

Proposition 3.87 can be rewritten as follows: The matrix S_u^λ (where $n \in \mathbb{N}$, where $u \in \{1, 2, \ldots, n\}$, and where λ is a number) is the $n \times n$-identity matrix I_n with the (u, u)-th entry replaced by λ.

Proof of Proposition 3.87. Recall that $E_{u,u}$ is the $n \times n$-matrix whose (u, u)-th entry is 1 and whose all other entries are 0 (indeed, this is how $E_{u,u}$ was defined). Since matrices are scaled entry by entry, we can therefore conclude how $(\lambda - 1) E_{u,u}$ looks like: Namely, $(\lambda - 1) E_{u,u}$ is the $n \times n$-matrix whose (u, u)-th entry is $(\lambda - 1) \cdot 1 = \lambda - 1$ and whose all other entries are $(\lambda - 1) \cdot 0 = 0$. Thus, we know the following:

- The matrix I_n is the $n \times n$-matrix whose diagonal entries[62] are 1, and whose all other entries are 0.
- The matrix $(\lambda - 1) E_{u,u}$ is the $n \times n$-matrix whose (u, u)-th entry is $\lambda - 1$, and whose all other entries are 0.

Since matrices are added entry by entry, we can thus infer how $I_n + (\lambda - 1) E_{u,u}$ looks like: Namely, the matrix $I_n + (\lambda - 1) E_{u,u}$ is the $n \times n$-matrix whose (u, u)-th entry is $1 + (\lambda - 1) = \lambda$, whose all other diagonal entries are $1 + 0 = 1$, and whose all other entries are $0 + 0 = 0$. Since $I_n + (\lambda - 1) E_{u,u} = S_u^\lambda$, this rewrites as follows: The matrix S_u^λ is the $n \times n$-matrix whose (u, u)-th entry is λ, whose all other diagonal entries are 1, and whose all other entries are 0. This proves Proposition 3.87. □

We can next see what happens to a matrix when it is multiplied by S_u^λ:

Proposition 3.88. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a number. Let C be an $n \times m$-matrix. Then, $S_u^\lambda C$ is the $n \times m$-matrix obtained from C by scaling the u-th row by λ.

(Recall that the rows of C are row vectors, and thus are scaled entry by entry. Hence, scaling the u-th row by λ means multiplying each entry of the u-th row of C by λ.)

Example 3.89. Let $n = 3$ and $m = 2$. Let C be the 3×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix}$. Let λ be a number. Then, Proposition 3.88 (applied to $u = 2$) claims that $S_2^\lambda C$ is the 3×2-matrix obtained from C by scaling the 2-nd row by λ. A computation confirms this claim:

$$S_2^\lambda C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix} = \begin{pmatrix} a & b \\ \lambda a' & \lambda b' \\ a'' & b'' \end{pmatrix}.$$

[62] This includes the (u, u)-th entry.

Proposition 3.90. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a number. Let C be an $m \times n$-matrix. Then, CS_u^λ is the $m \times n$-matrix obtained from C by scaling the u-th column by λ.

I will prove Proposition 3.88 in Section 3.14 below (although by now, this proof could be a simple exercise).

We shall refer to the matrix S_u^λ defined in Definition 3.85 as a "λ-scaling matrix"; it is the second of the previously announced three kinds of elementary matrices.

Here are a few more properties of λ-scaling matrices:

Proposition 3.91. Let $n \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a number. Then, $\left(S_u^\lambda\right)^T = S_u^\lambda$.

Proposition 3.92. Let $n \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$.

(a) We have $S_u^1 = I_n$.

(b) If λ and μ are two numbers, then $S_u^\lambda S_u^\mu = S_u^{\lambda\mu}$.

(c) Let λ be a nonzero number. Then, the matrix S_u^λ is invertible, and its inverse is $\left(S_u^\lambda\right)^{-1} = S_u^{1/\lambda}$.

A proof of this proposition will also be given in Section 3.14.

3.14. (*) Some proofs about the λ-scaling matrices

Proof of Proposition 3.88. Clearly, $S_u^\lambda C$ is an $n \times m$-matrix.

Proposition 3.44 (applied to n, m and u instead of m, p and v) shows that $E_{u,u}C$ is the $n \times m$-matrix whose u-th row is the u-th row of C, and whose all other rows are filled with zeroes. Thus,

$$\operatorname{row}_u\left(E_{u,u}C\right) = \operatorname{row}_u C \tag{90}$$

(since the u-the row of $E_{u,u}C$ is the u-th row of C) and

$$\operatorname{row}_i\left(E_{u,u}C\right) = 0_{1\times m} \qquad \text{for every } i \in \{1, 2, \ldots, n\} \text{ satisfying } i \neq u \tag{91}$$

(since all other rows of $E_{u,u}C$ are filled with zeroes).

But recall that matrices are added entry by entry. Thus, matrices are also added by row by row – i.e., if U and V are two $n \times m$-matrices, then any row of $U + V$ is the sum of the corresponding rows of U and of V. In other words, if U and V are two $n \times m$-matrices, then

$$\operatorname{row}_i(U + V) = \operatorname{row}_i U + \operatorname{row}_i V \qquad \text{for every } i \in \{1, 2, \ldots, n\}. \tag{92}$$

Also, if U is an $n \times m$-matrix, then every number μ satisfies

$$\operatorname{row}_i(\mu U) = \mu \operatorname{row}_i U \qquad \text{for every } i \in \{1, 2, \ldots, n\} \tag{93}$$

(since matrices are scaled entry by entry).

We have

$$\underbrace{S_u^\lambda}_{=I_n+(\lambda-1)E_{u,u}} C = (I_n + (\lambda-1)\,E_{u,u})\,C = \underbrace{I_n C}_{=C} + (\lambda-1)\,E_{u,u}C$$
$$= C + (\lambda-1)\,E_{u,u}C.$$

Hence, for each $i \in \{1,2,\ldots,n\}$, we have

$$\operatorname{row}_i\left(\underbrace{S_u^\lambda C}_{=C+(\lambda-1)E_{u,u}C}\right) = \operatorname{row}_i\left(C + (\lambda-1)\,E_{u,u}C\right) = \operatorname{row}_i C + \underbrace{\operatorname{row}_i\left((\lambda-1)\,E_{u,u}C\right)}_{\substack{=(\lambda-1)\operatorname{row}_i(E_{u,u}C) \\ \text{(by (93), applied} \\ \text{to } \mu=\lambda-1)}}$$
$$\text{(by (92), applied to } U = C \text{ and } V = (\lambda-1)\,E_{u,u}C)$$
$$= \operatorname{row}_i C + (\lambda-1)\operatorname{row}_i\left(E_{u,u}C\right). \tag{94}$$

Now, we must prove that $S_u^\lambda C$ is the $n \times m$-matrix obtained from C by scaling the u-th row by λ. In other words, we must prove the following two claims:

> *Claim 1:* The u-th row of the $n \times m$-matrix $S_u^\lambda C$ equals λ times the u-th row of C.
>
> *Claim 2:* For each $i \in \{1,2,\ldots,n\}$ satisfying $i \neq u$, the i-th row of the $n \times m$-matrix $S_u^\lambda C$ equals the i-th row of C.

Proof of Claim 1: The u-th row of the $n \times m$-matrix $S_u^\lambda C$ is

$$\operatorname{row}_u\left(S_u^\lambda C\right) = \operatorname{row}_u C + (\lambda-1)\underbrace{\operatorname{row}_u\left(E_{u,u}C\right)}_{\substack{=\operatorname{row}_u C \\ \text{(by (90))}}} \qquad \text{(by (94), applied to } i = u)$$
$$= \operatorname{row}_u C + (\lambda-1)\operatorname{row}_u C = \underbrace{(1 + (\lambda-1))}_{=\lambda}\operatorname{row}_u C = \lambda \operatorname{row}_u C.$$

In other words, the u-th row of the $n \times m$-matrix $S_u^\lambda C$ equals λ times the u-th row of C. This proves Claim 1.

Proof of Claim 2: Let $i \in \{1,2,\ldots,n\}$ be such that $i \neq u$. Then, the i-th row of the $n \times m$-matrix $S_u^\lambda C$ is

$$\operatorname{row}_i\left(S_u^\lambda C\right) = \operatorname{row}_i C + (\lambda-1)\underbrace{\operatorname{row}_i\left(E_{u,u}C\right)}_{\substack{=0_{1\times m} \\ \text{(by (91))}}} \qquad \text{(by (94))}$$
$$= \operatorname{row}_i C + \underbrace{(\lambda-1)\,0_{1\times m}}_{=0_{1\times m}} = \operatorname{row}_i C + 0_{1\times m} = \operatorname{row}_i C.$$

In other words, the i-th row of the $n \times m$-matrix $S_u^\lambda C$ equals the i-th row of C. This proves Claim 2.

Now, we have proven both Claim 1 and Claim 2; this completes the proof of Proposition 3.88. □

The proof of Proposition 3.90 is analogous.

Proof of Proposition 3.91. Very easy and left to the reader. (See the proof of Proposition 3.59 for inspiration.) □

Proof of Proposition 3.92. **(a)** The definition of S_u^1 yields $S_u^1 = I_n + \underbrace{(1-1)\, E_{u,u}}_{=0E_{u,u}=0_{n\times n}} = I_n + 0_{n\times n} = I_n$. This proves Proposition 3.92 **(a)**.

(b) *First proof:* Let λ and μ be two numbers. Proposition 3.87 (applied to μ instead of λ) tells us how the matrix S_u^μ looks like: Its (u,u)-th entry is μ; all its other diagonal entries are 1; all its remaining entries are 0. In particular, its u-th row is

$$\operatorname{row}_u\left(S_u^\mu\right) = (0,0,\ldots,0,\mu,0,0,\ldots,0) \tag{95}$$

(where the lonely μ stands in the u-th position).

Proposition 3.88 (applied to $m = n$ and $C = S_u^\mu$) shows that $S_u^\lambda S_u^\mu$ is the $n \times n$-matrix obtained from S_u^μ by scaling the u-th row by λ. Thus, its u-th row is

$$\begin{aligned}
\operatorname{row}_u\left(S_u^\lambda S_u^\mu\right) &= \lambda \underbrace{\operatorname{row}_u\left(S_u^\mu\right)}_{\substack{=(0,0,\ldots,0,\mu,0,0,\ldots,0) \\ \text{(by (95))}}} = \lambda\,(0,0,\ldots,0,\mu,0,0,\ldots,0) \\
&= (\lambda 0, \lambda 0, \ldots, \lambda 0, \lambda\mu, \lambda 0, \lambda 0, \ldots, \lambda 0) \\
&= (0,0,\ldots,0,\lambda\mu,0,0,\ldots,0)
\end{aligned}$$

(where the lonely μ or $\lambda\mu$ stands in the u-th position, as before). This is the same as the u-th row of S_u^μ, except that the u-th entry has become $\lambda\mu$ (whereas in the u-th row of S_u^μ it used to be μ). All other rows of $S_u^\lambda S_u^\mu$ are equal to the corresponding rows of S_u^μ (since $S_u^\lambda S_u^\mu$ was obtained from S_u^μ by scaling the u-th row by λ). Summarizing, we thus conclude that the matrix $S_u^\lambda S_u^\mu$ differs from the matrix S_u^μ in only one entry, namely the (u,u)-th entry[63]; and this (u,u)-th entry is $\lambda\mu$ (for the matrix $S_u^\lambda S_u^\mu$). Since we already know how the matrix S_u^μ looks like (namely, its (u,u)-th entry is μ; all its other diagonal entries are 1; all its remaining entries are 0), we thus can conclude how the matrix $S_u^\lambda S_u^\mu$ looks like: Its (u,u)-th entry is $\lambda\mu$; all its other diagonal entries are 1; all its remaining entries are 0. But this is precisely how the matrix $S_u^{\lambda\mu}$ looks like (because of Proposition 3.87, applied to $\lambda\mu$ instead of λ). Hence, $S_u^\lambda S_u^\mu = S_u^{\lambda\mu}$. This proves Proposition 3.92 **(b)**.

Second proof: We can also prove Proposition 3.92 **(b)** easily using Proposition 3.52: Indeed, we have $u = u$ and thus $\delta_{u,u} = 1$. But Proposition 3.52 (applied to $m = n$,

[63] If it differs from it at all! If $\lambda = 1$ or $\mu = 0$, then the matrices $S_u^\lambda S_u^\mu$ and S_u^μ are completely equal (including in their (u,u)-th entries).

$p = n$, $x = u$ and $y = u$) yields $E_{u,u,n,n}E_{u,u,n,n} = \delta_{u,u}E_{u,u,n,n}$. Since $E_{u,u,n,n} = E_{u,u}$, this rewrites as $E_{u,u}E_{u,u} = \underbrace{\delta_{u,u}}_{=1} E_{u,u} = 1E_{u,u} = E_{u,u}$. Now, the definitions of S_u^λ and S_u^μ yield $S_u^\lambda = I_n + (\lambda - 1)\, E_{u,u}$ and $S_u^\mu = I_n + (\mu - 1)\, E_{u,u}$. Multiplying these two equalities, we find

$$\begin{aligned}
S_u^\lambda S_u^\mu &= (I_n + (\lambda - 1)\, E_{u,u})\,(I_n + (\mu - 1)\, E_{u,u}) \\
&= \underbrace{I_n\,(I_n + (\mu - 1)\, E_{u,u})}_{=I_n + (\mu-1)E_{u,u}} + \underbrace{(\lambda - 1)\, E_{u,u}\,(I_n + (\mu - 1)\, E_{u,u})}_{=(\lambda-1)E_{u,u}I_n + (\lambda-1)E_{u,u}(\mu-1)E_{u,u}} \\
&= I_n + (\mu - 1)\, E_{u,u} + (\lambda - 1)\underbrace{E_{u,u}I_n}_{=E_{u,u}} + \underbrace{(\lambda - 1)\, E_{u,u}\,(\mu - 1)\, E_{u,u}}_{=(\lambda-1)(\mu-1)E_{u,u}E_{u,u}} \\
&= I_n + \underbrace{(\mu - 1)\, E_{u,u} + (\lambda - 1)\, E_{u,u}}_{=((\mu-1)+(\lambda-1))E_{u,u}} + (\lambda - 1)\,(\mu - 1)\underbrace{E_{u,u}E_{u,u}}_{=E_{u,u}} \\
&= I_n + \underbrace{((\mu - 1) + (\lambda - 1))\, E_{u,u} + (\lambda - 1)\,(\mu - 1)\, E_{u,u}}_{=((\mu-1)+(\lambda-1)+(\lambda-1)(\mu-1))E_{u,u}} \\
&= I_n + \underbrace{((\mu - 1) + (\lambda - 1) + (\lambda - 1)\,(\mu - 1))}_{=\lambda\mu-1} E_{u,u} = I_n + (\lambda\mu - 1)\, E_{u,u}.
\end{aligned}$$

Comparing this with

$$S_u^{\lambda\mu} = I_n + (\lambda\mu - 1)\, E_{u,u} \qquad \left(\text{by the definition of } S_u^{\lambda\mu}\right),$$

we obtain $S_u^\lambda S_u^\mu = S_u^{\lambda\mu}$. This proves Proposition 3.92 **(b)** again.

(c) Notice that $1/\lambda$ is a well-defined number (since λ is nonzero). Proposition 3.92 **(b)** (applied to $\mu = 1/\lambda$) yields $S_u^\lambda S_u^{1/\lambda} = S_u^{\lambda(1/\lambda)} = S_u^1 = I_n$ (by Proposition 3.92 **(a)**).

But we can also apply Proposition 3.92 **(b)** to $1/\lambda$ and λ instead of λ and μ. We thus obtain $S_u^{1/\lambda} S_u^\lambda = S_u^{(1/\lambda)\lambda} = S_u^1 = I_n$ (by Proposition 3.92 **(a)**).

The two equalities $S_u^\lambda S_u^{1/\lambda} = I_n$ and $S_u^{1/\lambda} S_u^\lambda = I_n$ show that $S_u^{1/\lambda}$ is an inverse of S_u^λ. This proves Proposition 3.92 **(c)**. □

3.15. Invertibly triangular matrices are products of S_u^λ's and $A_{u,v}^\lambda$'s

In the same way as the matrices $A_{u,v}^\lambda$ served as "building blocks" for unitriangular matrices (in Theorem 3.63), we can obtain "building blocks" for invertibly triangular matrices if we accompany these matrices $A_{u,v}^\lambda$ by the matrices S_u^λ with nonzero λ. We shall state this soon in some precision; let us first introduce terminology:

Definition 3.93. Let $n \in \mathbb{N}$. A *scaling $n \times n$-matrix* means a matrix of the form S_u^λ, where λ is a **nonzero** number, and where u is an element of $\{1, 2, \ldots, n\}$. When n is clear from the context, we shall omit the "$n \times n$-" and simply say "scaling matrix".

The name "scaling matrix" is, again, not standard, but it will be useful for me in this chapter.

Example 3.94. If $n = 3$, then the scaling 3×3-matrices are the matrices of the form

$$S_1^\lambda = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad S_2^\lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
$$S_3^\lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

for all numbers λ.

It is clear that each scaling matrix is diagonal (and therefore lower-triangular and upper-triangular). Thus, every product of scaling matrices is a product of diagonal matrices, and thus itself must be diagonal[64]. Moreover, both scaling matrices and lower addition matrices[65] are invertibly lower-triangular[66]. Hence, every product of scaling matrices and lower addition matrices is a product of invertibly lower-triangular matrices, and thus itself must be invertibly lower-triangular[67]. It turns out that the converse is also true: Every invertibly lower-triangular matrix is a product of scaling matrices and lower addition matrices! This is again a simple particular case of Gaussian elimination, similar to Theorem 3.63; let me state it as a theorem:

Theorem 3.95. Let $n \in \mathbb{N}$. An $n \times n$-matrix C is invertibly lower-triangular if and only if C is a product of scaling matrices and lower addition matrices.

Example 3.96. (a) The invertibly lower-triangular 2×2-matrix $\begin{pmatrix} 2 & 0 \\ 5 & 3 \end{pmatrix}$ is a product of scaling matrices and lower addition matrices: Namely, it equals $S_1^2 S_2^3 A_{2,1}^{5/3}$.

[64] Here we are using the fact that any product of diagonal matrices is diagonal. This is not hard to check.

[65] Those were defined in Definition 3.61.

[66] Here, we are using the requirement that λ is nonzero in Definition 3.93. Without this requirement, S_u^λ would not be invertibly lower-triangular!

[67] because Corollary 3.40 shows that any product of invertibly lower-triangular matrices is invertibly lower-triangular

(b) The invertibly lower-triangular 1×1-matrix $\begin{pmatrix} 5 \end{pmatrix}$ is a product of scaling matrices and lower addition matrices: Namely, it is S_1^5.

(c) Let C be the invertibly lower-triangular 3×3-matrix $\begin{pmatrix} u & 0 & 0 \\ a & v & 0 \\ b & c & w \end{pmatrix}$. Then, C is a product of lower addition matrices: Namely, it equals $S_1^u S_2^v S_3^w A_{2,1}^{a/v} A_{3,1}^{b/w} A_{3,2}^{c/w}$.

Let us actually see how this representation of C can be found. We shall proceed by writing C as a product of one scaling matrix with a second matrix C', which is still invertibly lower-triangular but has one diagonal entry equal to 1. We then will do the same with C', obtaining a third matrix C''; then, do the same with C'', and so on. At the end, we will be left with an invertibly lower-triangular matrix whose **all** diagonal entries are 1. This means that we will be left with a lower-unitriangular matrix. But from Theorem 3.63, we already know that this latter matrix must be a product of lower addition matrices. In more detail: We first observe that the diagonal entries u, v, w of C are nonzero (since C is invertibly lower-triangular). Now, we proceed in several steps:

Step 1: Let us turn the $(1,1)$-st entry of C into 1 by scaling row 1 by $1/u$. Denote the resulting matrix by C'. Thus, $C' = \begin{pmatrix} 1 & 0 & 0 \\ a & v & 0 \\ b & c & w \end{pmatrix}$. Notice that the new matrix C' is still invertibly lower-triangular (since only row 1 has been changed, and since the scaling has left all the zero entries in place), but now has its first diagonal entry equal to 1. Since C' was obtained from C by scaling row 1 by $1/u$, we can conversely obtain C from C' by scaling row 1 by u. According to Proposition 3.88 (applied to n, 1, u and C' instead of m, u, λ and C), this means that $C = S_1^u C'$.

Step 2: Let us turn the $(2,2)$-st entry of C' into 1 by scaling row 2 by $1/v$. Denote the resulting matrix by C''. Thus, $C'' = \begin{pmatrix} 1 & 0 & 0 \\ a/v & 1 & 0 \\ b & c & w \end{pmatrix}$. Again, the new matrix C'' is still invertibly lower-triangular (since only row 2 has been changed, and all zeroes have survived the scaling), and moreover the first diagonal entry is still 1 (because only row 2 has been changed); but now the second diagonal entry is also 1. Similarly to how we found that $C = S_1^u C'$ in Step 1, we now obtain $C' = S_2^v C''$.

Step 3: Let us turn the $(3,3)$-st entry of C'' into 1 by scaling row 3 by $1/w$. Denote the resulting matrix by C'''. Thus, $C''' = \begin{pmatrix} 1 & 0 & 0 \\ a/v & 1 & 0 \\ b/w & c/w & 1 \end{pmatrix}$. Again, the new matrix C''' is still invertibly lower-triangular, and the first two diagonal entries are still equal to 1; and now the third diagonal entry has

become 1 as well. Similarly to how we found that $C = S_1^u C'$ in Step 1, we now obtain $C'' = S_3^w C'''$.

We have thus turned all diagonal entries into 1. Our final matrix C''' is thus upper-unitriangular. Thus, Theorem 3.63 shows that C''' is a product of lower addition matrices. Explicitly, it can be written as follows:

$$C''' = A_{2,1}^{a/v} A_{3,1}^{b/w} A_{3,2}^{c/w}.$$

(Indeed, this follows from (79), applied to a/v, b/w and c/w instead of a, b and c). Combining the three equalities we have found, we obtain

$$\begin{aligned} C &= S_1^u \underbrace{C'}_{=S_1^v C''} = S_1^u S_1^v \underbrace{C''}_{=S_1^w C'''} = S_1^u S_1^v S_1^w \underbrace{C'''}_{=A_{2,1}^{a/v} A_{3,1}^{b/w} A_{3,2}^{c/w}} \\ &= S_1^u S_1^v S_1^w A_{2,1}^{a/v} A_{3,1}^{b/w} A_{3,2}^{c/w}. \end{aligned}$$

Thus we have represented C as a product of scaling matrices and lower addition matrices.

The general proof of Theorem 3.95 follows the idea outlined in Example 3.96 **(c)**. The proof is so similar to the proof of Theorem 3.63 that I shall be copying the structure of the latter proof:

Proof of Theorem 3.95. $\Longleftarrow$: We have already proven that every product of scaling matrices and lower addition matrices is invertibly lower-triangular. Hence, if C is a product of scaling matrices and lower addition matrices, then C is invertibly lower-triangular. This proves the $\Longleftarrow$ direction of Theorem 3.95.

$\Longrightarrow$: We need to prove that if C is invertibly lower-triangular, then C is a product of scaling matrices and lower addition matrices.

So let us assume that C is invertibly lower-triangular. Our goal is to prove that C is a product of scaling matrices and lower addition matrices.

Let me introduce a notation first: A *row scaling* shall mean a transformation that changes an $n \times m$-matrix (for some $m \in \mathbb{N}$) by scaling one of its rows by some nonzero number. In more formal terms: A *row scaling* means a transformation of the form "scale the u-th row by λ", for some fixed nonzero number λ and some fixed $u \in \{1, 2, \ldots, n\}$. As we know from Proposition 3.88, this transformation amounts to multiplying a matrix by S_u^λ from the left (i.e., this transformation sends any $n \times m$-matrix B to $S_u^\lambda B$); we shall therefore denote this transformation itself by S_u^λ as well (hoping that the reader will not confuse the transformation with the matrix).

Here is an example (for $n = 4$): The row scaling S_3^λ is the transformation that changes a $4 \times m$-matrix by scaling the 3-rd row by λ. For example, it transforms

the 4×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \\ a''' & b''' \end{pmatrix}$ into $\begin{pmatrix} a & b \\ a' & b' \\ \lambda a'' & \lambda b'' \\ a''' & b''' \end{pmatrix}$.

Notice that any row scaling S_u^λ is invertible: Namely, it can be undone by the row scaling $S_u^{1/\lambda}$. [68]

Notice that, for any row scaling S_u^λ, the matrix S_u^λ is a row scaling matrix.

I claim that we can transform the invertibly lower-triangular $n \times n$-matrix C into a lower-unitriangular matrix by performing a sequence of row scalings. Namely, we should proceed by the following method:[69]

- At first, our matrix is

$$C = \begin{pmatrix} C_{1,1} & 0 & 0 & \cdots & 0 \\ C_{2,1} & C_{2,2} & 0 & \cdots & 0 \\ C_{3,1} & C_{3,2} & C_{3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & C_{n,n} \end{pmatrix}.$$

  Its diagonal entries $C_{1,1}, C_{2,2}, \ldots, C_{n,n}$ are nonzero (since C is invertibly lower-triangular).

- Now, we perform the row scaling S_1^λ (for an appropriate choice of λ, namely for $\lambda = 1/C_{1,1}$) to turn the $(1,1)$-th entry of the matrix into 1. This operation preserves the invertibly lower-triangular nature of the matrix (i.e., the matrix remains invertibly lower-triangular[70]). As the result, we have turned the $(1,1)$-th entry of the matrix into 1. In other words, our matrix now looks

[68] Here are two ways to prove this:

First proof: The row scaling S_u^λ transforms a matrix by scaling the u-th row by λ, i.e., by multiplying each entry of the u-th row by λ. The row scaling $S_u^{1/\lambda}$ transforms a matrix by scaling the u-th row by $1/\lambda$, i.e., by dividing each entry of the u-th row by λ. Hence, these two row scalings undo each other (i.e., if we perform one and then the other, then we arrive back at the matrix we have started with), because each entry of the u-th row is multiplied by λ by the former row scaling and divided by λ by the latter (whereas entries in other rows are left unchanged by both scalings). So we have shown that the row scaling S_u^λ can be undone by the row scaling $S_u^{1/\lambda}$. Qed.

Second proof: Proposition 3.92 **(c)** shows that the matrix $S_u^{1/\lambda}$ is the inverse of the matrix S_u^λ. Hence, multiplying a matrix by $S_u^{1/\lambda}$ undoes multiplying a matrix by S_u^λ. In other words, the row scaling $S_u^{1/\lambda}$ undoes the row scaling S_u^λ. Qed.

[69] See the three-step procedure in Example 3.96 **(c)** for an illustration of this method.

[70] In fact, this holds for any row scaling: If B is any invertibly lower-triangular matrix and S_u^μ is a row scaling, then the result of applying S_u^μ to B will still be invertibly lower-triangular. To prove this, just observe that all zero entries of B remain zero when the row scaling S_u^μ is applied (since S_u^μ merely scales a row), whereas all nonzero diagonal entries of B remain nonzero (since S_u^μ scales a row by the **nonzero** number μ).

as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_{2,1} & C_{2,2} & 0 & \cdots & 0 \\ C_{3,1} & C_{3,2} & C_{3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & C_{n,n} \end{pmatrix}.$$

- Next, we similarly perform the row scaling S_2^λ (for an appropriate choice of λ, namely for $\lambda = 1/C_{2,2}$) to turn the $(2,2)$-th entry of the matrix into 1. Again, this operation preserves the invertibly lower-triangular nature of the matrix (i.e., the matrix remains invertibly lower-triangular)[71]. Furthermore, the $(1,1)$-th entry of the matrix has not been changed by S_2^λ (since S_2^λ only changes the 2-nd row), and thus is still 1. As the result, we have turned the $(2,2)$-th entry of the matrix into 1. In other words, our matrix now looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_{2,1}/C_{2,2} & 1 & 0 & \cdots & 0 \\ C_{3,1} & C_{3,2} & C_{3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & C_{n,n} \end{pmatrix}.$$

- Next, we similarly perform the row scaling S_3^λ (for an appropriate choice of λ, namely for $\lambda = 1/C_{3,3}$) to turn the $(3,3)$-th entry of the matrix into 1. As the result, we have turned the $(3,3)$-th entry of the matrix into 1. In other words, our matrix now looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_{2,1}/C_{2,2} & 1 & 0 & \cdots & 0 \\ C_{3,1}/C_{3,3} & C_{3,2}/C_{3,3} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & C_{n,n} \end{pmatrix}.$$

- We continue this process, changing each diagonal entry of the matrix into 1 (one at a time). At the end, our matrix looks as follows:

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ C_{2,1}/C_{2,2} & 1 & 0 & \cdots & 0 \\ C_{3,1}/C_{3,3} & C_{3,2}/C_{3,3} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1}/C_{n,n} & C_{n,2}/C_{n,n} & C_{n,3}/C_{n,n} & \cdots & 1 \end{pmatrix}.$$

[71] **Warning:** Unlike in our proof of Theorem 3.63, this operation will (usually) change not only the $(2,2)$-th entry, but also the $(2,1)$-nd entry of our matrix. In our proof of Theorem 3.63, each of the operations that we performed changed only one entry of our matrix; but here in the proof of Theorem 3.95, this is not the case. Nevertheless, the proof works, because the exact values of the entries below the diagonal are not important.

This is a lower-unitriangular matrix. Denote it by D. Theorem 3.63 (applied to D instead of C) shows that D is lower-unitriangular if and only if D is a product of lower addition matrices. Hence, D is product of lower addition matrices (since D is lower-unitriangular).

Thus, we have found an algorithm to transform our matrix C into a lower-unitriangular matrix D by a sequence of row scalings. Therefore, we can conversely transform the matrix D into C by a sequence of row scalings[72]. Let us denote these row scalings (used to transform D into C) by $S_{u_1}^{\lambda_1}, S_{u_2}^{\lambda_2}, \ldots, S_{u_k}^{\lambda_k}$, **numbered backwards** (i.e., starting from the one used last). Since each row scaling S_u^λ amounts to multiplying a matrix by the matrix S_u^λ (that is, it sends any $n \times m$-matrix B to $S_u^\lambda B$), we thus conclude that

$$C = S_{u_1}^{\lambda_1} S_{u_2}^{\lambda_2} \cdots S_{u_k}^{\lambda_k} D. \tag{96}$$

But $S_{u_1}^{\lambda_1} S_{u_2}^{\lambda_2} \cdots S_{u_k}^{\lambda_k}$ is a product of scaling matrices (because each of the matrices $S_{u_1}^{\lambda_1}, S_{u_2}^{\lambda_2}, \ldots, S_{u_k}^{\lambda_k}$ is a scaling matrix[73]). Hence, (96) rewrites as follows:

$$C = \underbrace{S_{u_1}^{\lambda_1} S_{u_2}^{\lambda_2} \cdots S_{u_k}^{\lambda_k}}_{\substack{\text{this is a product of} \\ \text{scaling matrices}}} \underbrace{D}_{\substack{\text{this is a product of} \\ \text{lower addition matrices}}}.$$

Hence, C is a product of scaling matrices and lower addition matrices. This is precisely what we had to prove. This proves the $\Longrightarrow$ direction of Theorem 3.95. Hence, the proof of Theorem 3.95 is complete. □

Remark 3.97. Our proof of Theorem 3.95 (specifically, of its $\Longrightarrow$ direction) actually gives an explicit representation of an invertibly lower-triangular $n \times n$-matrix C as a product of scaling matrices and lower addition matrices. We leave the details to the reader.

We can use Theorem 3.95 to prove the following analogue of Theorem 3.67, in the same way as we used Theorem 3.63 to prove Theorem 3.67 itself:

Theorem 3.98. Let $n \in \mathbb{N}$. Let A be an invertibly lower-triangular $n \times n$-matrix. Then, A is invertible, and its inverse A^{-1} is again invertibly lower-triangular.

Proof of Theorem 3.98. Theorem 3.95 (applied to $C = A$) shows that A is invertibly lower-triangular if and only if A is a product of scaling matrices and lower addition matrices. Hence, A is product of scaling matrices and lower addition matrices (since A is invertibly lower-triangular). In other words, A has the form $A = A_1 A_2 \cdots A_k$ for some $k \in \mathbb{N}$ and some k matrices $A_1, A_2, \ldots, A_k$, where each of $A_1, A_2, \ldots, A_k$ is either a scaling matrix or a lower addition matrix. Consider these k and $A_1, A_2, \ldots, A_k$.

Observe the following fact:

[72] because (as we have shown) any row scaling S_u^λ is invertible, and can be undone by another row scaling

[73] Here we are using the fact that, for any row scaling S_u^λ, the matrix S_u^λ is a scaling matrix.

Fact 1: Let $i \in \{1,2,\ldots,k\}$. Then, the matrix A_i is invertible, and its inverse A_i^{-1} is either a scaling matrix or a lower addition matrix.

[*Proof of Fact 1:* We know that A_i is either a scaling matrix or a lower addition matrix (since each of $A_1, A_2, \ldots, A_k$ is either a scaling matrix or a lower addition matrix). In other words, we are in one of the following two cases:

Case 1: The matrix A_i is a scaling matrix.

Case 2: The matrix A_i is a lower addition matrix.

Let us consider Case 1 first. In this case, A_i is a scaling matrix. In other words, A_i has the form $A_i = S_u^\lambda$ for some nonzero number λ and some $u \in \{1,2,\ldots,n\}$. Consider these λ and u. Proposition 3.92 **(c)** shows that the matrix S_u^λ is invertible, and its inverse is $\left(S_u^\lambda\right)^{-1} = S_u^{1/\lambda}$. Since $S_u^\lambda = A_i$, this rewrites as follows: The matrix A_i is invertible, and its inverse is $A_i^{-1} = S_u^{1/\lambda}$. But $S_u^{1/\lambda}$ is a scaling matrix. In other words, A_i^{-1} is a scaling matrix (since $A_i^{-1} = S_u^{1/\lambda}$). Hence, A_i^{-1} is either a scaling matrix or a lower addition matrix. Thus, Fact 1 is proven in Case 1 (since we have already shown that A_i is invertible).

Let us now consider Case 2. In this case, A_i is a lower addition matrix. In other words, A_i has the form $A_i = A_{u,v}^\lambda$, where λ is a number, and where u and v are two elements of $\{1,2,\ldots,n\}$ satisfying $u > v$ (by the definition of a "lower addition matrix"). Consider these λ, u and v. Proposition 3.60 **(c)** shows that the matrix $A_{u,v}^\lambda$ is invertible, and its inverse is $\left(A_{u,v}^\lambda\right)^{-1} = A_{u,v}^{-\lambda}$. Since $A_{u,v}^\lambda = A_i$, this rewrites as follows: The matrix A_i is invertible, and its inverse is $A_i^{-1} = A_{u,v}^{-\lambda}$. But $A_{u,v}^{-\lambda}$ is a lower addition matrix (since $u > v$). In other words, A_i^{-1} is a lower addition matrix (since $A_i^{-1} = A_{u,v}^{-\lambda}$). Hence, A_i^{-1} is either a scaling matrix or a lower addition matrix. Thus, Fact 1 is proven in Case 2 (since we have already shown that A_i is invertible).

We have now proven Fact 1 in each of the two Cases 1 and 2. Thus, Fact 1 is proven.]

We are in one of the following two cases:

Case 1: We have $k \neq 0$.

Case 2: We have $k = 0$.

Let us deal with Case 1. In this case, we have $k \neq 0$; thus, k is a positive integer.

Fact 1 shows that, for each $i \in \{1,2,\ldots,k\}$, the matrix A_i is invertible, and its inverse A_i^{-1} is either a scaling matrix or a lower addition matrix. In other words, the matrices $A_1, A_2, \ldots, A_k$ are invertible, and each of their inverses $A_1^{-1}, A_2^{-1}, \ldots, A_k^{-1}$ is either a scaling matrix or a lower addition matrix. In other words, each of $A_k^{-1}, A_{k-1}^{-1}, \ldots, A_1^{-1}$ is either a scaling matrix or a lower addition matrix.

Now, Proposition 3.12 shows that the matrix $A_1A_2\cdots A_k$ is invertible, and its inverse is $(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}$. Since $A = A_1A_2\cdots A_k$, this rewrites as follows: The matrix A is invertible, and its inverse is $A^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}$.

The equality $A^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}$ shows that A^{-1} is a product of scaling matrices and lower addition matrices (since each of $A_k^{-1}, A_{k-1}^{-1}, \ldots, A_1^{-1}$ is either a scaling matrix or a lower addition matrix). But Theorem 3.95 (applied to $C = A^{-1}$) shows that A^{-1} is invertibly lower-triangular if and only if A^{-1} is a product of scaling matrices and lower addition matrices. Hence, A^{-1} is invertibly lower-triangular (since A^{-1} is a product of scaling matrices and lower addition matrices). This completes the proof of Theorem 3.98 in Case 1.

Case 2 is trivial (indeed, $A = I_n$ in this case) and is left to the reader.[74] Thus, Theorem 3.98 is proven in both Cases 1 and 2; this shows that Theorem 3.98 is always valid. □

Again, a similar result holds for upper-triangular matrices (and, again, has a similar proof):

Theorem 3.99. Let $n \in \mathbb{N}$. Let A be an invertibly upper-triangular $n \times n$-matrix. Then, A is invertible, and its inverse A^{-1} is again invertibly upper-triangular.

3.16. (*) Yet another proof of triangular invertibility

Let us now give a second proof of Theorem 3.98. We first shall show a lemma:

Lemma 3.100. Let $n \in \mathbb{N}$. Let A be an invertibly lower-triangular $n \times n$-matrix. Let $b = (b_1, b_2, \ldots, b_n)^T$ be a column vector of size n (that is, an $n \times 1$-matrix). Let $r \in \{1, 2, \ldots, n\}$ be such that $b_1 = b_2 = \cdots = b_{r-1} = 0$. (Notice that if $r = 1$, then the equality $b_1 = b_2 = \cdots = b_{r-1} = 0$ claims nothing, and thus is automatically true – i.e., the numbers $b_1, b_2, \ldots, b_n$ can be arbitrary in this case.)

Then, there exists a column vector $v = (v_1, v_2, \ldots, v_n)^T$ of size n satisfying $Av = b$ and $v_1 = v_2 = \cdots = v_{r-1} = 0$ and $v_r = \dfrac{1}{A_{r,r}} b_r$.

Proof of Lemma 3.100. The matrix A is invertibly lower-triangular. In other words, A is lower-triangular and all its diagonal entries are nonzero (because this is what "invertibly lower-triangular" means). Since A is lower-triangular, we have

$$A_{i,j} = 0 \qquad \text{whenever } i < j \tag{97}$$

(by the definition of "lower-triangular"). Since all diagonal entries of A are nonzero, we have

$$A_{i,i} \neq 0 \qquad \text{for each } i \in \{1, 2, \ldots, n\}. \tag{98}$$

[74] Alternatively, our proof for Case 1 can be made to work in Case 2 as well, because Proposition 3.12 holds for $k = 0$ (as long as we define the empty product to be I_n). See Remark 3.13 for the details.

We shall now define n numbers $v_1, v_2, \ldots, v_n$. Our definition is recursive: For each $i \in \{1, 2, \ldots, n\}$, we assume that the first $i-1$ numbers $v_1, v_2, \ldots, v_{i-1}$ are already defined, and we define the next number v_i by

$$v_i = \frac{1}{A_{i,i}} \left(b_i - \left(A_{i,1} v_1 + A_{i,2} v_2 + \cdots + A_{i,i-1} v_{i-1} \right) \right) \tag{99}$$

[75]. This is a valid recursive definition, because it gives us a way to compute the numbers $v_1, v_2, \ldots, v_n$ one by one (beginning with v_1, then proceeding to v_2, then to v_3, and so on). Here is how this computation will look like:

- The number v_1 is defined (and can be computed) by $v_1 = \dfrac{1}{A_{1,1}} b_1$. (This is the particular case of (99) for $i = 1$. [76])
- The number v_2 is defined (and can be computed) by $v_2 = \dfrac{1}{A_{2,2}} \left(b_2 - A_{2,1} v_1 \right)$.
- The number v_3 is defined (and can be computed) by $v_3 = \dfrac{1}{A_{3,3}} \left(b_3 - \left(A_{3,1} v_1 + A_{3,2} v_2 \right) \right)$.
- And so on, up to v_n.

We furthermore define a column vector v by $v = \left(v_1, v_2, \ldots, v_n \right)^T$. Now, we claim the following:

Claim 1: We have $Av = b$.

Claim 2: We have $v_1 = v_2 = \cdots = v_{r-1} = 0$.

Claim 3: We have $v_r = \dfrac{1}{A_{r,r}} b_r$.

[*Proof of Claim 1:* Multiplying the equalities $A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{pmatrix}$ and

[75] The division by $A_{i,i}$ is allowed because of (98).

[76] Note that the sum $A_{i,1} v_1 + A_{i,2} v_2 + \cdots + A_{i,i-1} v_{i-1}$ becomes an empty sum when $i = 1$, and thus equals 0; therefore, we have omitted it.

$v = (v_1, v_2, \ldots, v_n)^T = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$, we obtain

$$Av = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$
$$= \begin{pmatrix} A_{1,1}v_1 + A_{1,2}v_2 + \cdots + A_{1,n}v_n \\ A_{2,1}v_1 + A_{2,2}v_2 + \cdots + A_{2,n}v_n \\ \vdots \\ A_{n,1}v_1 + A_{n,2}v_2 + \cdots + A_{n,n}v_n \end{pmatrix}. \tag{100}$$

But for each $i \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned}
&A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,n}v_n \\
&= (A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i}v_i) \\
&\quad + \begin{pmatrix} \underbrace{A_{i,i+1}}_{\substack{=0 \\ \text{(by (97),} \\ \text{applied to } j=i+1\text{)}}} v_{i+1} + \underbrace{A_{i,i+2}}_{\substack{=0 \\ \text{(by (97),} \\ \text{applied to } j=i+2\text{)}}} v_{i+2} + \cdots + \underbrace{A_{i,n}}_{\substack{=0 \\ \text{(by (97),} \\ \text{applied to } j=n\text{)}}} v_n \end{pmatrix} \\
&= (A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i}v_i) + \underbrace{(0v_{i+1} + 0v_{i+2} + \cdots + 0v_n)}_{=0} \\
&= A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i}v_i \\
&= (A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i-1}v_{i-1}) + \underbrace{A_{i,i}v_i}_{\substack{= b_i - \left(A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i-1}v_{i-1}\right) \\ \text{(by (99))}}} \\
&= (A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i-1}v_{i-1}) + b_i - (A_{i,1}v_1 + A_{i,2}v_2 + \cdots + A_{i,i-1}v_{i-1}) \\
&= b_i.
\end{aligned}$$

In other words, the n equalities

$$\begin{aligned}
A_{1,1}v_1 + A_{1,2}v_2 + \cdots + A_{1,n}v_n &= b_1; \\
A_{2,1}v_1 + A_{2,2}v_2 + \cdots + A_{2,n}v_n &= b_2; \\
&\vdots \\
A_{n,1}v_1 + A_{n,2}v_2 + \cdots + A_{n,n}v_n &= b_n
\end{aligned}$$

hold. In other words,

$$\begin{pmatrix} A_{1,1}v_1 + A_{1,2}v_2 + \cdots + A_{1,n}v_n \\ A_{2,1}v_1 + A_{2,2}v_2 + \cdots + A_{2,n}v_n \\ \vdots \\ A_{n,1}v_1 + A_{n,2}v_2 + \cdots + A_{n,n}v_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Thus, (100) becomes

$$Av = \begin{pmatrix} A_{1,1}v_1 + A_{1,2}v_2 + \cdots + A_{1,n}v_n \\ A_{2,1}v_1 + A_{2,2}v_2 + \cdots + A_{2,n}v_n \\ \vdots \\ A_{n,1}v_1 + A_{n,2}v_2 + \cdots + A_{n,n}v_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = b.$$

This proves Claim 1.]

[*Proof of Claim 2:* We shall show that

$$v_1 = v_2 = \cdots = v_{k-1} = 0 \qquad \text{for each } k \in \{1, 2, \ldots, r\}. \tag{101}$$

Our proof of (101) will proceed by induction over k:

Induction base: For $k = 1$, the equality (101) claims nothing (because there are no numbers $v_1, v_2, \ldots, v_{k-1}$ when $k = 1$), and thus is true (for the stupid reason that an empty claim is always true). This completes the induction base.

Induction step: Let $i \in \{1, 2, \ldots, r-1\}$. Assume that (101) holds for $k = i$. We must now prove that (101) holds for $k = i + 1$.

We have $i \in \{1, 2, \ldots, r-1\}$ and thus $b_i = 0$ (since $b_1 = b_2 = \cdots = b_{r-1} = 0$).

We have assumed that (101) holds for $k = i$. In other words, we have $v_1 = v_2 = \cdots = v_{i-1} = 0$. In other words,

$$v_\ell = 0 \qquad \text{for each } \ell \in \{1, 2, \ldots, i-1\}. \tag{102}$$

But now, (99) yields

$$\begin{aligned} v_i &= \frac{1}{A_{i,i}} \left(b_i - \left(A_{i,1} \underbrace{v_1}_{\substack{=0 \\ \text{(by (102), applied} \\ \text{to } \ell=1\text{)}}} + A_{i,2} \underbrace{v_2}_{\substack{=0 \\ \text{(by (102), applied} \\ \text{to } \ell=2\text{)}}} + \cdots + A_{i,i-1} \underbrace{v_{i-1}}_{\substack{=0 \\ \text{(by (102), applied} \\ \text{to } \ell=i-1\text{)}}} \right) \right) \\ &= \frac{1}{A_{i,i}} \left(b_i - \underbrace{\left(A_{i,1}0 + A_{i,2}0 + \cdots + A_{i,i-1}0 \right)}_{=0} \right) = \frac{1}{A_{i,i}} \underbrace{b_i}_{=0} = 0. \end{aligned}$$

Combining this with $v_1 = v_2 = \cdots = v_{i-1} = 0$, we obtain $v_1 = v_2 = \cdots = v_i = 0$. In other words, (101) holds for $k = i + 1$. This completes the induction step. Hence, (101) is proven.

Now that we have proven (101), we can apply (101) to $k = r$. We thus obtain $v_1 = v_2 = \cdots = v_{r-1} = 0$. This proves Claim 2.]

[*Proof of Claim 3:* Claim 2 shows that $v_1 = v_2 = \cdots = v_{r-1} = 0$. In other words,

$$v_\ell = 0 \qquad \text{for each } \ell \in \{1, 2, \ldots, r-1\}. \tag{103}$$

But (99) (applied to $i = r$) yields

$$\begin{aligned} v_r &= \frac{1}{A_{r,r}} \left(b_r - \left(A_{r,1} \underbrace{v_1}_{\substack{=0 \\ \text{(by (103), applied} \\ \text{to } \ell=1)}} + A_{r,2} \underbrace{v_2}_{\substack{=0 \\ \text{(by (103), applied} \\ \text{to } \ell=2)}} + \cdots + A_{r,r-1} \underbrace{v_{r-1}}_{\substack{=0 \\ \text{(by (103), applied} \\ \text{to } \ell=r-1)}} \right) \right) \\ &= \frac{1}{A_{r,r}} \left(b_r - \underbrace{\left(A_{r,1}0 + A_{r,2}0 + \cdots + A_{r,r-1}0 \right)}_{=0} \right) = \frac{1}{A_{r,r}} b_r. \end{aligned}$$

This proves Claim 3.]

We have now proven all three claims. Thus, our column vector $v = (v_1, v_2, \ldots, v_n)^T$ satisfies $Av = b$ and $v_1 = v_2 = \cdots = v_{r-1} = 0$ and $v_r = \dfrac{1}{A_{r,r}} b_r$. This shows that such a vector exists; in other words, Lemma 3.100 is proven. □

Next, we prove a lemma which is "almost" Theorem 3.98:

Lemma 3.101. Let $n \in \mathbb{N}$. Let A be an invertibly lower-triangular $n \times n$-matrix. Then, there exists an invertibly lower-triangular $n \times n$-matrix B such that $AB = I_n$.

The matrix B in Lemma 3.101 will turn out to be the inverse of A; but Lemma 3.101 does not yet claim this (instead, Lemma 3.101 only guarantees that B is a right inverse of A).

Proof of Lemma 3.101. For each $j \in \{1, 2, \ldots, n\}$, we construct a column vector $v_{[j]}$ of size n as follows:

Consider the vector $\operatorname{col}_j (I_n)$; this is the j-th column of the identity matrix I_n. Since $I_n = \left(\delta_{i,j}\right)_{1 \le i \le n,\ 1 \le j \le n}$, we have

$$\operatorname{col}_j (I_n) = \begin{pmatrix} \delta_{1,j} \\ \delta_{2,j} \\ \vdots \\ \delta_{n,j} \end{pmatrix} = \left(\delta_{1,j}, \delta_{2,j}, \ldots, \delta_{n,j}\right)^T.$$

This vector $\operatorname{col}_j (I_n)$ is a column vector of size n, and satisfies $\delta_{1,j} = \delta_{2,j} = \cdots = \delta_{j-1,j} = 0$ (since none of the numbers $1, 2, \ldots, j-1$ equals j). Hence, Lemma 3.100

(applied to $b = \operatorname{col}_j(I_n)$ and $b_i = \delta_{i,j}$ and $r = j$) says that there exists a column vector $v = (v_1, v_2, \ldots, v_n)^T$ of size n satisfying $Av = \operatorname{col}_j(I_n)$ and $v_1 = v_2 = \cdots = v_{j-1} = 0$ and $v_j = \dfrac{1}{A_{j,j}}\delta_{j,j}$. Denote this vector v by $v_{[j]}$, and denote its entries $v_1, v_2, \ldots, v_n$ by $v_{1,j}, v_{2,j}, \ldots, v_{n,j}$ (in order to stress that they depend on j).

Now, forget that we fixed j. Thus, for each $j \in \{1, 2, \ldots, n\}$, we have found a column vector

$$v_{[j]} = \left(v_{1,j}, v_{2,j}, \ldots, v_{n,j}\right)^T \tag{104}$$

of size n satisfying

$$Av_{[j]} = \operatorname{col}_j(I_n) \tag{105}$$

and

$$v_{1,j} = v_{2,j} = \cdots = v_{j-1,j} = 0 \tag{106}$$

and

$$v_{j,j} = \frac{1}{A_{j,j}}\delta_{j,j}. \tag{107}$$

Altogether, these are n column vectors $v_{[1]}, v_{[2]}, \ldots, v_{[n]}$ of size n. We can assemble them into a matrix B: Namely, set

$$B = \left(v_{i,j}\right)_{1 \le i \le n,\ 1 \le j \le n}.$$

This matrix B therefore satisfies

$$B_{i,j} = v_{i,j} \qquad \text{for all } i \in \{1, 2, \ldots, n\} \text{ and } j \in \{1, 2, \ldots, n\}. \tag{108}$$

Hence, the matrix B is lower-triangular[77] and all its diagonal entries are nonzero[78]. In other words, the matrix B is invertibly lower-triangular.

Also, recall that $B = \left(v_{i,j}\right)_{1 \le i \le n,\ 1 \le j \le n}$. Hence, every $j \in \{1, 2, \ldots, n\}$ satisfies

$$\operatorname{col}_j B = \begin{pmatrix} v_{1,j} \\ v_{2,j} \\ \vdots \\ v_{n,j} \end{pmatrix} = \left(v_{1,j}, v_{2,j}, \ldots, v_{n,j}\right)^T = v_{[j]} \qquad (\text{by (104)}). \tag{109}$$

[77] *Proof.* Let $j \in \{1, 2, \ldots, n\}$. Then, $v_{1,j} = v_{2,j} = \cdots = v_{j-1,j} = 0$ (by (106)). In other words, $v_{i,j} = 0$ for every $i \in \{1, 2, \ldots, n\}$ satisfying $i < j$. In light of (108), this rewrites as $B_{i,j} = 0$ for every $i \in \{1, 2, \ldots, n\}$ satisfying $i < j$.

Now, forget that we fixed j. We thus have shown that $B_{i,j} = 0$ whenever $i < j$. In other words, the matrix B is lower-triangular.

[78] *Proof.* Let $i \in \{1, 2, \ldots, n\}$. Applying (107) to $j = i$, we find $v_{i,i} = \dfrac{1}{A_{i,i}} \underbrace{\delta_{i,i}}_{\substack{=1 \\ (\text{since } i=i)}} = \dfrac{1}{A_{i,i}} \neq 0$. But (108) (applied to $j = i$) yields $B_{i,i} = v_{i,i} \neq 0$.

Now, forget that we fixed i. We thus have learnt that $B_{i,i} \neq 0$ for each $i \in \{1, 2, \ldots, n\}$. In other words, the diagonal entries of the matrix B are nonzero.

Now, Proposition 2.19 **(d)** (applied to $m = n$ and $p = n$) shows that, for each $j \in \{1, 2, \ldots, n\}$, we have

$$\operatorname{col}_j (AB) = A \cdot \underbrace{\operatorname{col}_j B}_{\substack{= v_{[j]} \\ \text{(by (109))}}} = A v_{[j]} = \operatorname{col}_j (I_n) \qquad \text{(by (105))}.$$

In other words, each column of the $n \times n$-matrix AB equals the corresponding column of I_n. Thus, $AB = I_n$.

Hence, we have found an invertibly lower-triangular $n \times n$-matrix B such that $AB = I_n$. This proves Lemma 3.101. □

We are now ready to prove Theorem 3.98 by a rather cunning trick:

Proof of Theorem 3.98. Lemma 3.101 shows that there exists an invertibly lower-triangular $n \times n$-matrix B such that $AB = I_n$. Fix such a B, and denote it by C. Thus, C is an invertibly lower-triangular $n \times n$-matrix such that $AC = I_n$.

But we can also apply Lemma 3.101 to C instead of B. As the result, we conclude that there exists an invertibly lower-triangular $n \times n$-matrix B such that $CB = I_n$. Fix such a B, and denote it by D. Thus, D is an invertibly lower-triangular $n \times n$-matrix such that $CD = I_n$.

Now, the matrix A is a left inverse of C (since $AC = I_n$), whereas the matrix D is a right inverse of C (since $CD = I_n$). Hence, Proposition 3.6 **(a)** (applied to n, C, A and D instead of m, A, L and R) reveals that $A = D$. Hence, $C \underbrace{A}_{=D} = CD = I_n$. Combining $CA = I_n$ with $AC = I_n$, we conclude that C is an inverse of A. Thus, the matrix A is invertible, and its inverse is $A^{-1} = C$. Thus, A^{-1} is invertibly lower-triangular (since C is invertibly lower-triangular, but $A^{-1} = C$). This proves Theorem 3.98. □

We have thus proven Theorem 3.98 again.

We could similarly prove Theorem 3.99, though this would require an analogue of Lemma 3.100 in which (for example) the condition $b_1 = b_2 = \cdots = b_{r-1} = 0$ would be replaced by $b_{r+1} = b_{r+2} = \cdots = b_n = 0$ (and in the proof, we would have to define the $v_1, v_2, \ldots, v_n$ by "reverse recursion", beginning with v_n and then proceeding with v_{n-1} and so on). It is probably quicker to derive Theorem 3.99 from Theorem 3.98 in the same way as we derived Theorem 3.69 from Theorem 3.67. Either way, the proof is straightforward (given what has already been shown), and is left to the reader.

Theorem 3.67 and Theorem 3.69 can also be proven in the same way; the changes usually boil down to replacing "$\neq 0$" by "$= 1$". Again, this proof can safely be left to the reader (who would thus obtain a third proof of Theorem 3.67!).

3.17. The swapping matrices $T_{u,v}$

The λ-addition matrices $A^{\lambda}_{u,v}$ from Definition 3.53 and the λ-scaling matrices S^{λ}_{u} from Definition 3.85 are two of the three kinds of matrices commonly called "elementary matrices". The third kind are the *swapping matrices*:

Definition 3.102. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Then, $T_{u,v}$ shall denote the $n \times n$-matrix $I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$. (Here, all the matrices of the form $E_{p,q}$ are meant to be $n \times n$-matrices: that is, $E_{p,q} = E_{p,q,n,n}$ for all $p \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, n\}$.)

Again, the notation $T_{u,v}$ is hiding the dependency on n, and we ought to write $T_{u,v,n}$ instead; but we will not, because there will not be any real occasion for confusion.

Example 3.103. Let $n = 4$. Then,

$$T_{1,3} = I_n - E_{1,1} - E_{3,3} + E_{1,3} + E_{3,1} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and}$$

$$T_{2,3} = I_n - E_{2,2} - E_{3,3} + E_{2,3} + E_{3,2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The pattern that you see on these examples is true in general:

Proposition 3.104. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Then, the matrix $T_{u,v}$ has the following entries:

- Its (u, u)-th and (v, v)-th entries are 0.
- All its other diagonal entries are 1.
- Its (u, v)-th and (v, u)-th entries are 1.
- All its remaining entries are 0.

Proof of Proposition 3.104. For each $p \in \{1, 2, \ldots, n\}$ and $q \in \{1, 2, \ldots, n\}$, the matrix $E_{p,q}$ is the $n \times n$-matrix whose (p, q)-th entry is 1 and whose all other entries are 0 (indeed, this is how $E_{p,q}$ was defined). Hence, we obtain the following two facts:

> *Fact 1:* Adding $E_{p,q}$ to an $n \times n$-matrix C has the effect that the (p, q)-th entry of C is increased by 1 (while all other entries remain unchanged).

Fact 2: Subtracting $E_{p,q}$ from an $n \times n$-matrix C has the effect that the (p,q)-th entry of C is decreased by 1 (while all other entries remain unchanged).

But recall that $T_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$. Thus, the matrix $T_{u,v}$ is obtained from the matrix I_n by first subtracting $E_{u,u}$, then subtracting $E_{v,v}$, then adding $E_{u,v}$, and then adding $E_{v,u}$. Using Fact 1 and Fact 2, we can see how these subtractions and additions affect the entries of a matrix; thus, we can find all entries of the matrix $T_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$:

- The matrix I_n is the $n \times n$-matrix whose diagonal entries[79] are 1, and whose all other entries are 0.
- Subtracting $E_{u,u}$ from I_n has the effect that the (u,u)-th entry is decreased by 1; thus, it becomes $1 - 1 = 0$ (because it was 1 in I_n). Hence, the (u,u)-th entry of the matrix $I_n - E_{u,u}$ is 0, all its other diagonal entries are 1, and all its remaining entries are 0.
- Subtracting $E_{v,v}$ from $I_n - E_{u,u}$ has the effect that the (v,v)-th entry is decreased by 1; thus, it becomes $1 - 1 = 0$ (because it was 1 in $I_n - E_{u,u}$). Hence, the (u,u)-th and (v,v)-th entries of the matrix $I_n - E_{u,u} - E_{v,v}$ are 0, all its other diagonal entries are 1, and all its remaining entries are 0.
- Adding $E_{u,v}$ to $I_n - E_{u,u} - E_{v,v}$ has the effect that the (u,v)-th entry is increased by 1; thus, it becomes $0 + 1 = 1$ (because it was 0 in $I_n - E_{u,u} - E_{v,v}$). Hence, the (u,u)-th and (v,v)-th entries of the matrix $I_n - E_{u,u} - E_{v,v} + E_{u,v}$ are 0, all its other diagonal entries are 1, its (u,v)-th entry is 1, and all its remaining entries are 0.
- Adding $E_{v,u}$ to $I_n - E_{u,u} - E_{v,v} + E_{u,v}$ has the effect that the (v,u)-th entry is increased by 1; thus, it becomes $0 + 1 = 1$ (because it was 0 in $I_n - E_{u,u} - E_{v,v} + E_{u,v}$). Hence, the (u,u)-th and (v,v)-th entries of the matrix $I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$ are 0, all its other diagonal entries are 1, its (u,v)-th and (v,u)-th entries are 1, and all its remaining entries are 0. Since $I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u} = T_{u,v}$, this rewrites as follows: The (u,u)-th and (v,v)-th entries of the matrix $T_{u,v}$ are 0, all its other diagonal entries are 1, its (u,v)-th and (v,u)-th entries are 1, and all its remaining entries are 0. This proves Proposition 3.104.

□

We can next see what happens to a matrix when it is multiplied by $T_{u,v}$:

[79] This includes the (u,u)-th entry.

Proposition 3.105. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let C be an $n \times m$-matrix. Then, $T_{u,v}C$ is the $n \times m$-matrix obtained from C by swapping the u-th row with the v-th row.

Example 3.106. Let $n = 3$ and $m = 2$. Let C be the 3×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix}$. Then, Proposition 3.105 (applied to $u = 1$ and $v = 3$) claims that $T_{1,3}C$ is the 3×2-matrix obtained from C by swapping the 1-st row with the 3-rd row. A computation confirms this claim:

$$T_{1,3}C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix} = \begin{pmatrix} a'' & b'' \\ a' & b' \\ a & b \end{pmatrix}.$$

Proposition 3.107. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let C be an $m \times n$-matrix. Then, $CT_{u,v}$ is the $m \times n$-matrix obtained from C by swapping the u-th column with the v-th column.

Corollary 3.108. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Then:

(a) The matrix $T_{u,v}$ can be obtained from I_n by swapping the u-th row with the v-th row.

(b) The matrix $T_{u,v}$ can be obtained from I_n by swapping the u-th column with the v-th column.

I will prove Proposition 3.105 and Corollary 3.108 in Section 3.18 below.

We shall refer to the matrix $T_{u,v}$ defined in Definition 3.102 as a "swapping matrix".[80] It is the third type of elementary matrices.

Here are a few more properties of swapping matrices:

Proposition 3.109. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Then, $(T_{u,v})^T = T_{u,v}$.

Proposition 3.110. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$.

(a) We have $T_{u,v} = I_n$.

(b) The matrix $T_{u,v}$ is invertible, and its inverse is $(T_{u,v})^{-1} = T_{u,v}$.

(c) We have $T_{v,u} = T_{u,v}$.

These facts will be proven in Section 3.18.

[80] The letter T in "$T_{u,v}$" stands for "transposition", but this is not really related to the transpose of a matrix. It is instead due to the fact that the word "transposition" means "changing the positions (of something)", and in mathematics is often used for swapping two things.

3.18. (*) Some proofs about the swapping matrices

Proof of Proposition 3.105. Clearly, $T_{u,v}C$ is an $n \times m$-matrix.

Let x and y be two elements of $\{1,2,\ldots,n\}$. Proposition 3.44 (applied to n, m, x and y instead of m, p, u and v) shows that $E_{x,y}C$ is the $n \times m$-matrix whose x-th row is the y-th row of C, and whose all other rows are filled with zeroes. Thus,

$$\operatorname{row}_x \left(E_{x,y}C\right) = \operatorname{row}_y C \tag{110}$$

(since the x-the row of $E_{x,y}C$ is the y-th row of C) and

$$\operatorname{row}_i \left(E_{x,y}C\right) = 0_{1\times m} \qquad \text{for every } i \in \{1,2,\ldots,n\} \text{ satisfying } i \neq x \tag{111}$$

(since all other rows of $E_{x,y}C$ are filled with zeroes).

Now, forget that we fixed x and y. We thus have proven (110) and (111) for every two elements x and y of $\{1,2,\ldots,n\}$. In particular, every $y \in \{1,2,\ldots,n\}$ satisfies

$$\operatorname{row}_u \left(E_{v,y}C\right) = 0_{1\times m} \tag{112}$$

(by (111), applied to $i = u$ and $x = v$ (since $u \neq v$)) and

$$\operatorname{row}_v \left(E_{u,y}C\right) = 0_{1\times m} \tag{113}$$

(by (111), applied to $i = v$ and $x = u$ (since $v \neq u$)).

But recall that matrices are added entry by entry. Thus, matrices are also added by row by row – i.e., if U and V are two $n \times m$-matrices, then any row of $U + V$ is the sum of the corresponding rows of U and of V. In other words, if U and V are two $n \times m$-matrices, then

$$\operatorname{row}_i (U+V) = \operatorname{row}_i U + \operatorname{row}_i V \qquad \text{for every } i \in \{1,2,\ldots,n\}. \tag{114}$$

Similarly,

$$\operatorname{row}_i (U-V) = \operatorname{row}_i U - \operatorname{row}_i V \qquad \text{for every } i \in \{1,2,\ldots,n\}. \tag{115}$$

Using (114) and (115) repeatedly, we can show that

$$\operatorname{row}_i (U-V-W+X+Y) = \operatorname{row}_i U - \operatorname{row}_i V - \operatorname{row}_i W + \operatorname{row}_i X + \operatorname{row}_i Y \tag{116}$$

for every $i \in \{1,2,\ldots,n\}$.

We have

$$\begin{aligned} \underbrace{T_{u,v}}_{=I_n-E_{u,u}-E_{v,v}+E_{u,v}+E_{v,u}} C &= \left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right) C \\ &= \underbrace{I_n C}_{=C} - E_{u,u}C - E_{v,v}C + E_{u,v}C + E_{v,u}C \\ &= C - E_{u,u}C - E_{v,v}C + E_{u,v}C + E_{v,u}C. \end{aligned}$$

Hence, for each $i \in \{1,2,\ldots,n\}$, we have

$$\begin{aligned}
&\operatorname{row}_i \left(\underbrace{T_{u,v}C}_{=C-E_{u,u}C-E_{v,v}C+E_{u,v}C+E_{v,u}C} \right) \\
&= \operatorname{row}_i \left(C - E_{u,u}C - E_{v,v}C + E_{u,v}C + E_{v,u}C \right) \\
&= \operatorname{row}_i C - \operatorname{row}_i \left(E_{u,u}C \right) - \operatorname{row}_i \left(E_{v,v}C \right) + \operatorname{row}_i \left(E_{u,v}C \right) + \operatorname{row}_i \left(E_{v,u}C \right) \qquad (117)
\end{aligned}$$

(by (116), applied to $U = C$, $V = E_{u,u}C$, $W = E_{v,v}C$, $X = E_{u,v}C$ and $Y = E_{v,u}C$).

Now, we must prove that $T_{u,v}C$ is the $n \times m$-matrix obtained from C by swapping the u-th row with the v-th row. In other words, we must prove the following three claims:

Claim 1: The u-th row of the $n \times m$-matrix $T_{u,v}C$ equals the v-th row of C.

Claim 2: The v-th row of the $n \times m$-matrix $T_{u,v}C$ equals the u-th row of C.

Claim 3: For each $i \in \{1,2,\ldots,n\}$ satisfying $i \neq u$ and $i \neq v$, the i-th row of the $n \times m$-matrix $T_{u,v}C$ equals the i-th row of C.

Proof of Claim 1: The u-th row of the $n \times m$-matrix $T_{u,v}C$ is

$$\begin{aligned}
&\operatorname{row}_u \left(T_{u,v}C \right) \\
&= \operatorname{row}_u C - \underbrace{\operatorname{row}_u \left(E_{u,u}C \right)}_{\substack{=\operatorname{row}_u C \\ \text{(by (110), applied to} \\ x=u \text{ and } y=u)}} - \underbrace{\operatorname{row}_u \left(E_{v,v}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (112),} \\ \text{applied to } y=v)}} + \underbrace{\operatorname{row}_u \left(E_{u,v}C \right)}_{\substack{=\operatorname{row}_v C \\ \text{(by (110), applied to} \\ x=u \text{ and } y=v)}} + \underbrace{\operatorname{row}_u \left(E_{v,u}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (112),} \\ \text{applied to } y=u)}} \\
&\qquad \text{(by (117), applied to } i = u) \\
&= \operatorname{row}_u C - \operatorname{row}_u C - 0_{1\times m} + \operatorname{row}_v C + 0_{1\times m} = \operatorname{row}_v C.
\end{aligned}$$

In other words, the u-th row of the $n \times m$-matrix $T_{u,v}C$ equals the v-th row of C. This proves Claim 1.

Proof of Claim 2: The v-th row of the $n \times m$-matrix $T_{u,v}C$ is

$$\begin{aligned}
&\operatorname{row}_i \left(T_{u,v}C \right) \\
&= \operatorname{row}_i C - \underbrace{\operatorname{row}_i \left(E_{u,u}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (111), applied to} \\ x=u \text{ and } y=u)}} - \underbrace{\operatorname{row}_i \left(E_{v,v}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (111), applied to} \\ x=v \text{ and } y=v)}} + \underbrace{\operatorname{row}_i \left(E_{u,v}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (111), applied to} \\ x=u \text{ and } y=v)}} + \underbrace{\operatorname{row}_v \left(E_{v,u}C \right)}_{\substack{=0_{1\times m} \\ \text{(by (111), applied to} \\ x=v \text{ and } y=u)}} \\
&\qquad \text{(by (117))} \\
&= \operatorname{row}_i C - 0_{1\times m} - 0_{1\times m} + 0_{1\times m} + 0_{1\times m} = \operatorname{row}_i C.
\end{aligned}$$

In other words, the i-th row of the $n \times m$-matrix $T_{u,v}C$ equals the i-th row of C. This proves Claim 3.

Now, we have proven Claim 1, Claim 2 and Claim 3; this completes the proof of Proposition 3.105. $\square$

The proof of Proposition 3.107 is analogous.

Proof of Corollary 3.108. **(a)** Proposition 3.105 (applied to $m = n$ and $C = I_n$) shows that $T_{u,v} I_n$ is the $n \times n$-matrix obtained from I_n by swapping the u-th row with the v-th row. In other words, $T_{u,v}$ is the $n \times n$-matrix obtained from I_n by swapping the u-th row with the v-th row (since $T_{u,v} I_n = T_{u,v}$). This proves Corollary 3.108 **(a)**.

(b) The proof of Corollary 3.108 **(b)** is similar, except that now we have to use Proposition 3.107. □

Proof of Proposition 3.109. Proposition 3.18 **(d)** shows that any two $n \times m$-matrices A and B (where $m \in \mathbb{N}$) satisfy

$$(A + B)^T = A^T + B^T. \tag{118}$$

Similarly, any two $n \times m$-matrices A and B (where $m \in \mathbb{N}$) satisfy

$$(A - B)^T = A^T - B^T. \tag{119}$$

By repeated application of (118) and (119), we can show the following fact: If A, B, C, D and E are five $n \times m$-matrices (for some $m \in \mathbb{N}$), then

$$(A - B - C + D + E)^T = A^T - B^T - C^T + D^T + E^T. \tag{120}$$

The definition of $T_{u,v}$ yields $T_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$. Hence,

$$\begin{aligned}
(T_{u,v})^T &= (I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u})^T \\
&= \underbrace{(I_n)^T}_{\substack{= I_n \\ \text{(by Proposition 3.18 (a))}}} - \underbrace{(E_{u,u})^T}_{\substack{= E_{u,u} \\ \text{(by Proposition 3.49, applied to } n, u \text{ and } u \\ \text{instead of } m, u \text{ and } v)}} - \underbrace{(E_{v,v})^T}_{\substack{= E_{v,v} \\ \text{(by Proposition 3.49, applied to } n, v \text{ and } v \\ \text{instead of } m, u \text{ and } v)}} \\
&\quad + \underbrace{(E_{u,v})^T}_{\substack{= E_{v,u} \\ \text{(by Proposition 3.49, applied to } m{=}n)}} + \underbrace{(E_{v,u})^T}_{\substack{= E_{u,v} \\ \text{(by Proposition 3.49, applied to } n, v \text{ and } u \\ \text{instead of } m, u \text{ and } v)}} \\
&\quad \left(\begin{array}{c} \text{by (120), applied to } m = n, A = I_n, B = E_{u,u}, \\ C = E_{v,v}, D = E_{u,v} \text{ and } E = E_{v,u} \end{array} \right) \\
&= I_n - E_{u,u} - E_{v,v} + E_{v,u} + E_{u,v} \\
&= I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u} = T_{u,v}.
\end{aligned}$$

This proves Proposition 3.109. □

Proof of Proposition 3.110. **(a)** The definition of $T_{u,u}$ yields $T_{u,u} = I_n - E_{u,u} - E_{u,u} + E_{u,u} + E_{u,u} = I_n$. This proves Proposition 3.110 **(a)**.

(b) *First proof:* Corollary 3.108 **(a)** shows that $T_{u,v}$ is the $n \times n$-matrix obtained from I_n by swapping the u-th row with the v-th row. Proposition 3.105 (applied

to $m = n$ and $C = T_{u,v}$) shows that $T_{u,v}T_{u,v}$ is the $n \times n$-matrix obtained from $T_{u,v}$ by swapping the u-th row with the v-th row. Thus, in order to obtain the matrix $T_{u,v}T_{u,v}$ from I_n, we have to do the following procedure:

- swap the u-th row with the v-th row;
- then, **again** swap the u-th row with the v-th row.

But this procedure clearly returns the matrix I_n with which we started (since the second swap undoes the first swap). Thus, $T_{u,v}T_{u,v} = I_n$.

From $T_{u,v}T_{u,v} = I_n$ and $T_{u,v}T_{u,v} = I_n$ (yes, this is one equality repeated twice), it follows that the matrix $T_{u,v}$ is an inverse of $T_{u,v}$. Hence, the matrix $T_{u,v}$ is invertible, and its inverse is $(T_{u,v})^{-1} = T_{u,v}$. This proves Proposition 3.110 **(b)**.

Second proof: We can also prove Proposition 3.110 **(b)** using Proposition 3.52, provided that we can tolerate some rather lengthy computations.

If i, j, x and y are four elements of $\{1, 2, \ldots, n\}$, then

$$E_{i,j}E_{x,y} = \delta_{j,x}E_{i,y} \tag{121}$$

(where all three matrices $E_{i,j}$, $E_{x,y}$ and $E_{i,y}$ have to be understood as $n \times n$-matrices)[81]. In particular, every $i \in \{1, 2, \ldots, n\}$ and $y \in \{1, 2, \ldots, n\}$ satisfy

$$E_{i,u}E_{v,y} = 0_{n\times n} \tag{122}$$

[82] and

$$E_{i,v}E_{u,y} = 0_{n\times n} \tag{123}$$

[83]. Furthermore, if i, x and y are three elements of $\{1, 2, \ldots, n\}$, then

$$E_{i,x}E_{x,y} = E_{i,y} \tag{124}$$

[84].

[81] *Proof of (121):* Let i, j, x and y be four elements of $\{1, 2, \ldots, n\}$. Then, Proposition 3.52 (applied to n, n, i and j instead of m, p, u and v) shows that $E_{i,j,n,n}E_{x,y,n,n} = \delta_{j,x}E_{i,y,n,n}$. Since we abbreviate the matrices $E_{i,j,n,n}$, $E_{x,y,n,n}$ and $E_{i,y,n,n}$ as $E_{i,j}$, $E_{x,y}$ and $E_{i,y}$ (respectively), this can be rewritten as $E_{i,j}E_{x,y} = \delta_{j,x}E_{i,y}$. Thus, (121) is proven.

[82] *Proof of (122):* Let $i \in \{1, 2, \ldots, n\}$ and $y \in \{1, 2, \ldots, n\}$. Then, (121) (applied to $j = u$ and $x = v$) yields $E_{i,u}E_{v,y} = \underbrace{\delta_{u,v}}_{\substack{=0 \\ (\text{since } u \neq v)}} E_{i,y} = 0E_{i,y} = 0_{n\times n}$, qed.

[83] *Proof of (123):* Let $i \in \{1, 2, \ldots, n\}$ and $y \in \{1, 2, \ldots, n\}$. Then, (121) (applied to $j = v$ and $x = u$) yields $E_{i,v}E_{u,y} = \underbrace{\delta_{v,u}}_{\substack{=0 \\ (\text{since } v \neq u)}} E_{i,y} = 0E_{i,y} = 0_{n\times n}$, qed.

[84] *Proof of (124):* Let i, j and x be three elements of $\{1, 2, \ldots, n\}$. Then, (121) (applied to $j = x$) yields $E_{i,x}E_{x,y} = \underbrace{\delta_{x,x}}_{\substack{=1 \\ (\text{since } x = x)}} E_{i,y} = 1E_{i,y} = E_{i,y}$, qed.

Now, the definition of $T_{u,v}$ yields $T_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$. Hence,

$$
\begin{aligned}
&= \underbrace{T_{u,v}}_{= I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}} \underbrace{T_{u,v}}_{= I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}} \\
&= \left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right) \\
&= \underbrace{I_n\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)}_{= I_n I_n - I_n E_{u,u} - I_n E_{v,v} + I_n E_{u,v} + I_n E_{v,u}} \\
&\quad - \underbrace{E_{u,u}\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)}_{= E_{u,u} I_n - E_{u,u} E_{u,u} - E_{u,u} E_{v,v} + E_{u,u} E_{u,v} + E_{u,u} E_{v,u}} \\
&\quad - \underbrace{E_{v,v}\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)}_{= E_{v,v} I_n - E_{v,v} E_{u,u} - E_{v,v} E_{v,v} + E_{v,v} E_{u,v} + E_{v,v} E_{v,u}} \\
&\quad + \underbrace{E_{u,v}\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)}_{= E_{u,v} I_n - E_{u,v} E_{u,u} - E_{u,v} E_{v,v} + E_{u,v} E_{u,v} + E_{u,v} E_{v,u}} \\
&\quad + \underbrace{E_{v,u}\left(I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}\right)}_{= E_{v,u} I_n - E_{v,u} E_{u,u} - E_{v,u} E_{v,v} + E_{v,u} E_{u,v} + E_{v,u} E_{v,u}}
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\underbrace{I_n I_n}_{=I_n} - \underbrace{I_n E_{u,u}}_{=E_{u,u}} - \underbrace{I_n E_{v,v}}_{=E_{v,v}} + \underbrace{I_n E_{u,v}}_{=E_{u,v}} + \underbrace{I_n E_{v,u}}_{=E_{v,u}} \right) \\
&\quad - \left(\underbrace{E_{u,u} I_n}_{=E_{u,u}} - \underbrace{E_{u,u} E_{u,u}}_{\substack{=E_{u,u} \\ \text{(by (124), applied} \\ \text{to } i=u,\ x=u \text{ and } y=u)}} - \underbrace{E_{u,u} E_{v,v}}_{\substack{=0_{n\times n} \\ \text{(by (122), applied} \\ \text{to } i=u \text{ and } y=v)}} + \underbrace{E_{u,u} E_{u,v}}_{\substack{=E_{u,v} \\ \text{(by (124), applied} \\ \text{to } i=u,\ x=u \text{ and } y=v)}} + \underbrace{E_{u,u} E_{v,u}}_{\substack{=0_{n\times n} \\ \text{(by (122), applied} \\ \text{to } i=u \text{ and } y=u)}} \right) \\
&\quad - \left(\underbrace{E_{v,v} I_n}_{=E_{v,v}} - \underbrace{E_{v,v} E_{u,u}}_{\substack{=0_{n\times n} \\ \text{(by (123), applied} \\ \text{to } i=v \text{ and } y=u)}} - \underbrace{E_{v,v} E_{v,v}}_{\substack{=E_{v,v} \\ \text{(by (124), applied} \\ \text{to } i=v,\ x=v \text{ and } y=v)}} + \underbrace{E_{v,v} E_{u,v}}_{\substack{=0_{n\times n} \\ \text{(by (123), applied} \\ \text{to } i=v \text{ and } y=v)}} + \underbrace{E_{v,v} E_{v,u}}_{\substack{=E_{v,u} \\ \text{(by (124), applied} \\ \text{to } i=v,\ x=v \text{ and } y=u)}} \right) \\
&\quad + \left(\underbrace{E_{u,v} I_n}_{=E_{u,v}} - \underbrace{E_{u,v} E_{u,u}}_{\substack{=0_{n\times n} \\ \text{(by (123), applied} \\ \text{to } i=u \text{ and } y=u)}} - \underbrace{E_{u,v} E_{v,v}}_{\substack{=E_{u,v} \\ \text{(by (124), applied} \\ \text{to } i=u,\ x=v \text{ and } y=v)}} + \underbrace{E_{u,v} E_{u,v}}_{\substack{=0_{n\times n} \\ \text{(by (123), applied} \\ \text{to } i=u \text{ and } y=v)}} + \underbrace{E_{u,v} E_{v,u}}_{\substack{=E_{u,u} \\ \text{(by (124), applied} \\ \text{to } i=u,\ x=v \text{ and } y=u)}} \right) \\
&\quad + \left(\underbrace{E_{v,u} I_n}_{=E_{v,u}} - \underbrace{E_{v,u} E_{u,u}}_{\substack{=E_{v,u} \\ \text{(by (124), applied} \\ \text{to } i=v,\ x=u \text{ and } y=u)}} - \underbrace{E_{v,u} E_{v,v}}_{\substack{=0_{n\times n} \\ \text{(by (122), applied} \\ \text{to } i=v \text{ and } y=v)}} + \underbrace{E_{v,u} E_{u,v}}_{\substack{=E_{v,v} \\ \text{(by (124), applied} \\ \text{to } i=v,\ x=u \text{ and } y=v)}} + \underbrace{E_{v,u} E_{v,u}}_{\substack{=0_{n\times n} \\ \text{(by (122), applied} \\ \text{to } i=v \text{ and } y=u)}} \right) \\
&= (I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}) - (E_{u,u} - E_{u,u} - 0_{n\times n} + E_{u,v} + 0_{n\times n}) \\
&\qquad - (E_{v,v} - 0_{n\times n} - E_{v,v} + 0_{n\times n} + E_{v,u}) + (E_{u,v} - 0_{n\times n} - E_{u,v} + 0_{n\times n} + E_{u,u}) \\
&\qquad + (E_{v,u} - E_{v,u} - 0_{n\times n} + E_{v,v} + 0_{n\times n}) \\
&= I_n.
\end{aligned}
$$

From $T_{u,v} T_{u,v} = I_n$ and $T_{u,v} T_{u,v} = I_n$ (yes, this is one equality repeated twice), it follows that the matrix $T_{u,v}$ is an inverse of $T_{u,v}$. Hence, the matrix $T_{u,v}$ is invertible, and its inverse is $(T_{u,v})^{-1} = T_{u,v}$. This proves Proposition 3.110 **(b)** again.

(c) The definition of $T_{v,u}$ yields $T_{v,u} = I_n - E_{v,v} - E_{u,u} + E_{v,u} + E_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$. Comparing this with $T_{u,v} = I_n - E_{u,u} - E_{v,v} + E_{u,v} + E_{v,u}$, we obtain $T_{v,u} = T_{u,v}$. This proves Proposition 3.110 **(c)**. □

3.19. Permutation matrices

Definition 3.111. Let $n \in \mathbb{N}$. A *swapping $n \times n$-matrix* means a matrix of the form $T_{u,v}$, where u and v are two distinct elements of $\{1, 2, \ldots, n\}$. When n is clear from the context, we shall omit the "$n \times n$-" and simply say "swapping matrix".

In Theorem 3.63, we have characterized lower-unitriangular matrices as products of lower addition matrices. Similarly, in Theorem 3.95, we have characterized invertibly lower-triangular matrices as products of scaling matrices and lower addition matrices. What kind of matrices are characterized as products of swapping matrices $T_{u,v}$? We should not expect anything with "triangular" in its name (after all, the matrices $T_{u,v}$ themselves are not triangular). Instead, we obtain the so-called permutation matrices.

Definition 3.112. Let $n \in \mathbb{N}$. An $n \times n$-matrix A is said to be a *permutation matrix* if it satisfies the following conditions:

(a) Each entry of A is either a 0 or a 1.

(b) Each row of A has exactly one entry equal to 1.

(c) Each column of A has exactly one entry equal to 1.

Example 3.113. **(a)** The 3×3-matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ is a permutation matrix.

(b) The 3×3-matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 3 \\ 1 & 0 & 0 \end{pmatrix}$ is not a permutation matrix, since it fails condition **(a)** of Definition 3.112.

(c) The 3×3-matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ is not a permutation matrix, since it fails condition **(c)** of Definition 3.112 (namely, the 3-rd column has two entries equal to 1, whereas the 1-st column has none).

(d) The 3×3-matrix $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ is not a permutation matrix, since it fails condition **(b)** of Definition 3.112 (namely, the 2-nd row has two entries equal to 1, whereas the 1-st row has none).

(e) For each $n \in \mathbb{N}$, the $n \times n$ identity matrix I_n is a permutation matrix. (This is Lemma 3.118 further below.)

(f) Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Then, the swapping matrix $T_{u,v}$ (defined in Definition 3.102) is a permutation matrix. (This is a particular case of Lemma 3.119 below.)

These examples do not exhaust the set of all permutation matrices. However (unlike, e.g., the lower-triangular matrices), this set is finite for each $n \in \mathbb{N}$. More precisely:

Proposition 3.114. Let $n \in \mathbb{N}$. Then, there are precisely $n!$ permutation matrices of size $n \times n$.

(Recall that $n!$ denotes the number $1 \cdot 2 \cdot 3 \cdot \cdots \cdot n$; it is called the "factorial of n". For instance, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$.)

We will outline a proof of Proposition 3.114 in Section 3.21.

Example 3.115. Proposition 3.114 (applied to $n = 3$) says that there are precisely $3! = 6$ permutation matrices of size 3×3. Here are they:

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad T_{1,2} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$T_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad T_{2,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

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

The two last matrices, which I have here denoted by A and B, are neither the identity matrix I_3 nor swapping matrices $T_{u,v}$. However, they can be written as products of swapping matrices:

$$A = T_{1,2}T_{2,3}, \qquad B = T_{2,3}T_{1,2}.$$

(Of course, they can also be written as products of swapping matrices in many other ways. For instance, $A = T_{1,3}T_{1,2} = T_{2,3}T_{2,3}T_{1,2}T_{2,3} = T_{2,3}T_{1,2}T_{2,3}T_{1,2}$.) This is not a coincidence: As we will see shortly (in Theorem 3.116), the permutation matrices are precisely the products of swapping matrices.

Theorem 3.116. Let $n \in \mathbb{N}$. An $n \times n$-matrix C is a permutation matrix if and only if C is a product of swapping matrices.

Some authors (e.g., Olver and Shakiban in [OlvSha06, Chapter 1, Definition 1.8]) use Theorem 3.116 as a definition of permutation matrices. (I.e., they **define** permutation matrices as products of swapping matrices, instead of using Definition 3.112.) More precisely, the following equivalent definitions of permutation matrices exist:

- Our Definition 3.112 above.
- An $n \times n$-matrix is called a *permutation matrix* if it is a product of swapping matrices. (This is the definition used in [OlvSha06, Chapter 1, Definition 1.8]; and Theorem 3.116 reveals that it is equivalent to our Definition 3.112.)

- An $n \times n$-matrix is called a *permutation matrix* if it has the same rows as I_n but (possibly) in a different order.
- An $n \times n$-matrix is called a *permutation matrix* if it has the same columns as I_n but (possibly) in a different order.
- An $n \times n$-matrix is called a *permutation matrix* if it has the form $\left(\delta_{w(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ for some bijective map $w : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. (This is the definition used in Stanley's [Stanle12, §1.5]. We will discuss it in some more detail in Section 3.21.)

The equivalence of all these definitions is easy to see (once Theorem 3.116 is proven); nevertheless, we shall be using Definition 3.112 only.

We shall give a complete proof of Theorem 3.116 in Section 3.20; but first, let us state some basic facts on which said proof relies:

Lemma 3.117. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix. Assume that A is a permutation matrix. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$. Let B be the matrix obtained from A by swapping the u-th row with the v-th row. Then, B is a permutation matrix.

Proof of Lemma 3.117. We know that A is a permutation matrix. According to the definition of a "permutation matrix", this means that A satisfies the following three statements:

Statement 1: Each entry of A is either a 0 or a 1.

Statement 2: Each row of A has exactly one entry equal to 1.

Statement 3: Each column of A has exactly one entry equal to 1.

Thus, we know that Statements 1, 2 and 3 are satisfied.

On the other hand, we want to prove that B is a permutation matrix. According to the definition of a "permutation matrix", this means proving that B satisfies the following three statements:

Statement 4: Each entry of B is either a 0 or a 1.

Statement 5: Each row of B has exactly one entry equal to 1.

Statement 6: Each column of B has exactly one entry equal to 1.

Hence, it remains to prove that Statements 4, 5 and 6 are satisfied.

Recall that the matrix B is obtained from A by swapping the u-th row with the v-th row. Hence, each row of B equals some row of A. Thus, Statement 5 follows from Statement 2. Therefore, Statement 5 is satisfied.

Also, each entry of B equals some entry of A (since each row of B equals some row of A). Thus, Statement 4 follows from Statement 1. Hence, Statement 4 is satisfied.

Recall again that the matrix B is obtained from A by swapping the u-th row with the v-th row. Hence, each column of B is obtained from the corresponding column of A by swapping the u-th entry with the v-th entry. Hence, each column of B has exactly as many entries equal to 1 as the corresponding column of A (because swapping two entries does not change the number of entries equal to 1). Therefore, Statement 6 follows from Statement 3. Hence, Statement 6 is satisfied.

We thus have shown that Statements 4, 5 and 6 are satisfied. As we have said, this completes the proof of Lemma 3.117. □

Lemma 3.118. Let $n \in \mathbb{N}$. Then, the identity matrix I_n is a permutation matrix.

Lemma 3.118 is very easy to prove (and will be proven in detail in Section 3.20 below).

Lemma 3.119. Let $n \in \mathbb{N}$. Then, any product of swapping $n \times n$-matrices is a permutation matrix.

(Here, we are again using the convention that the empty product of $n \times n$-matrices is I_n.)

Lemma 3.119 is, of course, the $\Longleftarrow$ direction of Theorem 3.116; it can be proven by induction using Lemma 3.117 and Proposition 3.105. The (rather straightforward) proof can be found in Section 3.20 below.

Now, let me give some examples for Theorem 3.116:

Example 3.120. **(a)** For each $n \in \mathbb{N}$, the $n \times n$ identity matrix I_n is a product of swapping matrices: Namely, it is the empty product (since the empty product of $n \times n$-matrices is I_n by definition).

(b) Each swapping matrix $T_{u,v}$ itself is a product of swapping matrices: namely, it is a product of itself.

(c) Let C be the permutation matrix $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. Then, C is a product of swapping matrices: Namely, it equals $T_{1,2}T_{2,4}T_{3,4}$.

Let us actually see how this representation of C can be found. We shall proceed by writing C as a product of one swapping matrix with a second matrix C', which is still a permutation matrix but has one diagonal entry equal to 1. We then will do the same with C', obtaining a third matrix C''; then, do the same with C'', and so on. At the end, we will be left with a permutation matrix whose **all** diagonal entries are 1. This means that we will be left with the identity matrix I_4.

In more detail: We proceed in several steps:

Step 1: Let us turn the $(1,1)$-th entry of C into 1 by swapping the 1-st row with the 2-nd row. Denote the resulting matrix by C'. Thus, $C' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. Notice that the new matrix C' is still a permutation matrix (by Lemma 3.117, applied to $A = C$, $B = C'$, $u = 1$ and $v = 2$), but now has its first diagonal entry equal to 1. Since C' was obtained from C by swapping the 1-st row with the 2-nd row, we can conversely obtain C from C' by swapping the 1-st row with the 2-nd row. According to Proposition 3.105 (applied to n, 1, 2 and C' instead of m, u, v and C), this means that $C = T_{1,2}C'$.

Step 2: Let us turn the $(2,2)$-th entry of C' into 1 by swapping the 2-nd row with the 4-th row. Denote the resulting matrix by C''. Thus, $C'' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$. Again, the new matrix C'' is still a permutation matrix (by Lemma 3.117, applied to $A = C'$, $B = C''$, $u = 2$ and $v = 4$), and still has its first diagonal entry equal to 1 (since the 1-st row has not been changed); but now its second diagonal entry is also 1. Similarly to how we found that $C = T_{1,2}C'$ in Step 1, we now obtain $C' = T_{2,4}C''$.

Step 3: Let us turn the $(3,3)$-th entry of C'' into 1 by swapping the 3-rd row with the 4-th row. Denote the resulting matrix by C'''. Thus, $C''' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Again, the new matrix C''' is still a permutation matrix, and the first two diagonal entries are still equal to 1; and now the third diagonal entry has become 1 as well. Similarly to how we found that $C = T_{1,2}C'$ in Step 1, we now obtain $C'' = T_{3,4}C'''$.

Step 4: We should now turn the $(4,4)$-th entry of C''' into 1, but fortunately this is unnecessary: It already is 1.

We have thus turned all diagonal entries into 1. Our final matrix C''' thus equals I_4 (since it is a permutation matrix). Combining the three equalities we have found, we obtain

$$\begin{aligned} C &= T_{1,2} \underbrace{C'}_{=T_{2,4}C''} = T_{1,2}T_{2,4} \underbrace{C''}_{=T_{3,4}C'''} = T_{1,2}T_{2,4}T_{3,4} \underbrace{C'''}_{=I_4} \\ &= T_{1,2}T_{2,4}T_{3,4}I_4 = T_{1,2}T_{2,4}T_{3,4}. \end{aligned}$$

Thus we have represented C as a product of swapping matrices.

Example 3.120 **(c)** essentially demonstrates how Theorem 3.116 (or, more precisely, the $\Longrightarrow$ direction of Theorem 3.116) can be proven in the general case (similarly to how Example 3.64 **(c)** outlines the proof of Theorem 3.63, and how Example 3.96 **(c)** outlines the proof of Theorem 3.95). Transforming the example into an actual rigorous proof, however, requires work: Not only would we have to formalize the algorithm, but we would also need to formally justify that the algorithm works[85]. In Section 3.20, we shall give a proof of Theorem 3.116 which is, **more or less**, the one suggested by Example 3.120 **(c)**; however, it will be organized rather differently (for the sake of easier readability)[86].

One easy corollary of Theorem 3.116 is the following:

Proposition 3.121. Let $n \in \mathbb{N}$. Let A and B be two $n \times n$-matrices that are permutation matrices. Then, AB is also a permutation matrix.

Furthermore, we can show:

Proposition 3.122. Let $n \in \mathbb{N}$. Let A be an $n \times n$-matrix that is a permutation matrix. Then:
(a) The matrix A is invertible.
(b) Its inverse is $A^{-1} = A^T$.
(c) This inverse A^{-1} is a permutation matrix.

Proposition 3.123. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let P be an $n \times n$-matrix that is a permutation matrix. Let C be an $n \times m$-matrix. Then, the $n \times m$-matrix PC can be obtained from C by rearranging the rows in a certain way that depends on P. (In more rigorous terms, this means that there exists a bijective map $w : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ such that every $i \in \{1, 2, \ldots, n\}$ satisfies $\operatorname{row}_i (PC) = \operatorname{row}_{w(i)} C$. If this sounds confusing to you, think of this map w as a way to match up the rows of PC with the rows of C such that each row of PC equals the corresponding row of C. We will go over this in more detail in Section 3.21.)

Example 3.124. Let $n = 3$ and $m = 2$. Let C be the 3×2-matrix $\begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix}$. Let P be the 3×3-matrix $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$; as we know, this is a permutation matrix. Then, Proposition 3.123 claims that the 3×2-matrix PC can be obtained from C

[85] For example, we would need to verify that each step results in a permutation matrix, and that the k-th step (for each k) leaves the first $k-1$ diagonal entries unchanged.

[86] Namely, instead of using a procedure with several steps, it will be based on an induction argument. The ideas will, of course, be the same; this is just an example of how algorithmic arguments can be rewritten as induction proofs.

by rearranging the rows in a certain way. And this can indeed be confirmed by a computation:

$$PC = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ a' & b' \\ a'' & b'' \end{pmatrix} = \begin{pmatrix} a' & b' \\ a'' & b'' \\ a & b \end{pmatrix}.$$

The rearrangement moves the first row to the very bottom, while letting the other two rows slide up one level. Other permutation matrices P would produce other rearrangements.

Proposition 3.123 is (in a sense) a partial generalization of Proposition 3.105 (although, of course, not a complete generalization, since it fails to specify the precise rearrangement). A similar partial generalization can be stated for Proposition 3.107; this time, of course, it will be the columns (not the rows) that get rearranged in CP.

We will prove Proposition 3.123 in Section 3.21.

3.20. (*) Proofs about permutation matrices

Let us now catch up on some proofs that we promised in the previous section.

Proof of Lemma 3.118. We want to prove that I_n is a permutation matrix. According to the definition of a "permutation matrix", this means proving that I_n satisfies the following three statements:

Statement 1: Each entry of I_n is either a 0 or a 1.

Statement 2: Each row of I_n has exactly one entry equal to 1.

Statement 3: Each column of I_n has exactly one entry equal to 1.

Hence, it remains to prove that Statements 1, 2 and 3 are satisfied.

But Statement 1 is obvious. Statement 2 is also clear (since each row of I_n has exactly one entry equal to 1 – namely, the diagonal entry), and Statement 3 is clear as well (for similar reasons). Thus, Statements 1, 2 and 3 are satisfied. This proves Lemma 3.118. □

Proof of Lemma 3.119. Let M be any product of swapping $n \times n$-matrices. We must show that M is a permutation matrix.

We have assumed that M is a product of swapping $n \times n$-matrices. In other words, $M = A_k A_{k-1} \cdots A_1$ for some $k \in \mathbb{N}$ and some k swapping $n \times n$-matrices $A_1, A_2, \ldots, A_k$. Consider this k and these $A_1, A_2, \ldots, A_k$.

We shall show that

$$A_i A_{i-1} \cdots A_1 \text{ is a permutation matrix} \tag{125}$$

for every $i \in \{0, 1, \ldots, k\}$ (where, as usual, $A_0 A_{-1} \cdots A_1$ has to be interpreted as an empty product and thus equals I_n).

[*Proof of (125):* We will prove (125) by induction over i:

Induction base: Lemma 3.118 says that I_n is a permutation matrix. In other words, $A_0 A_{-1} \cdots A_1$ is a permutation matrix[87]. In other words, (125) holds for $i = 0$. This completes the induction base.

Induction step: Let $j \in \{0, 1, \ldots, k\}$ be positive. Assume (as the *induction hypothesis*) that (125) holds for $i = j - 1$. We must show that (125) holds for $i = j$.

The induction hypothesis tells us that (125) holds for $i = j - 1$. In other words, $A_{j-1} A_{j-2} \cdots A_1$ is a permutation matrix. Set $C = A_{j-1} A_{j-2} \cdots A_1$. Thus, C is a permutation matrix (since $A_{j-1} A_{j-2} \cdots A_1$ is a permutation matrix).

But A_j is a swapping $n \times n$-matrix (since $A_1, A_2, \ldots, A_k$ are k swapping $n \times n$-matrices). In other words, A_j has the form $A_j = T_{u,v}$, where u and v are two distinct elements of $\{1, 2, \ldots, n\}$. Consider these u and v.

Now,

$$A_j A_{j-1} \cdots A_1 = \underbrace{A_j}_{=T_{u,v}} \underbrace{\left(A_1 A_2 \cdots A_{j-1}\right)}_{=C} = T_{u,v} C. \tag{126}$$

But Proposition 3.105 (applied to $m = n$) shows that $T_{u,v} C$ is the $n \times n$-matrix obtained from C by swapping the u-th row with the v-th row. Hence, Lemma 3.117 (applied to $A = C$ and $B = T_{u,v} C$) shows that $T_{u,v} C$ is a permutation matrix. In light of (126), this rewrites as follows: $A_j A_{j-1} \cdots A_1$ is a permutation matrix. In other words, (125) holds for $i = j$. This completes the induction step, and thus the inductive proof of (125).]

Now, (125) (applied to $i = k$) yields that $A_k A_{k-1} \cdots A_1$ is a permutation matrix. In other words, M is a permutation matrix (since $M = A_k A_{k-1} \cdots A_1$). This completes the proof of Lemma 3.119. □

The general proof of Theorem 3.116 follows the idea outlined in Example 3.120 **(c)**, but we are going to make it more manageable by introducing a convenient notion:

Definition 3.125. Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. An $n \times n$-matrix A is said to be *k-identical* if it satisfies $A_{1,1} = A_{2,2} = \cdots = A_{k,k} = 1$. (Note that the condition $A_{1,1} = A_{2,2} = \cdots = A_{k,k} = 1$ means "$A_{i,i} = 1$ for each $i \in \{1, 2, \ldots, k\}$". Thus, if $k = 0$, then this condition is vacuously true, since there exists no $i \in \{1, 2, \ldots, k\}$ in this case.)

This notion allows us to speak about our procedure from Example 3.120 **(c)** more crisply: We started with an arbitrary permutation matrix C, which was 0-identical. Then, in Step 1, we made it 1-identical by switching two rows. Then, in Step 2, we made it 2-identical by switching two further rows. Then, in Step 3, we made it 3-identical by switching two further rows. Then, in Step 4, we made it 4-identical

[87] since $A_0 A_{-1} \cdots A_1 = (\text{empty product of } n \times n\text{-matrices}) = I_n$

by doing nothing (since it already was 4-identical). At the end of the procedure, it was an identity matrix.

Here are some properties of k-identical permutation matrices:

Lemma 3.126. Let $n \in \mathbb{N}$ and $k \in \{0, 1, \ldots, n\}$. Let A be an $n \times n$-matrix. Assume that A is a k-identical permutation matrix.

(a) If $k = n$, then $A = I_n$.

(b) For each $v \in \{k+1, k+2, \ldots, n\}$, there exists some $u \in \{k+1, k+2, \ldots, n\}$ such that $A_{u,v} = 1$.

(c) If $k < n$ and $A_{k+1,k+1} = 1$, then A is a $(k+1)$-identical permutation matrix.

(d) If u and v are two distinct elements of $\{k+1, k+2, \ldots, n\}$ satisfying $v = k+1$ and $A_{u,v} = 1$, then $T_{u,v}A$ is a $(k+1)$-identical permutation matrix.

Proof of Lemma 3.126. We have assumed that A is a permutation matrix. According to the definition of a "permutation matrix", this means that A satisfies the following three statements:

Statement 1: Each entry of A is either a 0 or a 1.

Statement 2: Each row of A has exactly one entry equal to 1.

Statement 3: Each column of A has exactly one entry equal to 1.

Thus, we know that Statements 1, 2 and 3 are satisfied.

We have assumed that A is k-identical. In other words,

$$A_{1,1} = A_{2,2} = \cdots = A_{k,k} = 1 \tag{127}$$

(by the definition of "k-identical"). In other words,

$$A_{i,i} = 1 \qquad \text{for each } i \in \{1, 2, \ldots, k\}. \tag{128}$$

Next, we observe that

$$A_{i,j} = \delta_{i,j} \qquad \text{for all } i \in \{1, 2, \ldots, k\} \text{ and } j \in \{1, 2, \ldots, n\} \tag{129}$$

[88].

[88] *Proof of (129):* Let $i \in \{1, 2, \ldots, k\}$ and $j \in \{1, 2, \ldots, n\}$. We must prove that $A_{i,j} = \delta_{i,j}$.

We have $i \in \{1, 2, \ldots, k\}$. Hence, $A_{i,i} = 1$ (by (128)).

We are in one of the following two cases:

Case 1: We have $i = j$.

Case 2: We have $i \neq j$.

Let us first consider Case 1. In this case, we have $i = j$. Hence, $j = i$, so that $A_{i,j} = A_{i,i} = 1$. Comparing this with $\delta_{i,j} = 1$ (since $i = j$), we obtain $A_{i,j} = \delta_{i,j}$. Hence, $A_{i,j} = \delta_{i,j}$ is proven in Case 1.

Let us now consider Case 2. In this case, we have $i \neq j$. Thus, $\delta_{i,j} = 0$.

(a) Assume that $k = n$. Then, every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy $A_{i,j} = (I_n)_{i,j}$ [89]. In other words, each entry of the matrix A equals the corresponding entry of I_n. In other words, $A = I_n$. This proves Lemma 3.126 **(a)**.

(b) Let $v \in \{k+1, k+2, \ldots, n\}$. Statement 3 shows that each column of A has exactly one entry equal to 1. In particular, the v-th column has exactly one entry equal to 1. In other words, there exists exactly one $i \in \{1, 2, \ldots, n\}$ such that $A_{i,v} = 1$. Consider this i. We are going to show that $i \in \{k+1, k+2, \ldots, n\}$.

In fact, assume the contrary. Thus, $i \notin \{k+1, k+2, \ldots, n\}$. Combining $i \in \{1, 2, \ldots, n\}$ with $i \notin \{k+1, k+2, \ldots, n\}$, we obtain

$$i \in \{1, 2, \ldots, n\} \setminus \{k+1, k+2, \ldots, n\} = \{1, 2, \ldots, k\}.$$

Hence, $A_{i,v} = \delta_{i,v}$ (by (129), applied to $j = v$). But $i \in \{1, 2, \ldots, k\}$, so that $i \leq k$. However, $v \geq k+1$ (since $v \in \{k+1, k+2, \ldots, n\}$). Hence, $k+1 \leq v$, and thus $i \leq k < k+1 \leq v$. Thus, $i \neq v$, so that $\delta_{i,v} = 0$ and thus $A_{i,v} = \delta_{i,v} = 0$. This contradicts $A_{i,v} = 1 \neq 0$.

This contradiction shows that our assumption was false. Hence, $i \in \{k+1, k+2, \ldots, n\}$ is proven. Thus, there exists some $u \in \{k+1, k+2, \ldots, n\}$ such that $A_{u,v} = 1$ (namely, $u = i$). This proves Lemma 3.126 **(b)**.

(c) Assume that $k < n$ and $A_{k+1,k+1} = 1$. We need to show that A is a $(k+1)$-identical permutation matrix.

Combining (127) with $A_{k+1,k+1} = 1$, we obtain $A_{1,1} = A_{2,2} = \cdots = A_{k+1,k+1} = 1$.

But the matrix A is $(k+1)$-identical if and only if $A_{1,1} = A_{2,2} = \cdots = A_{k+1,k+1} = 1$ (because this is how "$(k+1)$-identical" is defined). Thus, the matrix A is $(k+1)$-

---

Now, assume (for the sake of contradiction) that $A_{i,j} \neq 0$. But $A_{i,j}$ is an entry of A, and thus is either a 0 or a 1 (by Statement 1). In other words, $A_{i,j} = 0$ or $A_{i,j} = 1$. Therefore, $A_{i,j} = 1$ (since $A_{i,j} \neq 0$). Combining this with $A_{i,i} = 1$, we conclude that the i-th row of A has at least two entries equal to 1: namely, the entries $A_{i,i}$ and $A_{i,j}$. (And these two entries actually lie in different cells, since $i \neq j$.)

But each row of A has exactly one entry equal to 1 (because Statement 2 is satisfied). In particular, the i-th row of A has exactly one entry equal to 1. This contradicts the fact that the i-th row of A has at least two entries equal to 1. This contradiction shows that our assumption (that $A_{i,j} \neq 0$) was false. Hence, we have $A_{i,j} = 0$. Compared with $\delta_{i,j} = 0$, this yields $A_{i,j} = \delta_{i,j}$. Thus, $A_{i,j} = \delta_{i,j}$ is proven in Case 2.

Now, $A_{i,j} = \delta_{i,j}$ is proven in each of the two Cases 1 and 2. Hence, $A_{i,j} = \delta_{i,j}$ always holds. In other words, (129) is proven.

[89] *Proof.* Recall that $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of I_n). Hence, every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy

$$(I_n)_{i,j} = \delta_{i,j}. \tag{130}$$

But $\{1, 2, \ldots, n\} = \{1, 2, \ldots, k\}$ (since $n = k$). Now, every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy

$$\begin{aligned} A_{i,j} &= \delta_{i,j} && \text{(by (129), since } i \in \{1, 2, \ldots, n\} = \{1, 2, \ldots, k\}\text{)} \\ &= (I_n)_{i,j} && \text{(by (130))}. \end{aligned}$$

identical (since $A_{1,1} = A_{2,2} = \cdots = A_{k+1,k+1} = 1$). Also, A is a permutation matrix. The proof of Lemma 3.126 **(c)** is thus complete.

(d) Let u and v be two distinct elements of $\{k+1, k+2, \ldots, n\}$ satisfying $v = k+1$ and $A_{u,v} = 1$. We must prove that $T_{u,v}A$ is a $(k+1)$-identical permutation matrix.

Let $B = T_{u,v}A$. Proposition 3.105 (applied to $m = n$ and $C = A$) shows that $T_{u,v}A$ is the $n \times n$-matrix obtained from A by swapping the u-th row with the v-th row. Since $B = T_{u,v}A$, this rewrites as follows: B is the $n \times n$-matrix obtained from A by swapping the u-th row with the v-th row. Hence, Lemma 3.117 shows that B is a permutation matrix. We shall next show that B is $(k+1)$-identical.

Recall that B is the $n \times n$-matrix obtained from A by swapping the u-th row with the v-th row. Hence, the following facts hold:

Fact 4: The u-th row of the matrix B equals the v-th row of A.

Fact 5: The v-th row of the matrix B equals the u-th row of A.

Fact 6: If $i \in \{1, 2, \ldots, n\}$ is such that $i \neq u$ and $i \neq v$, then the i-th row of the matrix B equals the i-th row of A.

Now, using Fact 6, we can easily see that $B_{i,i} = A_{i,i}$ for each $i \in \{1, 2, \ldots, k\}$ [90]. Hence, for each $i \in \{1, 2, \ldots, k\}$, we have $B_{i,i} = A_{i,i} = 1$ (by (128)). In other words,

$$B_{1,1} = B_{2,2} = \cdots = B_{k,k} = 1. \tag{131}$$

But Fact 5 shows that each entry of the v-th row of the matrix B equals the corresponding entry of the u-th row of A. In other words, $B_{v,j} = A_{u,j}$ for each $j \in \{1, 2, \ldots, n\}$. Applying this to $j = v$, we obtain $B_{v,v} = A_{u,v} = 1$. Since $v = k+1$, this rewrites as $B_{k+1,k+1} = 1$. Combining this with (131), we obtain $B_{1,1} = B_{2,2} = \cdots = B_{k+1,k+1} = 1$.

But the matrix B is $(k+1)$-identical if and only if $B_{1,1} = B_{2,2} = \cdots = B_{k+1,k+1} = 1$ (because this is how "$(k+1)$-identical" is defined). Thus, the matrix B is $(k+1)$-identical (since $B_{1,1} = B_{2,2} = \cdots = B_{k+1,k+1} = 1$). Hence, B is a $(k+1)$-identical permutation matrix (since we already know that B is a permutation matrix). In other words, $T_{u,v}A$ is a $(k+1)$-identical permutation matrix (since $B = T_{u,v}A$). This proves Lemma 3.126 **(d)**. □

Next, we show a slightly stronger version of the $\Longrightarrow$ direction of Theorem 3.116:

[90] *Proof.* Let $i \in \{1, 2, \ldots, k\}$. Then, $i \leq k < k+1 \leq u$ (since $u \geq k+1$ (since $u \in \{k+1, k+2, \ldots, n\}$)). Hence, $i \neq u$. The same argument (but made for v instead of u) shows that $i \neq v$. Hence, Fact 6 shows that the i-th row of the matrix B equals the i-th row of A. In other words, each entry of the i-th row of the matrix B equals the corresponding entry of the i-th row of A. In other words, $B_{i,j} = A_{i,j}$ for each $j \in \{1, 2, \ldots, n\}$. Applying this to $j = i$, we obtain $B_{i,i} = A_{i,i}$. Qed.

Lemma 3.127. Let $n \in \mathbb{N}$ and $p \in \{0,1,\ldots,n\}$. Let A be an $n \times n$-matrix. If A is an $(n-p)$-identical permutation matrix, then A is a product of at most p swapping matrices.

Proof of Lemma 3.127. We shall prove Lemma 3.127 by induction over p:

Induction base: If A is an $(n-0)$-identical permutation matrix, then A is a product of at most 0 swapping matrices[91]. In other words, Lemma 3.127 holds for $p = 0$. This completes the induction base.

Induction step: Let $q \in \{0,1,\ldots,n\}$ be positive. Assume (as the *induction hypothesis*) that Lemma 3.127 holds for $p = q-1$. We must now prove that Lemma 3.127 holds for $p = q$.

Assume that A is an $(n-q)$-identical permutation matrix. We shall show that

$$A \text{ is a product of at most } q \text{ swapping matrices.} \tag{132}$$

Set $k = n-q$. Thus, A is a k-identical permutation matrix (since A is an $(n-q)$-identical permutation matrix).

We have $k = n-q \in \{0,1,\ldots,n\}$ (since $q \in \{0,1,\ldots,n\}$). Furthermore, $k = n-q < n$ (since q is positive), so that $k \leq n-1$ (since k and n are integers).

Set $v = k+1$. Then, $v = k+1 \geq k+1$ and $v = k+1 \leq n$ (since $k \leq n-1$), so that $v \in \{k+1, k+2, \ldots, n\}$. Hence, Lemma 3.126 **(b)** shows that there exists some $u \in \{k+1, k+2, \ldots, n\}$ such that $A_{u,v} = 1$. Consider this u. We are in one of the following two cases:

Case 1: We have $u = k+1$.

Case 2: We have $u \neq k+1$.

Let us first consider Case 1. In this case, we have $u = k+1$. From $u = k+1$ and $v = k+1$, we obtain $A_{u,v} = A_{k+1,k+1}$, so that $A_{k+1,k+1} = A_{u,v} = 1$. Hence, Lemma 3.126 **(c)** shows that A is a $(k+1)$-identical permutation matrix. Since $\underbrace{k}_{=n-q} + 1 = n-q+1 = n-(q-1)$, this rewrites as follows: A is an $(n-(q-1))$-identical permutation matrix. But the induction hypothesis shows that Lemma 3.127 holds for $p = q-1$. Hence, Lemma 3.127 can be applied to $p = q-1$ (since A is an $(n-(q-1))$-identical permutation matrix). As a result, we conclude that A is a product of at most $q-1$ swapping matrices. In other words, there exists some $s \in \{0,1,\ldots,q-1\}$ such that A is a product of at most s swapping matrices. Consider this s. Thus, A is a product of at most q swapping matrices (since A is a product of s swapping matrices, but we have $s \in \{0,1,\ldots,q-1\} \subseteq \{0,1,\ldots,q\}$). In other words, (132) holds. Hence, (132) is proven in Case 1.

[91] *Proof.* Assume that A is an $(n-0)$-identical permutation matrix. In other words, A is an n-identical permutation matrix. Lemma 3.126 **(a)** (applied to $k = n$) thus yields $A = I_n$. Thus $A = I_n =$ (the empty product of swapping matrices) (since the empty product of swapping matrices is defined to be I_n). Hence, A is a product of 0 swapping matrices. Thus, A is a product of at most 0 swapping matrices. Qed.

Let us now consider Case 2. In this case, we have $u \neq k+1$. Thus, $u \neq k+1 = v$. Therefore, Lemma 3.126 **(d)** reveals that $T_{u,v}A$ is a $(k+1)$-identical permutation matrix. Since $\underbrace{k}_{=n-q}+1 = n-q+1 = n-(q-1)$, this rewrites as follows: $T_{u,v}A$ is an $(n-(q-1))$-identical permutation matrix. But the induction hypothesis shows that Lemma 3.127 holds for $p = q-1$. Hence, Lemma 3.127 can be applied to $q-1$ and $T_{u,v}A$ instead of p and A (since $T_{u,v}A$ is an $(n-(q-1))$-identical permutation matrix). As a result, we conclude that $T_{u,v}A$ is a product of at most $q-1$ swapping matrices. In other words, there exists some $s \in \{0,1,\ldots,q-1\}$ and some s swapping matrices $M_1, M_2, \ldots, M_s$ such that $T_{u,v}A = M_1M_2\cdots M_s$. Consider this s and these $M_1, M_2, \ldots, M_s$.

From $s \in \{0,1,\ldots,q-1\}$, we obtain $s+1 \in \{1,2,\ldots,q\} \subseteq \{0,1,\ldots,q\}$.

Proposition 3.110 **(b)** yields that the matrix $T_{u,v}$ is invertible, and its inverse is $(T_{u,v})^{-1} = T_{u,v}$. Thus, $T_{u,v}T_{u,v} = I_n$. Hence, $\underbrace{T_{u,v}T_{u,v}}_{=I_n}A = I_nA = A$, so that

$$A = T_{u,v}\underbrace{T_{u,v}A}_{=M_1M_2\cdots M_s} = T_{u,v}M_1M_2\cdots M_s.$$

But $T_{u,v}$ is a swapping matrix (by the definition of a "swapping matrix"), and the matrices $M_1, M_2, \ldots, M_s$ are swapping matrices as well. Hence, $T_{u,v}, M_1, M_2, \ldots, M_s$ are swapping matrices. Thus, $T_{u,v}M_1M_2\cdots M_s$ is a product of $s+1$ swapping matrices. In other words, A is a product of $s+1$ swapping matrices (since $A = T_{u,v}M_1M_2\cdots M_s$). Hence, A is a product of at most q swapping matrices (since $s+1 \in \{0,1,\ldots,q\}$). In other words, (132) holds. Hence, (132) is proven in Case 2.

Now, we have proven (132) in each of the two Cases 1 and 2. Hence, (132) always holds. In other words, A is a product of at most q swapping matrices.

Now, forget that we assumed that A is an $(n-q)$-identical permutation matrix. Thus, we have shown that if A is an $(n-q)$-identical permutation matrix, then A is a product of at most q swapping matrices. In other words, Lemma 3.127 holds for $p = q$. This completes the induction step. Thus, Lemma 3.127 is proven by induction. □

Theorem 3.116 is now an easy consequence of the above:

Proof of Theorem 3.116. $\Longleftarrow$: Lemma 3.119 says that any product of swapping matrices is a permutation matrix. Hence, if C is a product of swapping matrices, then C is a permutation matrix. This proves the $\Longleftarrow$ direction of Theorem 3.116.

$\Longrightarrow$: We need to prove that if C is a permutation matrix, then C is a product of swapping matrices.

So let us assume that C is a permutation matrix. Clearly, the matrix C is 0-identical[92]. In other words, the matrix C is $(n-n)$-identical (since $0 = n-n$).

[92] *Proof.* The matrix C satisfies the equality $C_{1,1} = C_{2,2} = \cdots = C_{0,0} = 1$ (in fact, this equality is vacuously true). But according to the definition of "0-identical", this means precisely that C is 0-identical. Qed.

Thus, Lemma 3.127 (applied to $p = n$ and $A = C$) shows that C is a product of at most n swapping matrices. Hence, C is a product of swapping matrices. This is precisely what we had to prove. This proves the $\Longrightarrow$ direction of Theorem 3.116. Hence, the proof of Theorem 3.116 is complete. □

Proof of Proposition 3.121. Theorem 3.116 (applied to $C = A$) shows that A is a permutation matrix if and only if A is a product of swapping matrices. Thus, A is a product of swapping matrices (since A is a permutation matrix). The same argument (applied to B instead of A) shows that B is a product of swapping matrices.

Now, we know that each of the two matrices A and B is a product of swapping matrices. Hence, their product AB is the product of two products of swapping matrices. Therefore, AB is itself a product of swapping matrices (since a product of two products of swapping matrices must itself be a product of swapping matrices).

Theorem 3.116 (applied to $C = AB$) shows that AB is a permutation matrix if and only if AB is a product of swapping matrices. Thus, AB is a permutation matrix (since AB is a product of swapping matrices). This proves Proposition 3.121. □

Let us now come to the proof of Proposition 3.122. Parts **(a)** and **(c)** of Proposition 3.122 could be verified similarly to how we have proved Theorem 3.67 (but using Theorem 3.116 now); but this would not help us proving part **(b)**. So we take a different path instead.

Proof of Proposition 3.122. We have assumed that A is a permutation matrix. According to the definition of a "permutation matrix", this means that A satisfies the following three statements:

Statement 1: Each entry of A is either a 0 or a 1.

Statement 2: Each row of A has exactly one entry equal to 1.

Statement 3: Each column of A has exactly one entry equal to 1.

Thus, we know that Statements 1, 2 and 3 are satisfied.
We have $A^T = \left(A_{j,i}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of A^T). Hence,

$$\left(A^T\right)_{i,j} = A_{j,i} \qquad \text{for every } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,n\}. \tag{133}$$

We are next going to show that

$$A^T \text{ is a permutation matrix.} \tag{134}$$

[*Proof of (134):* We must show that A^T is a permutation matrix. According to the definition of a "permutation matrix", this means proving that A^T satisfies the following three statements:

Statement 4: Each entry of A^T is either a 0 or a 1.

Statement 5: Each row of A^T has exactly one entry equal to 1.

Statement 6: Each column of A^T has exactly one entry equal to 1.

Hence, it remains to prove that Statements 4, 5 and 6 are satisfied.

The definition of A^T shows that the entries of A^T are precisely the entries of A, just moved to different cells. Thus, Statement 4 follows from Statement 1. Hence, Statement 4 is satisfied.

Let $i \in \{1,2,\ldots,n\}$. The i-th column of A has exactly one entry equal to 1 (by Statement 3). In other words, there exists exactly one $j \in \{1,2,\ldots,n\}$ satisfying $A_{j,i} = 1$. In view of (133), this rewrites as follows: There exists exactly one $j \in \{1,2,\ldots,n\}$ satisfying $\left(A^T\right)_{i,j} = 1$. In other words, the i-th row of A^T has exactly one entry equal to 1.

Now, forget that we fixed i. We thus have shown that for each $i \in \{1,2,\ldots,n\}$, the i-th row of A^T has exactly one entry equal to 1. In other words, each row of A^T has exactly one entry equal to 1. This proves Statement 5.

A similar argument (but with the roles of rows and columns switched, and also the roles of i and j switched, and using Statement 2 instead of Statement 3) can be used to prove Statement 6.

We thus have shown that Statements 4, 5 and 6 are satisfied. As we have said, this completes the proof of (134).]

Every $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$ satisfy

$$\left(AA^T\right)_{i,j} = \delta_{i,j}. \tag{135}$$

[*Proof of (135):* Let $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$.

The i-th row of A has exactly one entry equal to 1 (by Statement 2). In other words, there exists a unique $u \in \{1,2,\ldots,n\}$ satisfying $A_{i,u} = 1$. Consider this u. We have

$$A_{i,k} = \delta_{k,u} \qquad \text{for every } k \in \{1,2,\ldots,n\} \tag{136}$$

[93].

[93] *Proof of (136):* Let $k \in \{1,2,\ldots,n\}$. We must prove that $A_{i,k} = \delta_{k,u}$.

We will prove this by contradiction. Thus, assume the contrary. Hence, $A_{i,k} \neq \delta_{k,u}$.

If we had $k = u$, then we would have $A_{i,k} = A_{i,u} = 1 = \delta_{k,u}$ (since $\delta_{k,u} = 1$ (because $k = u$)), which would contradict $A_{i,k} \neq \delta_{k,u}$. Thus, we cannot have $k = u$. We therefore must have $k \neq u$. (Yes, we have just done a proof by contradiction inside a proof by contradiction. This is perfectly legitimate!)

Since $k \neq u$, we have $\delta_{k,u} = 0$. On the other hand, Statement 1 shows that each entry of A is either a 0 or a 1. In particular, $A_{i,k}$ is either a 0 or a 1 (since $A_{i,k}$ is an entry of A). Since $A_{i,k}$ is not a 0 (because $A_{i,k} \neq \delta_{k,u} = 0$), we therefore conclude that $A_{i,k}$ is a 1. In other words, $A_{i,k} = 1$.

So we know that both numbers $A_{i,k}$ and $A_{i,u}$ are equal to 1. These numbers $A_{i,k}$ and $A_{i,u}$ are two entries of the i-th row of A, and lie in different cells (since $k \neq u$). Thus, the i-th row of A has at least two entries equal to 1 (namely, the entries $A_{i,k}$ and $A_{i,u}$). This contradicts the fact that the i-th row of A has exactly one entry equal to 1. This contradiction shows that our assumption must have been false. Hence, $A_{i,k} = \delta_{k,u}$ is proven. In other words, (136) is proven.

On the other hand, the u-th column of A has exactly one entry equal to 1 (by Statement 3). We can use this to show that

$$A_{j,u} = \delta_{i,j} \tag{137}$$

[94].

But Proposition 2.28 (applied to $m = n$, $p = n$ and $B = A^T$) shows that

$$\left(AA^T\right)_{i,j} = \sum_{k=1}^{n} \underbrace{A_{i,k}}_{\substack{=\delta_{k,u} \\ \text{(by (136))}}} \underbrace{\left(A^T\right)_{k,j}}_{\substack{=A_{j,k} \\ \text{(by (133), applied} \\ \text{to } k \text{ instead of } i\text{)}}} = \sum_{k=1}^{n} \delta_{k,u} A_{j,k} = \sum_{k=1}^{n} A_{j,k} \delta_{k,u} = A_{j,u}$$

(by Proposition 2.40, applied to $p = 1$, $q = n$, $r = u$ and $a_k = A_{j,k}$). Hence, $\left(AA^T\right)_{i,j} = A_{j,u} = \delta_{i,j}$ (by (137)). This proves (135).]

On the other hand, $I_n = \left(\delta_{i,j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Hence, every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy $\left(I_n\right)_{i,j} = \delta_{i,j}$. Comparing this with (135), we find that every $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$ satisfy $\left(AA^T\right)_{i,j} = \left(I_n\right)_{i,j}$. In other words, each entry of the matrix AA^T equals the corresponding entry of I_n. Therefore, $AA^T = I_n$.

Now, let us change our point of view. We have just shown that $AA^T = I_n$. But (134) shows that A^T is a permutation matrix. Hence, the argument that we used to prove $AA^T = I_n$ can also be applied to A^T instead of A. It therefore yields $A^T \left(A^T\right)^T = I_n$. Since $\left(A^T\right)^T = A$ (indeed, this follows from Proposition 2.9), this rewrites as $A^T A = I_n$.

Now, the two equalities $AA^T = I_n$ and $A^T A = I_n$ (combined) show that the matrix A^T is an inverse of A. Hence, the matrix A is invertible, and its inverse is $A^{-1} = A^T$. This proves parts **(a)** and **(b)** of Proposition 3.122. It remains to prove part **(c)**. But this is obvious now: From (134), we know that A^T is a permutation

[94] *Proof of (137):* We will prove this by contradiction. Thus, assume the contrary. Hence, $A_{j,u} \neq \delta_{i,j}$.

If we had $i = j$, then we would have

$$\begin{aligned} A_{j,u} &= A_{i,u} && \left(\text{since } j = i\right) \\ &= 1 = \delta_{i,j} && \left(\text{since } \delta_{i,j} = 1 \text{ (because } i = j\text{)}\right), \end{aligned}$$

which would contradict $A_{j,u} \neq \delta_{i,j}$. Thus, we cannot have $i = j$. We therefore must have $i \neq j$. (Once again, we have just made a proof by contradiction within a proof by contradiction.)

Since $i \neq j$, we have $\delta_{i,j} = 0$. On the other hand, Statement 1 shows that each entry of A is either a 0 or a 1. In particular, $A_{j,u}$ is either a 0 or a 1 (since $A_{j,u}$ is an entry of A). Since $A_{j,u}$ is not a 0 (because $A_{j,u} \neq \delta_{i,j} = 0$), we therefore conclude that $A_{j,u}$ is a 1. In other words, $A_{j,u} = 1$.

So we know that both numbers $A_{i,u}$ and $A_{j,u}$ are equal to 1. These numbers $A_{i,u}$ and $A_{j,u}$ are two entries of the u-th column of A, and lie in different cells (since $i \neq j$). Thus, the u-th column of A has at least two entries equal to 1 (namely, the entries $A_{i,u}$ and $A_{j,u}$). This contradicts the fact that the u-th column of A has exactly one entry equal to 1. This contradiction shows that our assumption must have been false. Hence, $A_{j,u} = \delta_{i,j}$ is proven. In other words, (137) is proven.

matrix. Since $A^{-1} = A^T$, this rewrites as follows: A^{-1} is a permutation matrix. This proves Proposition 3.122 **(c)**. □

Let me add one more result, which somewhat extends the $\Longrightarrow$ direction of Theorem 3.116:

Proposition 3.128. Let $n \in \mathbb{N}$. A *simple swapping matrix* will mean an $n \times n$-matrix of the form $T_{k,k+1}$ for some $k \in \{1, 2, \ldots, n-1\}$.

Each permutation matrix of size $n \times n$ is a product of simple swapping matrices.

Example 3.129. Let $n = 3$. Then, there are three swapping matrices: $T_{1,2}$, $T_{1,3}$ and $T_{2,3}$. (You can also write down $T_{2,1}$, $T_{3,1}$ and $T_{3,2}$, but these are the same three matrices by different names.) Out of these three matrices, two (namely, $T_{1,2}$ and $T_{2,3}$) are simple swapping matrices (in the sense of Proposition 3.128), whereas the remaining one ($T_{1,3}$) is not.

Proposition 3.128 (applied to $n = 3$) says that each permutation matrix of size 3×3 is a product of simple swapping matrices. Let us check this for the permutation matrix $T_{1,3}$. Just writing $T_{1,3}$ as the product of itself alone does not count anymore, because (as we said) $T_{1,3}$ is not a simple swapping matrix. However, we can write $T_{1,3}$ as the product $T_{1,2}T_{2,3}T_{1,2}$, and this does the trick. (Alternatively, you can write $T_{1,3}$ as the product $T_{2,3}T_{1,2}T_{2,3}$. Again, there are many possibilities.)

Proof of Proposition 3.128 (sketched). We will not need Proposition 3.128, so let me only give a brief outline of its proof.

We want to prove that each permutation matrix of size $n \times n$ is a product of simple swapping matrices. But the $\Longrightarrow$ direction of Theorem 3.116 shows that each permutation matrix of size $n \times n$ is a product of swapping matrices. Hence, it suffices to prove that

$$\text{each swapping matrix is a product of simple swapping matrices.} \tag{138}$$

So let A be a swapping matrix. Thus, $A = T_{u,v}$ for two distinct elements u and v of $\{1, 2, \ldots, n\}$. Consider these u and v. We can WLOG assume that $u < v$ (since otherwise, we can simply switch u with v, without changing A because $T_{u,v} = T_{v,u}$). So, assume this. Then, I claim that

$$T_{u,v} = \left(T_{u,u+1}T_{u+1,u+2}\cdots T_{v-2,v-1}\right) T_{v-1,v} \left(T_{v-2,v-1}T_{v-3,v-2}\cdots T_{u,u+1}\right). \tag{139}$$

(To ease understanding, let me explain how the right hand side of (139) should be interpreted: The first parenthesized factor, $T_{u,u+1}T_{u+1,u+2}\cdots T_{v-2,v-1}$, is the product of all $T_{k,k+1}$ with k ranging over $\{u, u+1, \ldots, v-2\}$ **in increasing order**. If the set $\{u, u+1, \ldots, v-2\}$ is empty, then this product has to be understood as the empty product (which equals I_n as usual). The second parenthesized factor, $T_{v-2,v-1}T_{v-3,v-2}\cdots T_{u,u+1}$, is the product of all $T_{k,k+1}$ with k ranging

over $\{u, u+1, \ldots, v-2\}$ **in decreasing order**. Again, it is the empty product if $\{u, u+1, \ldots, v-2\}$ is empty (and thus equals I_n in this case).)

Clearly, the equality (139) represents $T_{u,v}$ as a product of simple swapping matrices; thus, once it is proven, the claim (138) will immediately follow.

One way to prove (139) is to rewrite the right hand side of (139) as

$$\left(T_{u,u+1}T_{u+1,u+2}\cdots T_{v-2,v-1}\right) T_{v-1,v} \left(T_{v-2,v-1}T_{v-3,v-2}\cdots T_{u,u+1}\right) I_n. \tag{140}$$

By using Proposition 3.105 repeatedly, we thus see that it is the matrix obtained from I_n by:

- first, swapping the u-th row with the $(u+1)$-th row,
- then, swapping the $(u+1)$-th row with the $(u+2)$-th row,
- and so on, until finally swapping the $(v-2)$-th row with the $(v-1)$-th row,
- then, swapping the $(v-1)$-th row with the v-th row,
- then, swapping the $(v-2)$-th row with the $(v-1)$-th row,
- then, swapping the $(v-3)$-th row with the $(v-2)$-th row,
- and so on, until finally swapping the u-th row with the $(u+1)$-th row.

This sequence of swappings has the following consequence: The u-th row has been moved further and further down (by repeatedly getting swapped with the next row) until it finally found rest in the position of the v-th row; then, the former v-th row has been moved further and further up (by repeatedly getting swapped with the previous row) until it finally found rest in the position of the u-th row. At the end of this process, the u-th and the v-th rows have traded places, but all the other rows have remained at the positions where they started out (although some of them were temporarily moved back and forth in the process). So the right hand side of (139) can be obtained from I_n by swapping the u-th row with the v-th row. But (according to Corollary 3.108 **(a)**) the matrix $T_{u,v}$ can be obtained in exactly the same way. Hence, the right hand side of (139) simply is $T_{u,v}$. Thus, (139) is proven. As explained, this proves Proposition 3.128.

(If you are looking for a more rigorous proof of (139) – without the swapping process so confusingly described above –, you can also proceed by induction over $v-u$. In the induction step, (139) is derived from the equality

$$T_{u+1,v} = \left(T_{u+1,u+2}T_{u+2,u+3}\cdots T_{v-2,v-1}\right) T_{v-1,v} \left(T_{v-2,v-1}T_{v-3,v-2}\cdots T_{u+1,u+2}\right),$$

which follows from the induction hypothesis. The details are, again, omitted.) $\square$

3.21. (*) Permutation matrices and permutations

In Definition 3.112, we have defined permutation matrices as matrices satisfying three particular conditions. This may not be their most natural definition, and certainly has the disadvantage of making them look like curiosities rather than like something important. In this section, I shall show a different approach to them which better demonstrates their significance. The idea is that permutation matrices are matrix representations of *permutations*, a fundamental notion of combinatorics (= the theory of finite sets).

First, let me recall some basic terminology:

- A map[95] $f : X \to Y$ between two sets X and Y is said to be *injective* if it has the following property:
  - If x_1 and x_2 are two elements of X satisfying $f(x_1) = f(x_2)$, then $x_1 = x_2$. (In words: If two elements of X are sent to one and the same element of Y by f, then these two elements of X must have been equal in the first place. In other words: An element of X is uniquely determined by its image under f.)

  Injective maps are often called "one-to-one maps" or "injections".

  For example:
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto 2x$ (this is the map that sends each integer x to $2x$) is injective, because if x_1 and x_2 are two integers satisfying $2x_1 = 2x_2$, then $x_1 = x_2$.
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x^2$ (this is the map that sends each integer x to x^2) is **not** injective, because if x_1 and x_2 are two integers satisfying $x_1^2 = x_2^2$, then we do not necessarily have $x_1 = x_2$. (For example, if $x_1 = -1$ and $x_2 = 1$, then $x_1^2 = x_2^2$ but not $x_1 = x_2$.)
- A map $f : X \to Y$ between two sets X and Y is said to be *surjective* if it has the following property:
  - For each $y \in Y$, there exists some $x \in X$ satisfying $f(x) = y$. (In words: Each element of Y is an image of some element of X under f.)

  Surjective maps are often called "onto maps" or "surjections".

  For example:
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x + 1$ (this is the map that sends each integer x to $x + 1$) is surjective, because each integer y has some integer satisfying $x + 1 = y$ (namely, $x = y - 1$).

[95] The words "map", "mapping", "function", "transformation" and "operator" are synonyms in mathematics. (That said, mathematicians often show some nuance by using one of them and not the other. However, we do not need to concern ourselves with this here.)

  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto 2x$ (this is the map that sends each integer x to $2x$) is **not** surjective, because not each integer y has some integer x satisfying $2x = y$. (For instance, $y = 1$ has no such x, since y is odd.)
  - The map $\{1,2,3,4\} \to \{1,2,3,4,5\}$, $x \mapsto x$ (this is the map sending each x to x) is **not** surjective, because not each $y \in \{1,2,3,4,5\}$ has some $x \in \{1,2,3,4\}$ satisfying $x = y$. (Namely, $y = 5$ has no such x.)

- A map $f : X \to Y$ between two sets X and Y is said to be *bijective* if it is both injective and surjective. Bijective maps are often called "one-to-one correspondences" or "bijections".

  For example:

  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x + 1$ is bijective, since it is both injective and surjective.
  - The map $\{1,2,3,4\} \to \{1,2,3,4,5\}$, $x \mapsto x$ is **not** bijective, since it is not surjective.
  - The map $\mathbb{Z} \to \mathbb{Z}$, $x \mapsto x^2$ is **not** bijective, since it is not injective. (It also is not surjective.)

- If X is a set, then id_X denotes the map from X to X that sends each $x \in X$ to x itself. (In words: id_X denotes the map which sends each element of X to itself.) The map id_X is often called the *identity map on* X, and often denoted by id (when X is clear from the context or irrelevant). The identity map id_X is always bijective.

- If $f : X \to Y$ and $g : Y \to Z$ are two maps, then the *composition* $g \circ f$ of the maps g and f is defined to be the map from X to Z that sends each $x \in X$ to $g(f(x))$. (In words: The composition $g \circ f$ is the map from X to Z that applies the map f **first** and **then** applies the map g.) You might find it confusing that this map is denoted by $g \circ f$ (rather than $f \circ g$), given that it proceeds by applying f first and g last; however, this has its reasons: It satisfies $(g \circ f)(x) = g(f(x))$. Had we denoted it by $f \circ g$ instead, this equality would instead become $(f \circ g)(x) = g(f(x))$, which would be even more confusing.

- If $f : X \to Y$ is a map between two sets X and Y, then an *inverse* of f means a map $g : Y \to X$ satisfying $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. (In words, the condition "$f \circ g = \mathrm{id}_Y$" means "if you start with some element $y \in Y$, then apply g, then apply f, then you get y back", or equivalently "the map f undoes the map g". Similarly, the condition "$g \circ f = \mathrm{id}_X$" means "if you start with some element $x \in X$, then apply f, then apply g, then you get x back", or equivalently "the map g undoes the map f". Thus, an inverse of f means a map $g : Y \to X$ that both undoes and is undone by f.)

The map $f : X \to Y$ is said to be *invertible* if and only if an inverse of f exists. If an inverse of f exists, then it is unique[96], and thus is called *the inverse of* f, and is denoted by f^{-1}.

For example:

- The map $\mathbb{Z} \to \mathbb{Z},\ x \mapsto x+1$ is invertible, and its inverse is $\mathbb{Z} \to \mathbb{Z},\ x \mapsto x-1$.
- The map $\mathbb{Q} \setminus \{1\} \to \mathbb{Q} \setminus \{0\},\ x \mapsto \dfrac{1}{1-x}$ is invertible, and its inverse is the map $\mathbb{Q} \setminus \{0\} \to \mathbb{Q} \setminus \{1\},\ x \mapsto 1 - \dfrac{1}{x}$.

A map $f : X \to Y$ is invertible if and only if it is bijective.

- If X is a set, then a *permutation* of X means a bijective (i.e., invertible) map $X \to X$.

Each permutation of the set $\{1, 2, \ldots, n\}$ (for any $n \in \mathbb{N}$) determines a permutation matrix:

Definition 3.130. Let $n \in \mathbb{N}$. Let w be a permutation of $\{1, 2, \ldots, n\}$ (that is, a bijection $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$). Then, we define an $n \times n$-matrix P_w by $P_w = \left(\delta_{w(i),j}\right)_{1 \le i \le n,\ 1 \le j \le n}$. In other words, P_w is the $n \times n$-matrix with the following entries:

- For each $i \in \{1, 2, \ldots, n\}$, the i-th row has an entry equal to 1 in the $w(i)$-th position (that is, the $(i, w(i))$-th entry of the matrix is 1).
- All other entries are 0.

Example 3.131. Let $n = 4$. There are 24 permutations of the set $\{1, 2, 3, 4\}$. One of them is the map $u : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$ that sends $1, 2, 3, 4$ to $3, 1, 4, 2$, respectively. The matrix P_u corresponding to this map u is $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$.

Proposition 3.132. Let $n \in \mathbb{N}$.
(a) If w is a permutation of $\{1, 2, \ldots, n\}$, then P_w is a permutation matrix.
(b) Let P be a permutation matrix of size $n \times n$. Then, there exists a **unique** permutation w of $\{1, 2, \ldots, n\}$ such that $P = P_w$.

[96] This is not hard to show. In fact, the situation is very similar to inverses of matrices; in particular, we can define "left inverses" and "right inverses", and prove analogues of Proposition 3.6 and Corollary 3.7 for maps instead of matrices.

Proof of Proposition 3.132 (sketched). **(a)** Let w be a permutation of $\{1,2,\ldots,n\}$. Thus, w is a bijective map, i.e., an injective and surjective map.

We want to show that P_w is a permutation matrix. According to the definition of a "permutation matrix", we therefor must prove the following three statements:

Statement 1: Each entry of P_w is either a 0 or a 1.

Statement 2: Each row of P_w has exactly one entry equal to 1.

Statement 3: Each column of P_w has exactly one entry equal to 1.

Statement 1 follows immediately from the definition of P_w.

Statement 2 follows from the definition as well: For each $i \in \{1,2,\ldots,n\}$, the i-th row of P_w has exactly one entry equal to 1 (namely, the entry in position $w(i)$).

It remains to prove Statement 3. Fix $j \in \{1,2,\ldots,n\}$. We must show that the j-th column of P_w has exactly one entry equal to 1. In other words, we must prove that there exists exactly one $i \in \{1,2,\ldots,n\}$ satisfying $w(i) = j$.

Since w is surjective, there exists **at least one** such i. Since w is injective, there exists **at most one** such i (because if i_1 and i_2 are two such i, then $w(i_1) = j$ and $w(i_2) = j$, so that $w(i_1) = w(i_2)$, and thus the injectivity of w leads to $i_1 = i_2$). Combining the previous two sentences, we conclude that there exists **exactly one** such i. Thus, Statement 3 is proven.

We have now proven all three Statements 1, 2 and 3. Hence, P_w is a permutation matrix, so that we have proven Proposition 3.132 **(a)**.

(b) We know that P is a permutation matrix. In other words, the following three statements hold:

Statement 1: Each entry of P is either a 0 or a 1.

Statement 2: Each row of P has exactly one entry equal to 1.

Statement 3: Each column of P has exactly one entry equal to 1.

Now, define a map $w : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ as follows:

Let $i \in \{1,2,\ldots,n\}$. Then, the i-th row of P has exactly one entry equal to 1 (by Statement 2). In other words, there exists exactly one $j \in \{1,2,\ldots,n\}$ such that $P_{i,j} = 1$. Define $w(i)$ to be this j.

Thus, we have defined $w(i)$ for each $i \in \{1,2,\ldots,n\}$. Hence, the map $w : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ is defined.

Described in words, the map w sends each $i \in \{1,2,\ldots,n\}$ to the position of the unique entry equal to 1 in the i-th row of P. Condition 1 shows that all the other entries are zeroes. Hence, the entries of the matrix P can be described as follows:

- For each $i \in \{1,2,\ldots,n\}$, the $(i, w(i))$-th entry of P is 1.
- All remaining entries of P are 0.

In other words, for each $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,n\}$, the (i,j)-th entry of P is $\begin{cases} 1, & \text{if } j = w(i); \\ 0, & \text{if } j \neq w(i) \end{cases} = \delta_{j,w(i)} = \delta_{w(i),j}$. Thus, the matrix P equals $\left(\delta_{w(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$.

Now, the map w is injective[97] and surjective[98]. Hence, w is bijective. Thus, w is a permutation of $\{1,2,\ldots,n\}$. The definition of P_w thus yields $P_w = \left(\delta_{w(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. Comparing this with $P = \left(\delta_{w(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$, we obtain $P = P_w$.

We thus have shown that there exists **at least one** permutation w of $\{1,2,\ldots,n\}$ such that $P = P_w$. It remains to prove that there exists **at most one** such permutation w. Fortunately, this is easy: The requirement that $P = P_w$ determines w uniquely, since the values of w can be read off the matrix P_w (namely, $w(i)$ is the position of the entry equal to 1 in the i-th row of P_w). Hence, Proposition 3.132 **(b)** is proven. □

Proposition 3.132 shows that the permutation matrices of size $n \times n$ (for any given $n \in \mathbb{N}$) are precisely the matrices of the form P_w with w being a permutation of $\{1,2,\ldots,n\}$. Moreover, it shows that each permutation matrix can be written in this form in a unique way. Hence, the map

$$\begin{aligned} \{\text{permutations of } \{1,2,\ldots,n\}\} &\to \{\text{permutation matrices of size } n \times n\}, \\ w &\mapsto P_w \end{aligned}$$

[99] is a bijection. Therefore, there are as many permutation matrices of size $n \times n$ as there are permutations of $\{1,2,\ldots,n\}$. This allows us to prove Proposition 3.114:

Proof of Proposition 3.114 (sketched). We have just seen that there are as many permutation matrices of size $n \times n$ as there are permutations of $\{1,2,\ldots,n\}$. Hence, it suffices to show that there are precisely $n!$ permutations of $\{1,2,\ldots,n\}$.

[97] *Proof.* Let i_1 and i_2 be two elements of $\{1,2,\ldots,n\}$ such that $w(i_1) = w(i_2)$. We must show that $i_1 = i_2$.

We have $P_{i_1,w(i_1)} = 1$ (by the definition of w) and $P_{i_2,w(i_2)} = 1$ (similarly). Since $w(i_1) = w(i_2)$, we have $P_{i_2,w(i_1)} = P_{i_2,w(i_2)} = 1$.

If $i_1 \neq i_2$, then the equalities $P_{i_1,w(i_1)} = 1$ and $P_{i_2,w(i_1)} = 1$ show that the $w(i_1)$-th column of P has (at least) two entries equal to 1 (namely, the entries in cells $(i_1, w(i_1))$ and $(i_2, w(i_1))$); but this flies in the face of the fact that the $w(i_1)$-th column of P has exactly one entry equal to 1 (which follows from Statement 3). Hence, we cannot have $i_1 \neq i_2$. We thus have $i_1 = i_2$. This completes the proof that w is injective.

[98] *Proof.* Let $j \in \{1,2,\ldots,n\}$. Then, the j-th column of P has exactly one entry equal to 1 (by Statement 3). In other words, there exists exactly one $i \in \{1,2,\ldots,n\}$ such that $P_{i,j} = 1$. Consider this i. Then, $P_{i,j} = 1$, so that $w(i) = j$ (by the definition of w). Hence, we have shown that, for each $j \in \{1,2,\ldots,n\}$, there exists some $i \in \{1,2,\ldots,n\}$ satisfying $w(i) = j$. In other words, the map w is surjective.

[99] By this, I mean: the map from the set of all permutations of $\{1,2,\ldots,n\}$ to the set of all permutation matrices of size $n \times n$ that sends each permutation w to the permutation matrix P_w.

How do we choose a permutation w of $\{1,2,\ldots,n\}$? Clearly, it suffices to choose each of its n values $w(1), w(2), \ldots, w(n)$ among the numbers in the set $\{1,2,\ldots,n\}$. These n values must be distinct (because w has to be a permutation, and thus injective). Also, these n values must cover all of the set $\{1,2,\ldots,n\}$ (since w has to be a permutation, and thus surjective); however, it turns out that this is already guaranteed if we choose these n values to be distinct (because n distinct values chosen among the numbers in $\{1,2,\ldots,n\}$ will **always** cover all of the set $\{1,2,\ldots,n\}$). Hence, we need to choose n distinct numbers $w(1), w(2), \ldots, w(n)$ among the numbers in the set $\{1,2,\ldots,n\}$. Here is one way to do this:

- First, choose a value for $w(1)$. There are n possible choices for this (since $w(1)$ has to belong to $\{1,2,\ldots,n\}$).
- Second, choose a value for $w(2)$. There are $n-1$ possible choices for this (since $w(2)$ has to belong to $\{1,2,\ldots,n\}$, but must not equal $w(1)$, since we want the n numbers $w(1), w(2), \ldots, w(n)$ to be distinct).
- Third, choose a value for $w(3)$. There are $n-2$ possible choices for this (since $w(3)$ has to belong to $\{1,2,\ldots,n\}$, but must not equal any of $w(1)$ and $w(2)$, since we want the n numbers $w(1), w(2), \ldots, w(n)$ to be distinct).
- Fourth, choose a value for $w(4)$. There are $n-3$ possible choices for this (since $w(4)$ has to belong to $\{1,2,\ldots,n\}$, but must not equal any of $w(1), w(2), w(3)$, since we want the n numbers $w(1), w(2), \ldots, w(n)$ to be distinct).
- And so on.
- At the last step, choose a value for $w(n)$. There are $n-(n-1)$ possible choices for this (since $w(n)$ has to belong to $\{1,2,\ldots,n\}$, but must not equal any of $w(1), w(2), \ldots, w(n-1)$, since we want the n numbers $w(1), w(2), \ldots, w(n)$ to be distinct).

We thus get a total of $n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \cdots \cdot (n-(n-1))$ possible choices. In other words, we get a total of $n!$ possible choices[100]. Hence, there are precisely $n!$ permutations of $\{1,2,\ldots,n\}$. As we have said, this proves Proposition 3.114. □

Here is a further property of permutation matrices constructed out of permutations:

[100] since

$$n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \cdots \cdot (n-(n-1)) = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \cdots \cdot 1 = 1 \cdot 2 \cdot \cdots \cdot n = n!$$

Proposition 3.133. Let $n \in \mathbb{N}$.

(a) If x and y are two permutations of $\{1,2,\ldots,n\}$, then $P_x P_y = P_{y \circ x}$.

(b) If w is a permutation of $\{1,2,\ldots,n\}$, then $(P_w)^{-1} = P_{w^{-1}}$.

(c) We have $P_{\mathrm{id}_{\{1,2,\ldots,n\}}} = I_n$.

(d) Let u and v be two distinct elements of $\{1,2,\ldots,n\}$. Let $\tau_{u,v}$ be the permutation of $\{1,2,\ldots,n\}$ which sends u to v, sends v to u, and sends each other element of $\{1,2,\ldots,n\}$ to itself. (For example, if $n = 7$, then $\tau_{2,5}$ sends $1,2,3,4,5,6,7$ to $1,5,3,4,2,6,7$, respectively.) Then, $P_{\tau_{u,v}} = T_{u,v}$.

We leave the easy proof to the reader.

As we have said, Proposition 3.132 shows that the map

$$\{\text{permutations of } \{1,2,\ldots,n\}\} \to \{\text{permutation matrices of size } n \times n\}, \quad w \mapsto P_w$$

is a bijection. This map provides a "dictionary" between permutations of $\{1,2,\ldots,n\}$ and permutation matrices of size $n \times n$ (so to speak). Proposition 3.133 **(a)** shows that (under this "dictionary") composition of permutations "corresponds" to multiplication of permutation matrices (except that the order of the factors is reversed). Proposition 3.133 **(b)** shows that the inverse of a permutation gets "translated" into the inverse of the corresponding permutation matrix. Proposition 3.133 **(c)** shows that the identity permutation $\mathrm{id}_{\{1,2,\ldots,n\}}$ "corresponds" to the identity matrix I_n. Finally, Proposition 3.133 **(d)** shows that the permutations $\tau_{u,v}$ "correspond" to the swapping matrices $T_{u,v}$. Using this "dictionary", we can translate theorems about permutations into theorems about permutation matrices, and vice versa. For instance, the translation of Proposition 3.128 into the language of permutations is the following fact:

Proposition 3.134. Let $n \in \mathbb{N}$. A *simple transposition* will mean a permutation in S_n that has the form $\tau_{k,k+1}$ for some $k \in \{1,2,\ldots,n-1\}$ (where $\tau_{k,k+1}$ is defined as in Proposition 3.133 **(d)**).

Each permutation of $\{1,2,\ldots,n\}$ is a product of simple transpositions.

Arguably, in this translated form, this proposition is rather obvious. It merely says that any rearrangement of the numbers $1,2,\ldots,n$ can be obtained from the list $(1,2,\ldots,n)$ by repeatedly swapping adjacent entries. This should be quite clear, at least intuitively (the "bubble sort" algorithm provides a way to obtain the list $(1,2,\ldots,n)$ from our rearrangement by repeatedly swapping adjacent entries; now, all that remains to be done is to perform these swaps backwards in order to get from $(1,2,\ldots,n)$ to the rearrangement).

Next, let us prove Proposition 3.123. Let us first state a more precise version of this proposition:

Proposition 3.135. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let w be a permutation of $\{1,2,\ldots,n\}$. Let P be the $n \times n$-matrix P_w. Let C be an $n \times m$-matrix. Then,

$$\operatorname{row}_i (PC) = \operatorname{row}_{w(i)} C \qquad \text{for every } i \in \{1,2,\ldots,n\}.$$

Proof of Proposition 3.135. We have $P = P_w = \left(\delta_{w(i),j}\right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$ (by the definition of P_w). Thus,

$$P_{i,j} = \delta_{w(i),j} \qquad \text{for all } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,n\}. \tag{141}$$

Now, Proposition 2.28 (applied to n, m, P and C instead of m, p, A and B) shows that

$$(PC)_{i,j} = \sum_{k=1}^{n} P_{i,k} C_{k,j} \qquad \text{for all } i \in \{1,2,\ldots,n\} \text{ and } j \in \{1,2,\ldots,m\}.$$

Hence, for all $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,m\}$, we have

$$(PC)_{i,j} = \sum_{k=1}^{n} \underbrace{P_{i,k}}_{\substack{=\delta_{w(i),k} \\ \text{(by (141), applied} \\ \text{to } k \text{ instead of } j)}} C_{k,j} = \sum_{k=1}^{n} \underbrace{\delta_{w(i),k} C_{k,j}}_{=C_{k,j}\delta_{w(i),k}} = \sum_{k=1}^{n} C_{k,j} \underbrace{\delta_{w(i),k}}_{\substack{=\delta_{k,w(i)} \\ \text{(since } \delta_{u,v}=\delta_{v,u} \\ \text{for any two} \\ \text{objects } u \text{ and } v)}}$$

$$= \sum_{k=1}^{n} C_{k,j} \delta_{k,w(i)} = C_{w(i),j} \tag{142}$$

(by Proposition 2.40, applied to $p = 1$, $q = n$, $r = w(i)$ and $a_k = C_{k,j}$).

Now, let $i \in \{1,2,\ldots,n\}$. Then, the definition of $\operatorname{row}_{w(i)} C$ yields $\operatorname{row}_{w(i)} C = \left(C_{w(i),y}\right)_{1 \leq x \leq 1,\ 1 \leq y \leq m}$. But the definition of $\operatorname{row}_i (PC)$ yields

$$\operatorname{row}_i (PC) = \left(\underbrace{(PC)_{i,y}}_{\substack{=C_{w(i),y} \\ \text{(by (142), applied to } j=y)}} \right)_{1 \leq x \leq 1,\ 1 \leq y \leq m} = \left(C_{w(i),y}\right)_{1 \leq x \leq 1,\ 1 \leq y \leq m}.$$

Comparing this with $\operatorname{row}_{w(i)} C = \left(C_{w(i),y}\right)_{1 \leq x \leq 1,\ 1 \leq y \leq m}$, we obtain $\operatorname{row}_i (PC) = \operatorname{row}_{w(i)} C$. This proves Proposition 3.135. □

Proof of Proposition 3.123. Proposition 3.132 **(b)** shows that there exists a **unique** permutation w of $\{1,2,\ldots,n\}$ such that $P = P_w$. Consider this w. Proposition 3.135 yields that

$$\operatorname{row}_i (PC) = \operatorname{row}_{w(i)} C \qquad \text{for every } i \in \{1,2,\ldots,n\}.$$

In other words, the i-th row of PC equals the $w(i)$-th row of C for each $i \in \{1,2,\ldots,n\}$. Since w is a permutation of $\{1,2,\ldots,n\}$, this shows that the rows of PC are the rows of C, rearranged. Hence, the matrix PC can be obtained from C by rearranging the rows. This proves Proposition 3.123. □

3.22. The standard row operations

I shall now introduce the standard row operations – certain transformations acting on matrices, changing some of their rows while leaving others unchanged. We have already encountered them in some proofs above (for instance, the "downward row additions" in the proof of Theorem 3.63 were one type of row operations), but now we shall give them the systematic treatment they deserve and see them collaborate on producing in Gaussian elimination.

Definition 3.136. For the rest of Chapter 3, we shall use the word "transformation" in a rather specific meaning: Let $n \in \mathbb{N}$. A *transformation of n-rowed matrices* will mean a map

$$\{\text{matrices with } n \text{ rows}\} \to \{\text{matrices with } n \text{ rows}\}.$$

In other words, a *transformation of n-rowed matrices* is a map that transforms each matrix with n rows into a new matrix with n rows.

We shall use the arrow notation for transformations: If $\mathcal{O}$ is some transformation of n-rowed matrices, and if C and D are two matrices, then we will use the notation "$C \xrightarrow{\mathcal{O}} D$" when we want to say that the transformation $\mathcal{O}$ transforms C into D. For example, if $\mathcal{M}$ denotes the transformation of 2-rowed matrices that multiplies each entry of a matrix by 3, then $\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix} \xrightarrow{\mathcal{M}} \begin{pmatrix} 3a & 3b & 3c \\ 3a' & 3b' & 3c' \end{pmatrix}$.

More generally, if $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_m$ is a sequence of transformations of n-rowed matrices, and if $C_0, C_1, \ldots, C_m$ are some matrices, then we will use the notation "$C_0 \xrightarrow{\mathcal{O}_1} C_1 \xrightarrow{\mathcal{O}_2} \cdots \xrightarrow{\mathcal{O}_m} C_m$" when we want to say that the transposition $\mathcal{O}_i$ transforms C_{i-1} into C_i for each $i \in \{1,2,\ldots,m\}$. Examples for this will appear below once we have defined some actual transformations.

We will also sometimes draw arrows in two directions. Namely: If $\mathcal{O}$ and $\mathcal{P}$ are two transformations of n-rowed matrices, and if C and D are two matrices, then we will use the notation "$C \underset{\mathcal{P}}{\overset{\mathcal{O}}{\rightleftarrows}} D$" when we want to say that the transformation $\mathcal{O}$ transforms C into D while the transformation $\mathcal{P}$ transforms D into C. This will often happen when two transformations $\mathcal{O}$ and $\mathcal{P}$ are inverse to each other (i.e., each of them undoes the other).

Definition 3.137. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1,2,\ldots,n\}$. Let λ be a number.

Consider the transformation which transforms an $n \times m$-matrix C (for some $m \in \mathbb{N}$) into the product $A_{u,v}^{\lambda} C$, where $A_{u,v}^{\lambda}$ is the λ-addition matrix as defined in Definition 3.53. It is the transformation that modifies an $n \times m$-matrix C by adding $\lambda \operatorname{row}_v C$ to the u-th row (according to Proposition 3.56). This transformation will be called *the row addition* $A_{u,v}^{\lambda}$. (We are using the same symbol $A_{u,v}^{\lambda}$ for the transformation and for the λ-addition matrix, since they are so closely related; nevertheless, they are not one and the same thing. I hope that the reader will be able to keep them apart.)

The row addition $A_{u,v}^{\lambda}$ is called a *downward row addition* if $u > v$, and is called an *upward row addition* if $u < v$.

Note that two matrices C and D (with n rows each) satisfy $C \xrightarrow{A_{u,v}^{\lambda}} D$ if and only if they satisfy $D = A_{u,v}^{\lambda} C$. (This is because the row addition $A_{u,v}^{\lambda}$ transforms each $n \times m$-matrix C into $A_{u,v}^{\lambda} C$.)

Note that the row addition $A_{u,v}^{\lambda}$ is inverse to the row addition $A_{u,v}^{-\lambda}$ (since the matrices $A_{u,v}^{\lambda}$ and $A_{u,v}^{-\lambda}$ are inverse). Hence, if two matrices C and D (with n rows each) satisfy $C \xrightarrow{A_{u,v}^{\lambda}} D$, then they satisfy $C \underset{A_{u,v}^{-\lambda}}{\overset{A_{u,v}^{\lambda}}{\rightleftarrows}} D$.

Example 3.138. The row addition $A_{1,3}^{5}$ modifies a $3 \times m$-matrix C by adding $5 \operatorname{row}_3 C$ to the 1-st row of C. Thus, this row addition transforms any 3×4-matrix $\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix}$ into $\begin{pmatrix} a_1 + 5a_3 & b_1 + 5b_3 & c_1 + 5c_3 & d_1 + 5d_3 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix}$. Using the arrow notation, we can write this as follows:

$$\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix} \xrightarrow{A_{1,3}^{5}} \begin{pmatrix} a_1 + 5a_3 & b_1 + 5b_3 & c_1 + 5c_3 & d_1 + 5d_3 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix}.$$

For example, $\begin{pmatrix} 1 & 3 & 2 & 1 \\ 0 & 4 & -1 & 1 \\ 2 & -1 & 3 & 1 \end{pmatrix} \xrightarrow{A_{1,3}^{5}} \begin{pmatrix} 11 & -2 & 17 & 6 \\ 0 & 4 & -1 & 1 \\ 2 & -1 & 3 & 1 \end{pmatrix}$.

Definition 3.139. Let $n \in \mathbb{N}$. Let $u \in \{1, 2, \ldots, n\}$. Let λ be a nonzero number.

Consider the transformation which transforms an $n \times m$-matrix C (for some $m \in \mathbb{N}$) into the product $S_u^{\lambda} C$, where S_u^{λ} is the λ-scaling matrix as defined in Definition 3.85. It is the transformation that modifies an $n \times m$-matrix C by scaling the u-th row by λ (according to Proposition 3.88). This transformation will be called *the row scaling* S_u^{λ}. (We are using the same symbol S_u^{λ} for the transformation and for the λ-scaling matrix. Again, this should not lead to confusion.)

Note that two matrices C and D (with n rows each) satisfy $C \xrightarrow{S_u^{\lambda}} D$ if and only if they satisfy $D = S_u^{\lambda} C$. (This is because the row scaling S_u^{λ} transforms each

$n \times m$-matrix C into $S_u^\lambda C$.)

Note that the row scaling S_u^λ is inverse to the row scaling $S_u^{1/\lambda}$ (since the matrices S_u^λ and $S_u^{1/\lambda}$ are inverse). Hence, if two matrices C and D (with n rows each) satisfy $C \xrightarrow{S_u^\lambda} D$, then they satisfy $C \underset{S_u^{1/\lambda}}{\overset{S_u^\lambda}{\longrightarrow\!\!\!\!\!\leftarrow}} D$.

Example 3.140. The row scaling S_2^5 modifies a $3 \times m$-matrix C by scaling the 2-nd row by 5. Thus, this row scaling transforms any 3×2-matrix $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}$ into $\begin{pmatrix} a_1 & b_1 \\ 5a_2 & 5b_2 \\ a_3 & b_3 \end{pmatrix}$.

Definition 3.141. Let $n \in \mathbb{N}$. Let u and v be two distinct elements of $\{1, 2, \ldots, n\}$.

Consider the transformation which transforms an $n \times m$-matrix C (for some $m \in \mathbb{N}$) into the product $T_{u,v}C$, where $T_{u,v}$ is the swapping matrix as defined in Definition 3.102. It is the transformation that modifies an $n \times m$-matrix C by swapping the u-th row with the v-th row (according to Proposition 3.105). This transformation will be called *the row swap* $T_{u,v}$. (We are using the same symbol $T_{u,v}$ for the transformation and for the swapping matrix.)

Note that two matrices C and D (with n rows each) satisfy $C \xrightarrow{T_{u,v}} D$ if and only if they satisfy $D = T_{u,v}C$. (This is because the row swap $T_{u,v}$ transforms each $n \times m$-matrix C into $T_{u,v}C$.)

Note that the row swap $T_{u,v}$ is inverse to itself (since the matrices $T_{u,v}$ and $T_{u,v}$ are inverse). Hence, if two matrices C and D (with n rows each) satisfy $C \xrightarrow{T_{u,v}} D$, then they satisfy $C \underset{T_{u,v}}{\overset{T_{u,v}}{\longrightarrow\!\!\!\!\!\leftarrow}} D$.

Example 3.142. The row swap $T_{1,3}$ modifies a $4 \times m$-matrix C by swapping the 1-st row of C with the 3-rd row of C. Thus, this row swap transforms any 4×2-matrix $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{pmatrix}$ into $\begin{pmatrix} a_3 & b_3 \\ a_2 & b_2 \\ a_1 & b_1 \\ a_4 & b_4 \end{pmatrix}$. Using the arrow notation, we can write this as follows:

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \\ a_4 & b_4 \end{pmatrix} \xrightarrow{T_{1,3}} \begin{pmatrix} a_3 & b_3 \\ a_2 & b_2 \\ a_1 & b_1 \\ a_4 & b_4 \end{pmatrix}.$$

For example, $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{pmatrix} \xrightarrow{T_{1,3}} \begin{pmatrix} 5 & 6 \\ 3 & 4 \\ 1 & 2 \\ 7 & 8 \end{pmatrix}$.

Definition 3.143. Let $n \in \mathbb{N}$. The *standard row operations* (on matrices with n rows) are:

- row additions (i.e., transformations of the form $A^{\lambda}_{u,v}$);
- row scalings (i.e., transformations of the form S^{λ}_{u});
- row swaps (i.e., transformations of the form $T_{u,v}$).

These row operations allow us to "reduce" any matrix to a certain simple form – called a row echelon form – that has many zeroes (in a sense, it is close to upper-triangular, although the two notions are not exactly the same) and that allows for easily solving linear systems.

3.23. Row-echelon matrices

Definition 3.144. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. An $n \times m$-matrix A is said to be *zero* if and only if $A = 0_{n \times m}$. In other words, an $n \times m$-matrix A is said to be *zero* if and only if all entries of A are zero.

An $n \times m$-matrix A is said to be *nonzero* if A is not zero. (This does not mean that all entries of A are nonzero!)

Definition 3.145. Let $m \in \mathbb{N}$. Let $v = (v_1, v_2, \ldots, v_m)$ be a $1 \times m$-matrix (i.e., a row vector of size m). If v is nonzero (i.e., if not all entries of v are zero), then the smallest $i \in \{1, 2, \ldots, m\}$ satisfying $v_i \neq 0$ is called the **pivot index** of v and will be denoted by $\operatorname{pivind} v$. Furthermore, the value of v_i for this smallest i is called the **pivot entry** of v. (Thus, the pivot entry of v is the first nonzero entry of v.)

Example 3.146. The row vector $(0, 6, 7, 0, 9)$ has pivot index 2 (because if we write it as $(v_1, v_2, v_3, v_4, v_5)$, then the smallest $i \in \{1, 2, 3, 4, 5\}$ satisfying $v_i \neq 0$ is 2) and pivot entry 6.

The row vector $(0, 0, -1, -1, 0)$ has pivot index 3 and pivot entry -1.

The row vector $(0, 0, 0, 0, 0)$ has no pivot index (since it is zero).

Definition 3.147. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.

(a) Let $i \in \{1, 2, \ldots, n\}$ be such that $\operatorname{row}_i A$ is nonzero. Then, the cell $(i, \operatorname{pivind}(\operatorname{row}_i A))$ is called the *pivot cell* in the i-th row of A. (In words, this is the leftmost cell in the i-th row of A that has a nonzero entry.)

(b) The *pivot cells* of A are the pivot cells in the i-th row of A, where i ranges over all elements of $\{1,2,\ldots,n\}$ for which $\operatorname{row}_i A$ is nonzero. The *pivot entries* of A are the entries of A in the pivot cells. (In other words, the pivot entries of A are the leftmost nonzero entries of all nonzero rows of A.)

Example 3.148. The pivot cells of the matrix $\begin{pmatrix} 7 & 3 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & -1 & 2 \end{pmatrix}$ are $(1,1)$, $(3,3)$ and $(4,2)$, and the respective pivot entries are 7, 4 and -1. There is no pivot cell in the 2-nd row, since this row is zero.

Definition 3.149. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix.

We say that A is a *row-echelon matrix* if for each $i \in \{1,2,\ldots,n-1\}$, the following holds:

- **either** $\operatorname{row}_{i+1} A$ is zero,
- **or** both rows $\operatorname{row}_i A$ and $\operatorname{row}_{i+1} A$ are nonzero, and satisfy $\operatorname{pivind}(\operatorname{row}_i A) < \operatorname{pivind}(\operatorname{row}_{i+1} A)$ (that is, the pivot cell in the i-th row of A is strictly further left than the pivot cell in the $(i+1)$-th row of A).

Instead of saying that "A is a row-echelon matrix", it is customary to say that "A is in row-echelon form".

Example 3.150. (a) Let A be the 3×4-matrix $\begin{pmatrix} -2 & 3 & 4 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$. Then, A is a row-echelon matrix. In fact, all rows of A are nonzero, and each $i \in \{1,2\}$ satisfies $\operatorname{pivind}(\operatorname{row}_i A) < \operatorname{pivind}(\operatorname{row}_{i+1} A)$ (indeed, we have $\operatorname{pivind}(\operatorname{row}_1 A) = 1$, $\operatorname{pivind}(\operatorname{row}_2 A) = 3$ and $\operatorname{pivind}(\operatorname{row}_3 A) = 4$). The pivot cells of A are $(1,1)$, $(2,3)$ and $(3,4)$.

(b) Let A be the 3×4-matrix $\begin{pmatrix} -2 & 3 & 4 & 5 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 4 & 3 \end{pmatrix}$ instead. Then, A is not a row-echelon matrix. Indeed, $i = 2$ fails to satisfy the condition in Definition 3.149 (because $\operatorname{pivind}(\operatorname{row}_2 A) = 3$ and $\operatorname{pivind}(\operatorname{row}_3 A) = 3$, but $\operatorname{row}_3 A$ is nonzero). The pivot cells of A are $(1,1)$, $(2,3)$ and $(3,3)$.

(c) Let A be the 5×4-matrix $\begin{pmatrix} 7 & 3 & 1 & 0 \\ 0 & -1 & 2 & 2 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ instead. Then, A is a row-echelon matrix. In fact, each $i \in \{1,2\}$ satisfies $\operatorname{pivind}(\operatorname{row}_i A) <$

$\operatorname{pivind}(\operatorname{row}_{i+1} A)$ (because $\operatorname{pivind}(\operatorname{row}_1 A) = 1$, $\operatorname{pivind}(\operatorname{row}_2 A) = 2$ and $\operatorname{pivind}(\operatorname{row}_3 A) = 3$), whereas for each $i \in \{3,4\}$, the vector $\operatorname{row}_{i+1} A$ is zero (since $\operatorname{row}_4 A$ and $\operatorname{row}_5 A$ is zero). The pivot cells of A are $(1,1)$, $(2,2)$ and $(3,3)$.

(d) Let A be the 3×4-matrix $\begin{pmatrix} 7 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}$ instead. Then, A is not a row-echelon matrix. Indeed, $i = 2$ fails to satisfy the condition in the definition of 3.149 (because $\operatorname{row}_i A$ is zero, but $\operatorname{row}_{i+1} A$ is nonzero). The pivot cells of A are $(1,1)$ and $(3,3)$.

(e) Let A be an invertibly upper-triangular $n \times n$-matrix (for some $n \in \mathbb{N}$) instead. (See Definition 3.30 **(b)** for the meaning of "invertibly upper-triangular".) Then, A is a row-echelon matrix. In fact, each $i \in \{1,2,\ldots,n-1\}$ satisfies $\operatorname{pivind}(\operatorname{row}_i A) < \operatorname{pivind}(\operatorname{row}_{i+1} A)$ (because $\operatorname{pivind}(\operatorname{row}_k A) = k$ for each $k \in \{1,2,\ldots,n\}$). The pivot cells of A are $(1,1), (2,2), \ldots, (n,n)$.

(f) On the other hand, not every upper-triangular $n \times n$-matrix is a row-echelon matrix. For instance, $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ is not a row-echelon matrix, since the pivot indices of its 2-nd and 3-rd rows are equal.

From the examples just given, it shouldn't be hard to build an intuition for row-echelon matrices. Roughly speaking, a matrix A is row-echelon if and only if

- a nonzero row of A cannot follow a zero row of A, and
- the first nonzero entry in each nonzero row of A is strictly further right than that in the row before.

Here are some first simple observations:

Proposition 3.151. Let A be an $n \times m$-matrix. Assume that A is a row-echelon matrix. Let k be the number of nonzero rows of A.

(a) The first k rows of A are nonzero, whereas the last $n-k$ rows of A are zero.

(b) The pivot cells of A are the k cells

$$(1, \operatorname{pivind}(\operatorname{row}_1 A)), (2, \operatorname{pivind}(\operatorname{row}_2 A)), \ldots, (k, \operatorname{pivind}(\operatorname{row}_k A)).$$

(c) We have

$$\operatorname{pivind}(\operatorname{row}_1 A) < \operatorname{pivind}(\operatorname{row}_2 A) < \cdots < \operatorname{pivind}(\operatorname{row}_k A).$$

Proof of Proposition 3.151. The matrix A is a row-echelon matrix. Thus, the condition in Definition 3.149 shows that, for each $i \in \{1,2,\ldots,n-1\}$, the following holds:

$$\text{if } \operatorname{row}_{i+1} A \text{ is nonzero, then } \operatorname{row}_i A \text{ is nonzero.} \tag{143}$$

Hence, a nonzero row of A cannot follow a zero row. Using this fact, it is easy to see that the nonzero rows of A are "concentrated at the top", i.e., there exists some $\ell \in \{0, 1, \ldots, n\}$ such that the first ℓ rows of A are nonzero whereas the last $n - \ell$ rows of A are zero. Consider this ℓ. (Note that $\ell = 0$ if all rows of A are zero, and $\ell = n$ if all rows of A are nonzero.) Clearly, the number of nonzero rows of A is ℓ. But we have denoted this number by k. Thus, $\ell = k$. So we conclude that the first k rows of A are nonzero whereas the last $n - k$ rows of A are zero (since the first ℓ rows of A are nonzero whereas the last $n - \ell$ rows of A are zero). This proves Proposition 3.151 **(a)**.[101]

(c) Proposition 3.151 **(a)** shows that the first k rows of A are nonzero. In other words, $\operatorname{row}_{i+1} A$ is nonzero for each $i \in \{0, 1, \ldots, k-1\}$. The condition in Definition 3.149 thus shows that each $i \in \{1, 2, \ldots, k-1\}$ satisfies $\operatorname{pivind}(\operatorname{row}_i A) < \operatorname{pivind}(\operatorname{row}_{i+1} A)$ (since A is a row echelon matrix). In other words,

$$\operatorname{pivind}(\operatorname{row}_1 A) < \operatorname{pivind}(\operatorname{row}_2 A) < \cdots < \operatorname{pivind}(\operatorname{row}_k A). \tag{144}$$

This proves Proposition 3.151 **(c)**.

(b) Proposition 3.151 **(a)** shows that the first k rows of A are nonzero whereas the last $n - k$ rows of A are zero. Thus, the pivot cells of A are the pivot cells in the first k rows of A. In other words, they are the k cells

$$(1, \operatorname{pivind}(\operatorname{row}_1 A)), (2, \operatorname{pivind}(\operatorname{row}_2 A)), \ldots, (k, \operatorname{pivind}(\operatorname{row}_k A)).$$

This proves Proposition 3.151 **(b)**. □

Proposition 3.152. Let A be a matrix.

(a) Each row of A has at most one pivot cell.

(b) Assume that A is a row-echelon matrix. Each column of A has at most one pivot cell.

[101] If you find this proof insufficiently rigorous, you can argue as follows instead: If all rows of A are zero, then $k = 0$ (since k is the number of nonzero rows of A), and therefore Proposition 3.151 **(a)** is obvious in this case. Hence, we WLOG assume that $k \neq 0$. Thus, there exists at least one nonzero row of A. In other words, there exists at least one $i \in \{1, 2, \ldots, n\}$ such that $\operatorname{row}_i A$ is nonzero. Let h be the **largest** such i. Thus, $\operatorname{row}_h A$ is nonzero, but all the rows $\operatorname{row}_{h+1} A, \operatorname{row}_{h+2} A, \ldots, \operatorname{row}_n A$ are zero.

We know that $\operatorname{row}_h A$ is nonzero. Hence, by applying (143) repeatedly, we can conclude that all the h rows $\operatorname{row}_h A, \operatorname{row}_{h-1} A, \operatorname{row}_{h-2} A, \ldots, \operatorname{row}_1 A$ are nonzero. (Strictly speaking, we are saying that $\operatorname{row}_{h-j} A$ is nonzero for each $j \in \{0, 1, \ldots, h-1\}$; this can be proven by induction over j.) In other words, all the h rows $\operatorname{row}_1 A, \operatorname{row}_2 A, \ldots, \operatorname{row}_h A$ are nonzero (since their order does not matter for us). In other words, the first h rows of A are nonzero (since the first h rows of A are $\operatorname{row}_1 A, \operatorname{row}_2 A, \ldots, \operatorname{row}_h A$).

But recall that $\operatorname{row}_{h+1} A, \operatorname{row}_{h+2} A, \ldots, \operatorname{row}_n A$ are zero. In other words, the last $n - h$ rows of A are zero (since the last $n - h$ rows of A are $\operatorname{row}_{h+1} A, \operatorname{row}_{h+2} A, \ldots, \operatorname{row}_n A$ are zero).

Thus, we know that the first h rows of A are nonzero, whereas the last $n - h$ rows of A are zero. Hence, the number of nonzero rows of A is h. But the number of nonzero rows of A is k (since this is how we defined k). Comparing the previous two sentences, we obtain $h = k$.

But recall that the first h rows of A are nonzero, whereas the last $n - h$ rows of A are zero. Since $h = k$, this rewrites sa follows: The first k rows of A are nonzero, whereas the last $n - k$ rows of A are zero. This proves Proposition 3.151 **(a)** again.

Proof of Proposition 3.152. **(a)** This is clear from the definition of a pivot cell. (More precisely, each nonzero row of A has exactly one pivot cell, whereas a nonzero row of A has no pivot cell.)

(b) Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that A is an $n \times m$-matrix. Let k be the number of nonzero rows of A. Proposition 3.151 **(c)** yields

$$\operatorname{pivind}\left(\operatorname{row}_1 A\right) < \operatorname{pivind}\left(\operatorname{row}_2 A\right) < \cdots < \operatorname{pivind}\left(\operatorname{row}_k A\right). \tag{145}$$

Proposition 3.151 **(b)** shows that the pivot cells of A are the k cells

$$\left(1, \operatorname{pivind}\left(\operatorname{row}_1 A\right)\right), \left(2, \operatorname{pivind}\left(\operatorname{row}_2 A\right)\right), \ldots, \left(k, \operatorname{pivind}\left(\operatorname{row}_k A\right)\right).$$

Therefore, (145) shows that each of these cells lies strictly further left than the next. Therefore, these cells lie in distinct columns. In other words, each column of A has at most one pivot cell. This proves Proposition 3.152 **(b)**. □

The main importance of row-echelon matrices is that when A is a row-echelon matrix, the equation $Ax = b$ can be solved by a simple method called "back-substitution". Before explaining this method in full generality, let me give four examples (in increasing order of complexity):[102]

Example 3.153. Let $A = \begin{pmatrix} 4 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 5 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. We want to solve the equation $Ax = b$; in other words, we want to find all column vectors x of size 3 that satisfy $Ax = b$.

Write the unknown vector x in the form $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. Multiplying the equalities $A = \begin{pmatrix} 4 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 5 \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, we obtain $Ax = \begin{pmatrix} 4 & 3 & 4 \\ 0 & -1 & 2 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4x_1 + 3x_2 + 4x_3 \\ -x_2 + 2x_3 \\ 5x_3 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. Hence, the equation $Ax = b$ rewrites as $\begin{pmatrix} 4x_1 + 3x_2 + 4x_3 \\ -x_2 + 2x_3 \\ 5x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. This, in turn, is

[102] Notice that in each of the following examples, A is a row-echelon matrix. However, the pivot cells differ:

- In Example 3.153, each row has a pivot cell, and each column has a pivot cell.
- In Example 3.154, each row has a pivot cell, but not each column does.
- In Example 3.155 and in Example 3.156, not each row has a pivot cell. These two examples differ in b (specifically, in the entry of b in the row in which A has no pivot cell).

equivalent to the following system of linear equations:

$$\begin{cases} 4x_1 + 3x_2 + 4x_3 = 3; \\ -x_2 + 2x_3 = 1; \\ 5x_3 = 5 \end{cases}. \tag{146}$$

The system (146) has a particularly simple form: each of its equations can be solved for some variable as long as we resolve them in the appropriate order. Namely:

1. We can solve the third equation in (146) for x_3, yielding the solution $x_3 = 1$.
2. Now that we have found the value of x_3, we can substitute it into the second equation in (146), so that the latter equation becomes $-x_2 + 2 \cdot 1 = 1$. We can now solve this equation for x_2, obtaining $x_2 = 1$.
3. Now that we have found the values of x_2 and x_3, we can substitute them into the first equation in (146), so that the latter equation becomes $4x_1 + 3 \cdot 1 + 4 \cdot 1 = 3$. We can now solve this equation for x_1, obtaining $x_1 = -1$.

Altogether, we have now found $x_1 = -1$, $x_2 = 1$ and $x_3 = 1$. In other words, $x = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$. Thus, we have shown that each column vector x satisfying $Ax = b$ must equal to $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

Conversely, the vector $x = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ actually **does** satisfy $Ax = b$; this is because we did not "throw away any requirements":

1. When we solved the third equation in (146) for x_3, we ensured that our solution $x_3 = 1$ actually does satisfy this third equation.
2. When we rewrote the second equation in (146) as $-x_2 + 2 \cdot 1 = 1$ and solved it for x_2, we ensured that our solution $x_2 = 1$ actually does satisfy the second equation, as long as $x_3 = 1$.
3. When we rewrote the first equation in (146) as $4x_1 + 3 \cdot 1 + 4 \cdot 1 = 3$ and solved it for x_1, we ensured that our solution $x_1 = -1$ actually does satisfy the second equation, as long as $x_2 = 1$ and $x_3 = 1$.

Looking back at these steps, we thus conclude that the values $x_1 = -1$, $x_2 = 1$ and $x_3 = 1$ (that is, $x = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$) do satisfy all the three equations in (146). (Of

course, we could also have checked this by computing Ax for these values; but it is important to know that this computation is unnecessary – the method we used comes with a guarantee that its result is correct!)

We can now conclude that the equation $Ax = b$ has a unique solution, namely $x = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$.

Example 3.154. Let $A = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 5 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. We want to solve the equation $Ax = b$; in other words, we want to find all column vectors x of size 4 that satisfy $Ax = b$.

Write the unknown vector x in the form $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$. Then, $Ax = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 5x_4 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. Hence, the equation $Ax = b$ rewrites as $\begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 5x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. This, in turn, is equivalent to the following system of linear equations:

$$\begin{cases} 4x_1 + 3x_2 + 4x_4 = 3; \\ -x_2 + 3x_3 + 2x_4 = 1; \\ 5x_4 = 5 \end{cases} . \tag{147}$$

The system (147) can be solved similarly to (146), with one more complication:

1. We can solve the third equation in (147) for x_4, yielding the solution $x_4 = 1$.
2. There is no equation (147) that "begins with x_3" (i.e., that has x_3 appear with nonzero coefficient, but x_1 and x_2 with zero coefficients). Thus, we have nothing to solve for x_3. (Of course, we could try solving the second equation in (147) for x_3; but this would require a value of x_2, which for now is just as unknown as x_3.)

   Instead of solving an equation for x_3, let us treat x_3 as a known! In other words, let us define a number r by $r = x_3$, and pretend that $x_3 = r$ is an explicit value for x_3.
3. Now that we have "found" the values of x_3 and x_4 (the word "found" is in scare quotes because all we know about x_3 is that $x_3 = r$), we can substitute

them into the second equation in (147), so that the latter equation becomes $-x_2 + 3r + 2 \cdot 1 = 1$. We can now solve this equation for x_2, obtaining $x_2 = 3r + 1$.

4. Now that we have "found" the values of x_2, x_3 and x_4, we can substitute them into the first equation in (147), so that the latter equation becomes $4x_1 + 3(3r+1) + 4 \cdot 1 = 3$. We can now solve this equation for x_1, obtaining $x_1 = -\frac{9}{4}r - 1$.

Altogether, we have now found $x_1 = -\frac{9}{4}r - 1$, $x_2 = 3r + 1$, $x_3 = r$ and $x_4 = 1$. In other words,

$$x = \begin{pmatrix} -\frac{9}{4}r - 1 \\ 3r + 1 \\ r \\ 1 \end{pmatrix}. \tag{148}$$

Thus, we have found that each column vector x satisfying $Ax = b$ must have the form (148) for **some** number $r \in \mathbb{R}$.

Conversely, any column vector x of the form (148) (for **any** real number $r \in \mathbb{R}$) actually **does** satisfy $Ax = b$. The reason for this is similar to the one we gave in Example 3.153: Namely, in the process of finding the solution (148), we did not "throw away any requirements".

So we have learned that a column vector x satisfies $Ax = b$ if and only if it has the form (148) for **some** number $r \in \mathbb{R}$. Thus, the equation $Ax = b$ has infinitely many solutions – namely, its solutions are all vectors x of the form (148) with $r \in \mathbb{R}$. This explains why we were unable to find an equation to solve for x_3: Different solutions x have different values of x_3.

Example 3.155. Let $A = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. We want to solve the equation $Ax = b$; in other words, we want to find all column vectors x of size 4 that satisfy $Ax = b$.

Write the unknown vector x in the form $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$. Then, $Ax = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 0 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. Hence,

the equation $Ax = b$ rewrites as $\begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}$. This, in turn, is equivalent to the following system of linear equations:

$$\begin{cases} 4x_1 + 3x_2 + 4x_4 = 3; \\ -x_2 + 3x_3 + 2x_4 = 1; \\ 0 = 5 \end{cases} . \tag{149}$$

The system (149) can be solved very easily: Its third equation (that is, $0 = 5$) is **never** satisfied (no matter what x is). Thus, the whole system (149) is never satisfied. In other words, $Ax = b$ is never satisfied. Hence, $Ax = b$ has no solutions.

Example 3.156. Let $A = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$. We want to solve the equation $Ax = b$; in other words, we want to find all column vectors x of size 4 that satisfy $Ax = b$.

Write the unknown vector x in the form $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$. Then, $Ax = \begin{pmatrix} 4 & 3 & 0 & 4 \\ 0 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 0 \end{pmatrix}$ and $b = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$. Hence, the equation $Ax = b$ rewrites as $\begin{pmatrix} 4x_1 + 3x_2 + 4x_4 \\ -x_2 + 3x_3 + 2x_4 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}$. This, in turn, is equivalent to the following system of linear equations:

$$\begin{cases} 4x_1 + 3x_2 + 4x_4 = 3; \\ -x_2 + 3x_3 + 2x_4 = 1; \\ 0 = 0 \end{cases} . \tag{150}$$

The system (150) can be solved similarly to (150), after a very simple first step:

1. The third equation in (150) is **always** satisfied (since it says that $0 = 0$), and thus places no restrictions on x. Hence, it can be discarded.

2. There is no equation (150) that "begins with x_4" (i.e., that has x_4 appear with nonzero coefficient, but x_1, x_2 and x_3 with zero coefficients). As in Example 3.154, we use this as an opportunity to treat x_4 as a known. In other words, let us define a number r by $r = x_4$, and pretend that $x_4 = r$ is an explicit value for x_4.

3. There is no equation (150) that "begins with x_3" (i.e., that has x_3 appear with nonzero coefficient, but x_1 and x_2 with zero coefficients). As in Example 3.154, we use this as an opportunity to treat x_3 as a known. In other words, let us define a number s by $s = x_3$, and pretend that $x_3 = s$ is an explicit value for x_3. (Of course, we must not reuse an existing variable such as r here.)

4. Now that we have "found" the values of x_3 and x_4, we can substitute them into the second equation in (150), so that the latter equation becomes $-x_2 + 3s + 2r = 1$. We can now solve this equation for x_2, obtaining $x_2 = 2r + 3s - 1$.

5. Now that we have "found" the values of x_2, x_3 and x_4, we can substitute them into the first equation in (150), so that the latter equation becomes $4x_1 + 3(2r + 3s - 1) + 4r = 3$. We can now solve this equation for x_1, obtaining $x_1 = \frac{3}{2} - \frac{9}{4}s - \frac{5}{2}r$.

Altogether, we have now found $x_1 = \frac{3}{2} - \frac{9}{4}s - \frac{5}{2}r$, $x_2 = 2r + 3s - 1$, $x_3 = s$ and $x_4 = r$. In other words,

$$x = \begin{pmatrix} \frac{3}{2} - \frac{9}{4}s - \frac{5}{2}r \\ 2r + 3s - 1 \\ s \\ r \end{pmatrix}. \tag{151}$$

Thus, we have found that each column vector x satisfying $Ax = b$ must have the form (151) for **some** numbers $r, s \in \mathbb{R}$.

Conversely, any column vector x of the form (151) (for **any** real numbers $r, s \in \mathbb{R}$) actually **does** satisfy $Ax = b$. This holds, again, for the same reason as in Example 3.153.

So we have learned that a column vector x satisfies $Ax = b$ if and only if it has the form (151) for **some** numbers $r, s \in \mathbb{R}$. Thus, the equation $Ax = b$ has infinitely many solutions – namely, its solutions are all vectors x of the form (151) with $r, s \in \mathbb{R}$.

We have now seen four examples of an algorithm for solving equations of the form $Ax = b$ where A is a row-echelon matrix. The general form of the algorithm should now be easy to guess, but let me pin it down:

[...] [THE BELOW IS INCOMPLETE]

Theorem 3.157. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $n \times m$-matrix. Let b be an $n \times 1$-vector. Assume that A is a row-echelon matrix.

Consider the problem of finding all solutions to the equation $Ax = b$. In other words, we want to find all $m \times 1$-matrices x satisfying $Ax = b$. This problem can be solved by the following algorithm (*called back-substitution*):

1. Set $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$. Multiplying the equalities $A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{pmatrix}$ and $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$, we obtain

$$\begin{aligned} Ax &= \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,m} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,m} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \\ &= \begin{pmatrix} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,m}x_m \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,m}x_m \\ \vdots \\ A_{n,1}x_1 + A_{n,2}x_2 + \cdots + A_{n,m}x_m \end{pmatrix}. \end{aligned}$$

Thus, the equation $Ax = b$ rewrites as

$$\begin{pmatrix} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,m}x_m \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,m}x_m \\ \vdots \\ A_{n,1}x_1 + A_{n,2}x_2 + \cdots + A_{n,m}x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

(since $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$). Hence, it is equivalent to the system

$$\begin{cases} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,m}x_m = b_1; \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,m}x_m = b_2; \\ \vdots \\ A_{n,1}x_1 + A_{n,2}x_2 + \cdots + A_{n,m}x_m = b_n \end{cases}. \tag{152}$$

This is a system of n linear equations in the m unknowns $x_1, x_2, \ldots, x_m$. We thus focus on solving this system.

2. Let k be the number of nonzero rows of A. Then, the last $n-k$ rows of A are zero (by Proposition 3.151 **(a)**). Hence, the last $n-k$ equations in the system (152) have the form $0 = b_i$ for some $i \in \{n-k+1, n-k+2, \ldots, n\}$. If **at**

least one of the values $b_{n-k+1}, b_{n-k+2}, \ldots, b_n$ is nonzero, then the equation $Ax = b$ has no solution. In this case, we should stop right here, since the problem is solved. (This is the situation we have encountered in Example 3.155.)

3. Assume now that none of the values $b_{n-k+1}, b_{n-k+2}, \ldots, b_n$ is nonzero. Thus, all of the values $b_{n-k+1}, b_{n-k+2}, \ldots, b_n$ are zero. Hence, the last $n-k$ equations in the system (152) have the form $0 = 0$, and therefore can be discarded (since they are automatically true). (Notice that if $n-k=0$, then we are discarding nothing.)

   We are thus left with the first k equations:

$$\begin{cases} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,m}x_m = b_1; \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,m}x_m = b_2; \\ \vdots \\ A_{k,1}x_1 + A_{k,2}x_2 + \cdots + A_{k,m}x_m = b_k \end{cases} \tag{153}$$

4. For each $i \in \{1, 2, \ldots, k\}$, set $\phi(i) = \operatorname{pivind}(\operatorname{row}_i A)$. Then, Proposition 3.151 **(c)** says that

$$\phi(1) < \phi(2) < \cdots < \phi(k).$$

   [...]

Proof of Theorem 3.157. We only need to show that in step 2 [...] □

[...]

3.24. <TODO> Gaussian elimination and the row echelon form

[...]
(to be continued)
TO-DO LIST! (This will appear eventually:)

TODO 3.158. Gaussian elimination: $C = E^* U$, where E^* is a product of elementary matrices ($A^{\lambda}_{u,v}$, S^{λ}_{i}, $T_{u,v}$) and U is a row-echelon matrix.

TODO 3.159. How this generalizes things shown before.

TODO 3.160. Example of Gaussian elimination: Start with the matrix $C = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 3 & 2 & 1 & 7 \\ 0 & 1 & -1 & 0 \\ -1 & 2 & -3 & 0 \end{pmatrix}$. Want to bring it into row echelon form by using downward

row additions (i.e., row operations $A^{\lambda}_{u,v}$ with $u > v$), row scalings (i.e., row operations S^{λ}_{u} with $\lambda \neq 0$) and row swappings (i.e., row operations $T_{u,v}$) only.

Subtracting row 1 from row 2 gives new matrix $C' = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ -1 & 2 & -3 & 0 \end{pmatrix}$. Thus $C = A^{1}_{2,1} C'$.

Subtracting $\frac{-1}{3}$ times row 1 from row 4 gives new matrix $C'' = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & \frac{8}{3} & -\frac{8}{3} & 2 \end{pmatrix}$. Thus $C' = A^{-1/3}_{4,1} C''$.

Now, the first column looks like row echelon form.

The entry 0 in cell $(2,2)$ prevents the first two columns from being row-echelon. Thus, swap it with a nonzero entry further down.

Swapping rows 2 and 3 gives new matrix $C''' = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{8}{3} & -\frac{8}{3} & 2 \end{pmatrix}$. Thus $C'' = T_{2,3} C'''$.

Subtracting $\frac{8}{3}$ times row 2 from row 4 gives new matrix $C'''' = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$. Thus $C''' = A^{8/3}_{4,2} C''''$.

Now, the first two columns look like row echelon form.

Usually we would have to do some operations for the third column, but this time we are in luck: The first three columns already look like row echelon form, so no operations are needed here.

However, the fourth column still needs an operation: It has a nonzero entry below its pivot 1.

Subtracting 2 times row 3 from row 4 gives new matrix $C''''' = \begin{pmatrix} 3 & 2 & 1 & 6 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Thus $C'''' = A^{2}_{4,3} C'''''$.

The matrix C''''' is in row echelon form; its pivot entries are in cells $(1,1)$, $(2,2)$

and $(3,4)$. Altogether,

$$\begin{aligned} C &= A_{2,1}^{1} \underbrace{C'}_{=A_{4,1}^{-1/3}C''} = A_{2,1}^{1} A_{4,1}^{-1/3} \underbrace{C''}_{=T_{2,3}C'''} = A_{2,1}^{1} A_{4,1}^{-1/3} T_{2,3} \underbrace{C'''}_{=A_{4,2}^{8/3}C''''} \\ &= A_{2,1}^{1} A_{4,1}^{-1/3} T_{2,3} A_{4,2}^{8/3} \underbrace{C''''}_{=A_{4,3}^{2}C'''''} = A_{2,1}^{1} A_{4,1}^{-1/3} T_{2,3} A_{4,2}^{8/3} A_{4,3}^{2} C'''''. \end{aligned}$$

We could tweak C''''' furthermore (by row operations) to obtain a *row-reduced echelon matrix*, which is characterized by the properties that **(a)** each pivot entry equals 1, and **(b)** in each column containing a pivot entry, all other entries are 0. To achieve **(a)**, we merely have to apply some row scaling operations. To achieve **(b)**, we need to apply *upward row additions*, i.e., row operations $A_{u,v}^{\lambda}$ with $u < v$; this way we can clear out the entries above each pivot.

But if we allow ourselves column operations as well, then we can even end up with a matrix which has entries 1 in cells $(1,1), (2,2), \ldots, (k,k)$ for some k, and entries 0 in all other cells. (Think of it as a truncated identity matrix, except that it is rectangular.)

TODO 3.161. History: [Grcar10] reference.

TODO 3.162. How to solve $Uv = b$ (back-substitution; beware of zero rows).

TODO 3.163. Thus, solve $Cv = b$.

TODO 3.164. How to see whether a matrix C is invertible using Gauss.

3.25. <TODO> PLU decomposition

TODO 3.165. PLU decomposition (not sure). (I didn't do PLU in class properly, and doing it right is subtle. [OlvSha06, Example 1.12] does it.)

TODO 3.166. Most of the time, $P = I_n$, because there is no need to permute rows when the entries are nonzero.

3.26. <TODO> Determinants (briefly)

Determinants are not a central player in these notes. Nevertheless, they are sufficiently important to be mentioned. I consider them to be one of the most beautiful objects in mathematics, and they are also one of the most useful in **pure** mathematics (particularly, in all breeds of algebra and combinatorics). However, their use in applied mathematics is rather limited (apart from "small" cases like matrices of size 2×2, 3×3 and 4×4, occurring in physics and geometry for well-known reasons). Thus, I shall give no proofs and barely state the most crucial results. Three good references on determinants are:

- [LaNaSc16, Chapter 8] does the definitions and the basic properties really well. I highly recommend it.
- [Heffer16, Chapter Four] also has a neat treatment of determinants, although unfortunately it requires Gauss-Jordan elimination (which I have not done above, and which I find out-of-place in an introduction to determinants). It has many examples and explains the geometric meaning of 2×2-determinants (as areas of triangles) and 3×3-determinants (as volumes of tetrahedra).
- [BarSch73, Chapter 4] doesn't always have the best proof (in particular, it uses some properties of singular matrices that we have not yet learned), but seems to be written well and with attention to detail.

Apart from that, lots of other texts define and discuss determinants: some sloppily, some rather well.[103] When reading other sources, be aware of a possible conflict of notations: Namely, we are using the notation $A_{i,j}$ for the (i,j)-th entry of a matrix A; but most authors use the notation $A_{i,j}$ for something different (namely, the (i,j)-th "cofactor" of A).

TODO 3.167. Recall Example 3.4 **(e)**: The number $ad - bc$ governs whether the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an inverse. Does such a number exist for larger matrices? Yes, it does, and is called the determinant.

TODO 3.168. The determinant of an $n \times n$-matrix A is a number, denoted by $\det A$ and defined later. (Some authors call it $|A|$, but we will not.)

TODO 3.169. Defining the determinant of an $n \times n$-matrix is not that easy. Let us first do it for small cases:

$$\det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = a_1 b_2 - a_2 b_1;$$

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = a_1 b_2 c_3 - a_1 c_2 b_3 - b_1 a_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3 - c_1 b_2 a_3;$$

$$\det \begin{pmatrix} a \end{pmatrix} = a;$$

$$\det () = 1.$$

(The last equality says that the determinant of a 0×0-matrix is 1, and yes, that's sometimes useful.)

[103] Olver's and Shakiban's treatment of determinants in [OlvSha06, §1.9] is very minimalistic and incomplete (in particular, they never bother proving that the determinant is well-defined). On the opposite side of the spectrum, two really rigorous and detailed treatments are Gill Williamson's notes [Gill12, Chapter 3] and my [Grinbe16]; but the downside of "rigorous" is "unmotivated and lacking intuition", and the downside of "detailed" is "extremely long". You will want to strike your own balance. As I said, I recommend [LaNaSc16, Chapter 8].

TODO 3.170. Let me give three versions of the definition of $\det A$ for all n:

- *First version* (informal and sloppy)*:* Set

$$\det \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \\ \vdots & \vdots & \ddots & \vdots \\ z_1 & z_2 & \cdots & z_n \end{pmatrix} = \text{sum of terms of the form } \pm ?_1 ?_2 \cdots ?_n,$$

  where the question marks in each term are some permutation (= rearrangement) of the letters $a, b, \ldots, z$, and the $\pm$ sign depends on the rearrangement. How exactly? Any rearrangement can be sorted into increasing order (i.e., into $ab \ldots z$) by switching pairs of letters; e.g. (for $n = 5$) we can sort the rearrangement *caebd* as follows:

$$\begin{aligned} \textit{caebd} &\to \textit{cabed} && \text{(here we switched } e \text{ with } b) \\ &\to \textit{baced} && \text{(here we switched } c \text{ with } b) \\ &\to \textit{abced} && \text{(here we switched } b \text{ with } a) \\ &\to \textit{abcde} && \text{(here we switched } e \text{ with } d)\,. \end{aligned}$$

  Notice that we used 4 switches here. The rule for the $\pm$ sign is: If the rearrangement needs an even number of switches, then it's a $+$; if an odd number, then it's a $-$.

- *Second version* (formal): Recall that each permutation matrix P is a product of several swapping matrices (i.e., of $T_{u,v}$'s). Define $\operatorname{sign}(P)$ to be $+1$ if P is the product of an even number of swapping matrices; define $\operatorname{sign}(P)$ to be -1 if P is the product of an odd number of swapping matrices. It is not obvious that this definition is legitimate (because if P would be representable both as an even product and as an odd product, then $\operatorname{sign}(P)$ would have to be $+1$ and -1 at the same time!), but it is nevertheless true (it can be shown that no P can ever be representable both as even product and as odd product). If P is a permutation matrix and A is an $n \times n$-matrix, then the *P-product of A* shall mean the product of all entries of A in those cells in which P has a 1. For example, if $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, then the P-product of $\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$ is $a_1 b_3 c_2$, because the cells in which P has a 1 are the cells $(1,1)$, $(2,3)$ and $(3,2)$, and the entries of A in these cells are a_1, b_3 and c_2. Now, the determinant $\det A$ of A is the sum of

$$\operatorname{sign}(P) \cdot (\text{the } P\text{-product of } A)$$

  for all permutation matrices P.

- *Third version* (formal, uses Section 3.21)*:* Let

$$\det A = \sum_{\sigma \in S_n} (-1)^\sigma A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)},$$

  where S_n denotes the set of all permutations of $\{1, 2, \ldots, n\}$, and the number $(-1)^\sigma$ is defined as follows: Proposition 3.134 shows that the permutation σ can be written as a product of k simple transpositions for some $k \in \mathbb{N}$. If this k is even, then set $(-1)^\sigma = 1$; if this k is odd, then set $(-1)^\sigma = -1$. (Again, we need to prove that this definition is legitimate, i.e., that $(-1)^\sigma$ is never defined to be 1 and -1 at the same time.)

Note that the third version of the definition is merely the formalization of the first version. The second version is just the translation of the third version into the language of permutation matrices.

TODO 3.171. State the main properties of determinants:

- We have $\det(0_{n\times n}) = 0$ for all $n > 0$.
- We have $\det(I_n) = 1$ for all n.
- We have $\det\left(A^T\right) = \det A$ for any square matrix A.
- The determinant is multilinear in the rows (i.e., linear **in each row**, as long as the other rows are fixed). For example,

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 + b_1' & b_2 + b_2' & b_3 + b_3' \\ c_1 & c_2 & c_3 \end{pmatrix}$$
$$= \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} + \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1' & b_2' & b_3' \\ c_1 & c_2 & c_3 \end{pmatrix}$$

  and

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ \lambda b_1 & \lambda b_2 & \lambda b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = \lambda \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

  See [LaNaSc16, Theorem 8.2.3, 3.] for a precise statement (but for columns instead of rows).
- Note that $\det(A + B) \neq \det A + \det B$ in general.
- Switching two rows of a matrix A negates $\det A$ (that is, if B is obtained from A by switching two rows, then $\det B = -\det A$).
- If A has two equal rows, then $\det A = 0$.

- If A has a row filled with zeroes, then $\det A = 0$.
- If we add a multiple of some row of A to another row, then $\det A$ does not change.
- All of the above properties that reference rows also hold for columns.
- If A is upper-triangular or lower-triangular, then $\det A$ is the product of the diagonal entries of A.
- We have $\det(AB) = \det A \cdot \det B$ for any two $n \times n$-matrices A and B. This fact is not obvious, and is one of the miracles of mathematics. It is probably the main reason why determinants are useful!
- An $n \times n$-matrix A is invertible if and only if $\det A \neq 0$.
- For each $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, n\}$, let $A_{\sim i, \sim j}$ denote the $(n-1) \times (n-1)$-matrix obtained from A by removing the i-th row and the j-th column. For example,

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}_{\sim 1, \sim 2} = \begin{pmatrix} b_1 & b_3 \\ c_1 & c_3 \end{pmatrix}.$$

  Let $\operatorname{adj} A$ be the $n \times n$-matrix $\left((-1)^{i+j} \det\left(A_{\sim j, \sim i} \right) \right)_{1 \leq i \leq n,\ 1 \leq j \leq n}$. (This is called the *adjugate* of A.) Then,

$$A \cdot \operatorname{adj} A = \operatorname{adj} A \cdot A = \det A \cdot I_n.$$

- For every $p \in \{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{q=1}^{n} (-1)^{p+q} A_{p,q} \det\left(A_{\sim p, \sim q} \right).$$

  This is called "Laplace expansion in the p-th row", and gives us a way to compute $\det A$ using determinants of $(n-1) \times (n-1)$-matrices.

- For every $q \in \{1, 2, \ldots, n\}$, we have

$$\det A = \sum_{p=1}^{n} (-1)^{p+q} A_{p,q} \det\left(A_{\sim p, \sim q} \right).$$

  This is called "Laplace expansion in the q-th column".

3.27. <TODO> The rest

TODO 3.172. (*) SageMath examples.

TODO 3.173. (*) What holds over rings, what over fields.

4. <TODO> Vector spaces

TODO 4.1. Introduction into vector spaces.

We shall now switch to what feels like a different subject: the theory of vector spaces. Soon, we will see that this is closely related to the matrix algebra that we have been doing above, and in some sense is like doing linear algebra in a more abstract language (as opposed to doing something completely different). But first, we introduce the relevant notions without reference to the matrix theory done above.

(..... This all needs to be filled in. So far, this is a stenographic lecture plan rather than a set of lecture notes .....)

4.1. <DRAFT> Vector spaces

References for vector spaces: [LaNaSc16, §4.1-§4.2] (possibly the best one), [OlvSha06, §2.1] (focusses on the analysis-related and applied stuff) and [Heffer16, Two.I.1].

Definition 4.2. A *vector space (over $\mathbb{R}$)* is a set V equipped with two binary operations:

- a binary operation called "$+$", which takes as its input two elements v and w of V, and yields an element of V called $v+w$;
- a binary operation called "$\cdot$", which takes as its input a number $\lambda \in \mathbb{R}$ and an element v of V, and yields an element of V called $\lambda \cdot v$ or λv,

as well as a chosen element of V called $\overrightarrow{0}$, with the following properties:

(a) We have $v+w = w+v$ for all $v \in V$ and $w \in V$. (This is called "commutativity of addition".)

(b) We have $u+(v+w) = (u+v)+w$ for all $u \in V$, $v \in V$ and $w \in V$. (This is called "associativity of addition".)

(c) We have $v+\overrightarrow{0} = \overrightarrow{0}+v = v$ for all $v \in V$. (This is called "neutrality of $\overrightarrow{0}$".)

(d) For each $v \in V$, there exists some element $w \in V$ satisfying $v+w = w+v = 0$. (This is called "existence of additive inverses". The element w is called the *additive inverse of v*, and is usually called $-v$; it is uniquely determined by v (this is not hard to check).)

(e) We have $(\lambda+\mu)\,v = \lambda v + \mu v$ for all $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $v \in V$. (This is called "right distributivity".)

(f) We have $\lambda\left(v+w\right)=\lambda v+\lambda w$ for all $\lambda\in\mathbb{R}$, $v\in V$ and $w\in V$. (This is called "left distributivity".)

(g) We have $\left(\lambda\mu\right)v=\lambda\left(\mu v\right)$ for all $\lambda\in\mathbb{R}$, $\mu\in\mathbb{R}$ and $v\in V$. (This is called "associativity of scaling". It allows us to write $\lambda\mu v$ for both $\left(\lambda\mu\right)v$ and $\lambda\left(\mu v\right)$.)

(h) We have $1v=v$ for all $v\in V$. (This is called "neutrality of 1".)

(i) We have $0v=\overrightarrow{0}$ for all $v\in V$.

(j) We have $\lambda\overrightarrow{0}=\overrightarrow{0}$ for all $\lambda\in\mathbb{R}$.

The operation $+$ is called the *addition* of the vector space; the operation $\cdot$ is called the *scaling* of the vector space; the element $\overrightarrow{0}$ is called the *zero vector* (or the *origin*) of the vector space. The elements of V are called *vectors*. (These are familiar-sounding names for abstract things. The operation $+$ may and may not have anything to do with addition of numbers. The operation $\cdot$ may and may not be related to scaling of numbers. The element $\overrightarrow{0}$ may and may not be the number 0. The elements of V, which we call "vectors", may and may not be the kinds of vectors we are used to seeing (i.e., row vectors and column vectors); they can just as well be polynomials or functions or numbers or matrices. You should think of the word "vector" as meaning "element of a vector space", not as meaning "a list of numbers"; the latter meaning is too restrictive. The properties **(a)**, **(b)**, ..., **(j)** are requiring that the operations $+$ and $\cdot$ and the element $\overrightarrow{0}$ "behave like" the addition and scaling of matrices and the zero matrix, at least as far as their most basic properties are concerned; however, they still leave a lot of freedom to decide what these operations and this elements should be.)

We shall often speak of "the vector space V" or say that "V is a vector space", but of course, the vector space is not just the set V; it is (as we have defined it) the set V endowed with the two operations $+$ and $\cdot$ and the element $\overrightarrow{0}$. The two operations and the element $\overrightarrow{0}$ are part of the data; if you modify them, then you end up with a different vector space, even if the set V is the same! However, most of the time, we will not have several different vector spaces sharing one and the same set V; therefore, we will be able to speak of "the vector space V" and assume that the reader knows what operations $+$ and $\cdot$ and what element $\overrightarrow{0}$ we mean.

We have just defined the notion of a "vector space over $\mathbb{R}$" (also known as a "real vector space"). We can similarly define a "vector space over $\mathbb{Q}$" (by replacing each "$\mathbb{R}$" in the above definition by "$\mathbb{Q}$") and a "vector space over $\mathbb{C}$" (by replacing each "$\mathbb{R}$" in the above definition by "$\mathbb{C}$"). Most of the vector spaces we shall be studying below are vector spaces over $\mathbb{R}$, but the other options are also useful. (If you have read Definition 2.52, you will have no trouble defining a "vector space over $\mathbb{K}$" for every field $\mathbb{K}$.)

Note that our definition of a vector space is somewhat similar to the definition of a commutative ring (Definition 2.50). At least as far as the operation $+$ alone

is concerned, the requirements on it are literally the same in the two definitions[104] (commutativity of addition, associativity of addition, neutrality of $\overrightarrow{0}$ and existence of additive inverses). However, the similarity ends here: The binary operation $\cdot$ ("multiplication") in Definition 2.50 works differently from the binary operation $\cdot$ ("scaling") in Definition 4.2. The former takes two inputs in the commutative ring, whereas the latter takes one input in $\mathbb{R}$ and one input in the vector space. The axioms still have certain similarities, but they should not fool you into believing that commutative rings are vector spaces (or vice versa).

If you compare our Definition 4.2 with other definitions of a "vector space" you find in the literature (for example, [LaNaSc16, Definition 4.1.1], [OlvSha06, Definition 2.1] or [Heffer16, Definition Two.I.1]), you will notice that they are slightly different: For example, [LaNaSc16, Definition 4.1.1], [OlvSha06, Definition 2.1] or [Heffer16, Definition Two.I.1] are lacking our properties **(i)** and **(j)**, whereas [Kowals16, Definition 2.3.1] is missing our properties **(d)** and **(j)**. However, the definitions are nevertheless **equivalent** (i.e., they define precisely the same notion of a vector space). The reason for that is some of the properties we required in Definition 4.2 are **redundant** (i.e., they follow from the other properties, so that nothing changes if we leave them out). For example:

Proposition 4.3. Property **(d)** in Definition 4.2 follows from properties **(e)**, **(h)** and **(i)**. Thus, we could leave out property **(d)** from the definition.

Proof of Proposition 4.3. Assume that properties **(e)**, **(h)** and **(i)** hold. We must now prove property **(d)**.

Let $v \in V$. Then,

$$(-1)\,v + \underbrace{v}_{\substack{=1v \\ \text{(by property \textbf{(h)})}}} = (-1)\,v + 1v = \underbrace{((-1)+1)}_{=0}\,v$$

(by property **(e)**, applied to $\lambda = -1$ and $\mu = 1$)

$$= 0v = \overrightarrow{0} \qquad \text{(by property \textbf{(i)})}$$

and

$$\underbrace{v}_{\substack{=1v \\ \text{(by property \textbf{(h)})}}} + (-1)\,v = 1v + (-1)\,v = \underbrace{(1+(-1))}_{=0}\,v$$

(by property **(e)**, applied to $\lambda = 1$ and $\mu = -1$)

$$= 0v = \overrightarrow{0} \qquad \text{(by property \textbf{(i)})}\,.$$

Hence, there exists some element $w \in V$ satisfying $v + w = w + v = 0$ (namely, $w = (-1)\,v$). Thus, property **(d)** is proven. Hence, we have shown that property **(d)** follows from properties **(e)**, **(h)** and **(i)**. This proves Proposition 4.3. □

[104] assuming that we identify the $\overrightarrow{0}$ in Definition 4.2 with the 0 in Definition 2.50

Proposition 4.4. Property **(j)** in Definition 4.2 follows from properties **(g)** and **(i)**. Thus, we could omit property **(j)** from the definition.

Proof of Proposition 4.4. Assume that properties **(g)** and **(i)** hold. We must now prove property **(j)**.

Let $\lambda \in \mathbb{R}$. Then, property **(i)** (applied to $v = \overrightarrow{0}$) yields $0 \cdot \overrightarrow{0} = \overrightarrow{0}$. Thus, $\lambda \left(0 \cdot \overrightarrow{0}\right) = \lambda \overrightarrow{0}$, so that

$$\lambda \overrightarrow{0} = \lambda \left(0 \cdot \overrightarrow{0}\right) = \underbrace{(\lambda \cdot 0)}_{=0} \overrightarrow{0} \qquad \left(\text{by property } \textbf{(g)}, \text{ applied to } \mu = 0 \text{ and } v = \overrightarrow{0}\right)$$
$$= 0 \cdot \overrightarrow{0} = \overrightarrow{0}.$$

Thus, property **(j)** is proven. Hence, we have shown that property **(j)** follows from properties **(g)** and **(i)**. This proves Proposition 4.4. □

Proposition 4.5. Property **(i)** in Definition 4.2 follows from properties **(b)**, **(c)**, **(d)** and **(e)**. Thus, we could omit property **(i)** from the definition.

Proof of Proposition 4.5. Assume that properties **(b)**, **(c)**, **(d)** and **(e)** hold. We must now prove property **(i)**.

Let $v \in V$. Property **(e)** (applied to $\lambda = 0$ and $\mu = 0$) yields $(0+0)\, v = 0v + 0v$. Hence, $0v + 0v = \underbrace{(0+0)}_{=0} v = 0v$.

But we can apply property **(d)** to $0v$ instead of v. We thus conclude that there exists some element $w \in V$ satisfying $0v + w = w + 0v = \overrightarrow{0}$. Consider this w.

Property **(b)** (applied to $0v$, $0v$ and w instead of u, v and w) yields $0v + (0v + w) = \underbrace{(0v + 0v)}_{=0v} + w = 0v + w$. Since $0v + w = \overrightarrow{0}$, this rewrites as $0v + \overrightarrow{0} = \overrightarrow{0}$.

Property **(c)** (applied to $0v$ instead of v) yields $0v + \overrightarrow{0} = \overrightarrow{0} + 0v = 0v$. Comparing $0v + \overrightarrow{0} = \overrightarrow{0}$ with $0v + \overrightarrow{0} = 0v$, we obtain $0v = \overrightarrow{0}$. Thus, property **(i)** is proven. Hence, we have shown that property **(i)** follows from properties **(b)**, **(c)**, **(d)** and **(e)**. This proves Proposition 4.5. □

However, we **cannot** simultaneously omit **both** properties **(d)** and **(i)** from the definition: Each of them is redundant provided the other stays in, but without the other they are not redundant.

The vector $\overrightarrow{0}$ in Definition 4.2 is often written as 0. This is an "abuse of notation" (because the vector $\overrightarrow{0}$ is not literally the number 0), but mostly harmless as there is rarely a possibility of confusion.

4.2. <DRAFT> Examples and constructions of vector spaces

Before we do anything interesting with vector spaces, let me show various examples of them:

Example 4.6. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. The set of all $n \times m$-matrices with real entries shall be denoted by $\mathbb{R}^{n \times m}$. This set $\mathbb{R}^{n \times m}$ is a vector space – or, more precisely, it becomes a vector space if we define the operation "$+$" as addition of matrices (conveniently, we have already been denoting this addition by "$+$" throughout these notes), define the operation "$\cdot$" as scaling of matrices (again, we have fortunately always been calling it "$\cdot$"), and define the zero vector $\overrightarrow{0}$ as the zero matrix $0_{n \times m}$. Proving that the properties in Definition 4.2 are satisfied is easy; in fact, they all boil down to simple facts about matrices.

Thus, $n \times m$-matrices are vectors (in the vector space $\mathbb{R}^{n \times m}$). This does not mean that they are row vectors or column vectors; we are instead using the word "vector" in its general meaning of "element of a vector space" here.

Of course, row vectors and column vectors are vectors as well (because they are matrices). The vector space of all column vectors of size n is $\mathbb{R}^{n \times 1}$ (since column vectors of size n are $n \times 1$-matrices). We will denote this vector space by $\mathbb{R}^n$. (Beware that some authors also use the notation $\mathbb{R}^n$ for the vector space of all row vectors of size n, which is $\mathbb{R}^{1 \times n}$. Some authors go as far as pretend that column vectors and row vectors are the same, but this is dangerous, since matrix multiplication treats them differently!)

Notice that multiplication of matrices does not matter for the vector space $\mathbb{R}^{n \times m}$. Even if there was no such thing as a product of two matrices, $\mathbb{R}^{n \times m}$ would still be a vector space, since vector spaces only need addition and scaling (and $\overrightarrow{0}$), but no multiplication of vectors.

Definition 4.7. The notations $\mathbb{R}^{n \times m}$ and $\mathbb{R}^n$ introduced in Example 4.6, as well as the vector space structure on $\mathbb{R}^{n \times m}$, will be used throughout these notes.

Example 4.8. Some simpler examples:

- The set $\mathbb{R}$ of all real numbers itself is a vector space (with addition being addition, scaling being multiplication, and $\overrightarrow{0}$ being the number 0.)
- The set $\mathbb{C}$ of all complex numbers is a vector space (with addition being addition, scaling being multiplication, and $\overrightarrow{0}$ being the number 0.)
- The one-element set $\{0\}$ is also a vector space (with addition, scaling and $\overrightarrow{0}$ being defined in the only possible way – there is only one choice).

Example 4.9. Consider the set $\mathbb{R}^\infty$ of all infinite sequences $(a_1, a_2, a_3, \ldots)$ of real numbers. This set $\mathbb{R}^\infty$ is also a vector space, with addition defined by

$$(a_1, a_2, a_3, \ldots) + (b_1, b_2, b_3, \ldots) = (a_1 + b_1, a_2 + b_2, a_3 + b_3, \ldots)$$

(that is, entry by entry), scaling defined by

$$\lambda (a_1, a_2, a_3, \ldots) = (\lambda a_1, \lambda a_2, \lambda a_3, \ldots)$$

(thus, again, entry by entry), and $\overrightarrow{0}$ defined by

$$\overrightarrow{0} = (0, 0, 0, \ldots).$$

(It makes sense to think of infinite sequences as $1 \times \infty$-matrices; thus, the above definitions of $+$, $\cdot$ and $\overrightarrow{0}$ are precisely the rules we set for matrices.)

Example 4.10. Let S be **any** set. Consider the set $\mathbb{R}^S$ of all maps from S to $\mathbb{R}$. Then, $\mathbb{R}^S$ becomes a vector space, if we define $+$, $\cdot$ and $\overrightarrow{0}$ as follows:

- If $f \in \mathbb{R}^S$ and $g \in \mathbb{R}^S$ are two maps, then their sum $f + g$ is defined to be the map from S to $\mathbb{R}$ that sends each $s \in S$ to $f(s) + g(s)$. Thus,

  $$(f + g)(s) = f(s) + g(s) \qquad \text{for every } s \in S.$$

  This is called *pointwise addition* (because it means that we add two maps by adding their values at each point).

- If $f \in \mathbb{R}^S$ is a map and λ is a number, then the map λf is defined to be the map from S to $\mathbb{R}$ that sends each $s \in S$ to $\lambda \cdot f(s)$. Thus,

  $$(\lambda f)(s) = \lambda \cdot f(s) \qquad \text{for every } s \in S.$$

  This is called *pointwise scaling* (because it means that we scale a map by scaling its value at each point).

- We define $\overrightarrow{0} \in \mathbb{R}^S$ to be the map from S to $\mathbb{R}$ that sends each $s \in S$ to 0. This is called the *constant-0 map*, since it is constant and all its values are 0.

This example can be viewed as a generalization of the previous examples. Namely:

- An $n \times m$-matrix A with real entries can be viewed as a map from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to $\mathbb{R}$ (namely, as the map which sends each $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to the (i, j)-th entry $A_{i,j}$ of the matrix A). Conversely, each map from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to $\mathbb{R}$ can be encoded by an $n \times m$-matrix (whose entries are the values of the map). Thus, we can identify the $n \times m$-matrices (with real entries) with the maps from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to $\mathbb{R}$. In other words, we can identify the set $\mathbb{R}^{n \times m}$ with the set $\mathbb{R}^{\{1,2,\ldots,n\} \times \{1,2,\ldots,m\}}$. It is easily seen that the addition $+$, the scaling $\cdot$ and the zero vector $\overrightarrow{0}$ of the vector space $\mathbb{R}^{n \times m}$ (as defined in Example 4.6) are identical with the addition $+$, the scaling $\cdot$ and the zero vector $\overrightarrow{0}$ of the vector space $\mathbb{R}^{\{1,2,\ldots,n\} \times \{1,2,\ldots,m\}}$ (as defined in the current example), once we identify the two sets. Hence, the **vector space** $\mathbb{R}^{n \times m}$ can be identified with the **vector space** $\mathbb{R}^{\{1,2,\ldots,n\} \times \{1,2,\ldots,m\}}$.

- The vector space $\mathbb{R}$ (from Example 4.8) can be identified with the vector space $\mathbb{R}^{\{1\}}$.
- The vector space $\mathbb{C}$ (from Example 4.8) can be identified with the vector space $\mathbb{R}^{\{1,2\}}$ (since a complex number is defined as a pair of two real numbers, and since the addition of two complex numbers is componentwise, and so is the scaling of a complex number by a real number).
- Recall that $\varnothing$ denotes the empty set (i.e., the set $\{\}$). The vector space $\{0\}$ (from Example 4.8) can be identified with the vector space $\mathbb{R}^{\varnothing}$. Indeed, the only element 0 of $\{0\}$ can be equated with the only map from $\varnothing$ to $\mathbb{R}$.
- The vector space $\mathbb{R}^{\infty}$ (from Example 4.9) can be identified with the vector space $\mathbb{R}^{\{1,2,3,\ldots\}}$.

I have not explained what it really means to "identify" two sets or vector spaces. Roughly speaking, identifying two sets S and T means matching up each element of S with an element of T (in such a way that each element of T ends up matched with exactly one element of S) and pretending that each element is identical to its match. When S and T are vector spaces, one additionally wants to ensure that this matching has the property that adding two elements of S has the same result as adding their matches in T, and similarly for scaling, and finally that the zero vector of S is matched with the zero vector of T. This is all satisfied in our above examples. Of course, this is still vague and imprecise. A rigorous way to formalize the idea of "identifying" is the concept of *isomorphisms* (i.e., invertible *linear maps*) between vector spaces; we shall introduce this concept later.

Example 4.11. The set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space as well. Again, addition and scaling are pointwise (i.e., defined in the same way as in Example 4.10), and again the zero vector is the constant-0 map. What makes this definition work is that:

- the sum of two continuous functions is continuous;
- the result of scaling a continuous function by a real λ is again continuous;
- the constant-0 map is continuous.

Example 4.12. The set of all polynomial functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space, too. (A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a *polynomial function* if there exist some $n \in \mathbb{N}$ and some $a_0, a_1, \ldots, a_n \in \mathbb{R}$ such that every $x \in \mathbb{R}$ satisfies $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$. For instance, the function f that sends each $x \in \mathbb{R}$ to $2(x+3)^2 x^3 + 2$ is a polynomial function, because it satisfies

$f(x) = 2(x+3)^2 x^3 + 2 = 2 + 0x + 0x^2 + 18x^3 + 12x^4 + 2x^5$ for all $x \in \mathbb{R}$.)

Example 4.13. The set of all constant functions from $\mathbb{R}$ to $\mathbb{R}$ is a vector space as well. But it can be identified with the vector space $\mathbb{R}$ (see Example 4.10 for the meaning of "identify"); in fact, any real number $r \in \mathbb{R}$ corresponds to the constant function that sends everything to r.

Example 4.14. You probably have seen at least two notions of "vectors" in analytic geometry: vectors in the plane, and vectors in space. The former kind of vectors can be identified with elements of $\mathbb{R}^2$, and the latter can be identified with elements of $\mathbb{R}^3$. Namely:

- In the plane, a vector that starts at the origin and ends at the point (a, b) is identified with the column vector $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$. (Every vector in the plane can be made to start at the origin, and its ending point is then uniquely determined; thus, this definition is legitimate.)
- In space, a vector that starts at the origin and ends at the point (a, b, c) is identified with the column vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$.

Let me show two ways to construct new vector spaces from given vector spaces; this, of course, can be used to obtain lots of further examples.

Definition 4.15. Let V and W be two sets. Recall that $V \times W$ denotes the set of all pairs (v, w) with $v \in V$ and $w \in W$. This set $V \times W$ is called the *Cartesian product* (or simply the *product*) of the two sets V and W.

Now, assume that V and W are two vector spaces. Then, we make the Cartesian product $V \times W$ into a vector space as well. Namely, we define

- its addition by
$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2) \tag{154}$$
(that is, entry by entry),
- its scaling by
$$\lambda (v, w) = (\lambda v, \lambda w) \tag{155}$$
(that is, entry by entry), and
- its zero vector by
$$\overrightarrow{0} = \left(\overrightarrow{0}, \overrightarrow{0} \right) \tag{156}$$
(that is, entry by entry again).

(Let me dwell on the precise meaning of the equalities (154), (155) and (156). The sign "$+$" appears three times in (154), and each time it has a slightly different meaning. Namely, the sign "$+$" on the left hand side of (154) stands for the addition of the vector space $V \times W$ (which we are defining). The sign "$+$" in "$v_1 + v_2$" stands for the addition of the vector space V. The sign "$+$" in "$w_1 + w_2$" stands for the addition of the vector space W. Similarly, the meanings of the expressions "$\lambda\,(v, w)$", "λv" and "λw" in (155) are slightly different, since each of these refers to the scaling of a different vector space. Finally, the three terms "$\overrightarrow{0}$" in (156) are the zero vectors of three different vector spaces (namely, of $V \times W$, of V and of W, in order of appearance).)

The vector space $V \times W$ is called the *Cartesian product* of the vector spaces V and W.

We can similarly define the Cartesian product of several vector spaces $V_1, V_2, \ldots, V_n$. It is again a vector space; its elements are n-tuples $(v_1, v_2, \ldots, v_n)$ for which each v_i belongs to the corresponding V_i; its addition, scaling and zero vector are again defined entry by entry.

Definition 4.16. Let V be a vector space. Let S be a set. Recall that V^S denotes the set of all maps from S to V. We make V^S into a vector space as well. Namely, we define its addition $+$, its scaling $\cdot$, and its zero vector $\overrightarrow{0}$ as follows:

- If $f \in V^S$ and $g \in V^S$ are two maps, then their sum $f + g$ is defined to be the map from S to V that sends each $s \in S$ to $f(s) + g(s)$ (where the "$+$" in "$f(s) + g(s)$" refers to the addition of V). Thus,

$$(f + g)(s) = f(s) + g(s) \qquad \text{for every } s \in S.$$

  This is called *pointwise addition* (because it means that we add two maps by adding their values at each point).

- If $f \in V^S$ is a map and λ is a number, then the map λf is defined to be the map from S to V that sends each $s \in S$ to $\lambda \cdot f(s)$ (where the "$\cdot$" in "$\lambda \cdot f(s)$" refers to the scaling of V). Thus,

$$(\lambda f)(s) = \lambda \cdot f(s) \qquad \text{for every } s \in S.$$

  This is called *pointwise scaling* (because it means that we scale a map by scaling its value at each point).

- We define $\overrightarrow{0} \in V^S$ to be the map from S to V that sends each $s \in S$ to $\overrightarrow{0}$ (the zero vector of V). This is called the *constant-$\overrightarrow{0}$ map*, since it is constant and all its values are $\overrightarrow{0}$.

This example, of course, generalizes Example 4.10: If we set $V = \mathbb{R}$, then V^S becomes precisely the vector space $\mathbb{R}^S$ defined in Example 4.10.

4.3. <DRAFT> (*) The summation sign for vectors

In Section 2.9, we have introduced the summation sign $\sum$ for sums of numbers; in Section 2.14, we have introduced the analogous summation sign $\sum$ for sums of matrices. In the same way, we can define the summation sign $\sum$ for sums of vectors in an arbitrary vector space:

Definition 4.17. Let V be a vector space. Let p and q be two integers such that $p \leq q+1$. Let $v_p, v_{p+1}, \ldots, v_q$ be some vectors in V. Then, $\sum\limits_{k=p}^{q} v_k$ means the sum $v_p + v_{p+1} + \cdots + v_q$. (This sum is well-defined, since each vector space satisfies a "general associativity law" similar to Proposition 2.23. Moreover, the order of its addends does not matter, as can be proven using property **(a)** in Definition 4.2.)

The notation $\sum\limits_{k=p}^{q} v_k$ is analogous to the notation $\sum\limits_{k=p}^{q} a_k$ introduced in Definition 2.26 (and to the notation $\sum\limits_{k=p}^{q} A_k$ introduced in Definition 2.43). The same remarks and clarifications done for the latter notation in Definition 2.27 apply to the former notation. There is only one difference: When the sum $\sum\limits_{k=p}^{q} v_k$ has no addends (i.e., when $p \geq q+1$), its value is defined to be the zero vector $\overrightarrow{0}$ (of the vector space V) rather than the number 0.

The summation sign $\sum$ for vectors (that we have just defined) generalizes both the summation sign $\sum$ for numbers and the summation sign $\sum$ for matrices (since numbers and matrices can both be viewed as vectors).

Most properties of the summation sign $\sum$ that hold for numbers can be straightforwardly generalized to vectors (in an arbitrary vector space). Some of them can even be generalized in several ways: For example, in Proposition 2.29, we can replace the numbers $a_p, a_{p+1}, \ldots, a_q$ by vectors (while b remains a number), **or** we can replace the number b by a vector (while $a_p, a_{p+1}, \ldots, a_q$ remain numbers; for this generalization, we should rewrite the products ba_k and $b\sum\limits_{k=p}^{q} a_k$ as $a_k b$ and $\left(\sum\limits_{k=p}^{q} a_k\right) b$). (We cannot replace both $a_p, a_{p+1}, \ldots, a_q$ and b by vectors, since the product of two vectors is not defined!) Let me state the two resulting generalizations explicitly (with slightly modified names):

Proposition 4.18. Let V be a vector space. Let p and q be two integers such that $p \leq q+1$. Let $v_p, v_{p+1}, \ldots, v_q$ be some vectors in V. Let b be a number. Then,

$$\sum_{k=p}^{q} bv_k = b \sum_{k=p}^{q} v_k.$$

(The expression $\sum_{k=p}^{q} bv_k$ has to be read as $\sum_{k=p}^{q} (bv_k)$.)

Proposition 4.19. Let V be a vector space. Let p and q be two integers such that $p \leq q+1$. Let $a_p, a_{p+1}, \ldots, a_q$ be some numbers. Let $v \in V$ be a vector. Then,

$$\sum_{k=p}^{q} a_k v = \left(\sum_{k=p}^{q} a_k \right) v.$$

(The expression $\sum_{k=p}^{q} a_k v$ has to be read as $\sum_{k=p}^{q} (a_k v)$.)

Let me also state the straightforward generalization of Proposition 2.40 to vectors:

Proposition 4.20. Let V be a vector space. Let p and q be two integers such that $p \leq q$. Let $r \in \{p, p+1, \ldots, q\}$. Let $v_p, v_{p+1}, \ldots, v_q$ be some vectors in V. Then,

$$\sum_{k=p}^{q} \delta_{k,r} v_k = v_r.$$

4.4. <TODO> Subspaces

References for subspaces: [LaNaSc16, §4.3] (possibly the best one), [OlvSha06, §2.2] (focusses on the analysis-related and applied stuff) and [Heffer16, Two.I.2].

When U is a subset of a vector space V, we can try to make U itself into a vector space by "inheriting" the addition, the scaling and the zero vector from V: That is, we define the sum $v+w$ of two vectors $v \in U$ and $w \in U$ to be the result of adding v and w as elements of V. Similarly, we define λv for $\lambda \in \mathbb{R}$ and $v \in U$ to be the result of scaling v by λ as element of V. Finally, we define the zero vector $\overrightarrow{0}$ of U as the zero vector $\overrightarrow{0}$ of V. However, this all works (i.e., makes U into a vector space) only if the vectors appearing in these definitions end up in U: For example, our definition of $v+w$ actually defines an addition on U only if the sum $v+w$ actually belongs to U for all $v \in U$ and $w \in U$. Similarly, scaling on U is well-defined only if $\lambda v \in U$ for all $\lambda \in \mathbb{R}$ and $v \in U$. Finally, the zero vector $\overrightarrow{0}$ of U is well-defined only if the zero vector $\overrightarrow{0}$ of V actually belongs to U. Let me collect all these conditions in one notion:

Definition 4.21. Let U be a subset of a vector space V. We say that U is a *subspace* (or *vector subspace*) of V if the following three conditions hold:

(a) We have $\overrightarrow{0} \in U$ (where $\overrightarrow{0}$ is the zero vector of V).

(b) We have $v+w \in U$ for all $v \in U$ and $w \in U$.

(c) We have $\lambda v \in U$ for all $\lambda \in \mathbb{R}$ and $v \in U$.

Note that condition **(b)** is often put into words as follows: "The subset U is closed under addition" (since it says, roughly speaking, that if you start with elements of U, then you cannot escape U by applying addition). Similarly, condition **(c)** is often stated as follows: "The subset U is closed under scaling". Hence, the subset U is a subspace of V if and only if U contains the zero vector, is closed under addition and is closed under scaling.

Proposition 4.22. Let U be a subspace of a vector space V. Then, U becomes a vector space, if we let it "inherit" the addition $+$, the scaling $\cdot$ and the zero vector $\overrightarrow{0}$ from V. Here, "inheriting" means that:

- we define the sum $v + w$ of two vectors $v \in U$ and $w \in U$ to be the result of adding v and w as elements of V.
- we define λv for $\lambda \in \mathbb{R}$ to $v \in U$ to be the result of scaling v by λ as element of V.
- we define the zero vector $\overrightarrow{0}$ of U as the zero vector $\overrightarrow{0}$ of V.

Proof of Proposition 4.22. The "inherited" operations on U are clearly well-defined. (For example, the condition **(b)** in Definition 4.21 shows that the "inherited" addition is actually an addition on U. Similarly, condition **(c)** shows that the "inherited" scaling is actually a scaling, and condition **(a)** shows that the "inherited" zero vector actually is in U.)

It remains to prove that the ten properties **(a)**, **(b)**, **(c)**, **(d)**, **(e)**, **(f)**, **(g)**, **(h)**, **(i)** and **(j)** from Definition 4.2 are satisfied for U instead of V. This is almost completely straightforward: We know that these ten properties are satisfied for V (since V is a vector space). Since the operations on U are "inherited" from V, this immediately yields that the nine properties **(a)**, **(b)**, **(c)**, **(e)**, **(f)**, **(g)**, **(h)**, **(i)** and **(j)** are also satisfied for U instead of V. However, this argument does not prove that property **(d)** is satisfied for U instead of V: We know that it holds for V, but we cannot immediately conclude that it must also hold for U instead of V (because there is no obvious reason why the element w whose existence is claimed in property **(d)** must belong to U). Fortunately, we can ignore property **(d)**, since Proposition 4.3 shows that this property is redundant. Hence, U is a vector space (since the nine properties **(a)**, **(b)**, **(c)**, **(e)**, **(f)**, **(g)**, **(h)**, **(i)** and **(j)** are satisfied for U instead of V). This proves Proposition 4.22. □

Remark 4.23. Some authors replace condition **(a)** in Definition 4.21 by the seemingly weaker condition that U be nonempty. However, this weaker condition is actually equivalent to condition **(a)** (as long as condition **(c)** is assumed). Indeed, if U is nonempty, then we can pick any $v \in U$, and then we can observe that $\overrightarrow{0} = 0v \in U$ (by condition **(c)**, applied to $\lambda = 0$); thus, condition **(a)** holds.

4.5. <DRAFT> Examples and constructions of subspaces

4.5.1. $\left\{ \overrightarrow{0} \right\}$ and V

Examples of subspaces, and general ways to construct subspaces, abound. Let us first go for the very lowest-hanging fruits:

Proposition 4.24. Let V be a vector space.
(a) The subset $\left\{ \overrightarrow{0} \right\}$ of V is a subspace of V.
(b) The subset V of V is a subspace of V.

Proof of Proposition 4.24. Straightforward, and left to the reader. (For instance, $\left\{ \overrightarrow{0} \right\}$ is closed under addition because the only possible sum of two elements of $\left\{ \overrightarrow{0} \right\}$ is $\overrightarrow{0} + \overrightarrow{0} = \overrightarrow{0}$, which of course belongs to $\left\{ \overrightarrow{0} \right\}$.) □

Please note that the subset $\left\{ \overrightarrow{0} \right\}$ in Proposition 4.24 **(a)** is **not** the empty set! (It contains the zero vector $\overrightarrow{0}$, whereas the empty set contains nothing.) That said, it is "as close to the empty set as a subspace of V can get": It contains only the zero vector $\overrightarrow{0}$, which any subspace of V is forced to contain. It is easy to see that $\left\{ \overrightarrow{0} \right\}$ is the intersection of all subspaces of V. (In contrast, the empty set $\varnothing$ is the intersection of all **subsets** of V.)

4.5.2. Some examples of subspaces of $\mathbb{R}^3$

Let us now get our hands dirty and verify some examples of subspaces and non-subspaces:

Example 4.25. Recall that $\mathbb{R}^n$ denotes the vector space of all column vectors of size n. (We defined this in Example 4.6.)
(a) Let A be the subset

$$\left\{ (x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1 - x_2 + 2x_3 = 0 \right\}$$

of $\mathbb{R}^3$. Then, A is a subspace of $\mathbb{R}^3$.
(b) Let B be the subset

$$\left\{ (x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1 = 2x_2 = 3x_3 \right\}$$

of $\mathbb{R}^3$. Then, B is a subspace of $\mathbb{R}^3$.
(c) Let C be the subset

$$\left\{ (x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1 - x_2 + 2x_3 = 1 \right\}$$

of $\mathbb{R}^3$. Then, C is **not** a subspace of $\mathbb{R}^3$.

(d) Let D be the subset

$$\left\{(u, 0, 2u+v)^T \mid u, v \in \mathbb{R}\right\}$$

of $\mathbb{R}^3$. (Note that the letter v stands for a number here, not for a vector.) Then, D is a subspace of $\mathbb{R}^3$.

(e) Let E be the subset

$$\left\{(u, 0, u+1)^T \mid u \in \mathbb{R}\right\}$$

of $\mathbb{R}^3$. Then, E is **not** a subspace of $\mathbb{R}^3$.

(f) Let F be the subset

$$\left\{(x_1, x_2, x_3)^T \mid x_1 x_2 x_3 = 0\right\}$$

of $\mathbb{R}^3$. Then, F is **not** a subspace of $\mathbb{R}^3$.

(g) Let G be the subset

$$\left\{(x_1, x_2, x_3)^T \mid x_1, x_2, x_3 \in \mathbb{Q}\right\}$$

of $\mathbb{R}^3$. (Recall that $\mathbb{Q}$ is the set of all rational numbers.) Then, G is **not** a subspace of $\mathbb{R}^3$.

(h) Let H be the subset

$$\left\{(u, 0, 2u+v+1)^T \mid u, v \in \mathbb{R}\right\}$$

of $\mathbb{R}^3$. Then, H is a subspace of $\mathbb{R}^3$. (Actually, $H = D$.)

Proof of Example 4.25. Recall how subspaces are defined (see Definition 4.21). Thus, in order to prove that some subset U of $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$, we must prove that it satisfies the three conditions **(a)**, **(b)** and **(c)** of Definition 4.21 (i.e., that it contains the zero vector, is closed under addition, and is closed under scaling). On the other hand, in order to prove that some subset U of $\mathbb{R}^3$ is **not** a subspace of $\mathbb{R}^3$, it suffices to prove that it fails **at least one** of these three conditions **(a)**, **(b)** and **(c)**. (Often, a subset U will fail two or all three of these conditions, but we do not need to check them all: Failing one is enough.) Let us now apply this strategy to the subsets A, B, C, D, E, F and G:

(a) We want to show that A is a subspace of $\mathbb{R}^3$. Thus, we need to show that A contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that A contains the zero vector: Recall that the zero vector of $\mathbb{R}^3$ is $0_{3\times 1} = (0,0,0)^T$. This vector $0_{3\times 1}$ lies in A if and only if $0 - 0 + 2 \cdot 0 = 0$ (by the definition

of A). Thus, $0_{3\times 1}$ lies in A (since $0-0+2\cdot 0=0$ holds). In other words, A contains the zero vector.

Proof that A is closed under addition: Let $v \in A$ and $w \in A$. We must prove that $v+w \in A$.

Write the vector $v \in A \subseteq \mathbb{R}^3$ in the form $v = (v_1, v_2, v_3)^T$ (for three real numbers $v_1, v_2, v_3 \in \mathbb{R}$). Since $(v_1, v_2, v_3)^T = v \in A$, we have $v_1 - v_2 + 2v_3 = 0$.

Write the vector $w \in A \subseteq \mathbb{R}^3$ in the form $w = (w_1, w_2, w_3)^T$ (for three real numbers $w_1, w_2, w_3 \in \mathbb{R}$). Since $(w_1, w_2, w_3)^T = w \in A$, we have $w_1 - w_2 + 2w_3 = 0$.

From $v = (v_1, v_2, v_3)^T$ and $w = (w_1, w_2, w_3)^T$, we obtain

$$v + w = (v_1, v_2, v_3)^T + (w_1, w_2, w_3)^T = (v_1 + w_1, v_2 + w_2, v_3 + w_3)^T$$

(since column vectors are added entry by entry). Thus, in order to prove that $v + w \in A$, we must show that $(v_1 + w_1) - (v_2 + w_2) + 2(v_3 + w_3) = 0$ (by the definition of A). But showing this is easy: Just notice that

$$\begin{aligned} &(v_1 + w_1) - (v_2 + w_2) + 2(v_3 + w_3) \\ &= \underbrace{(v_1 - v_2 + 2v_3)}_{=0} + \underbrace{(w_1 - w_2 + 2w_3)}_{=0} = 0 + 0 = 0. \end{aligned}$$

Thus, we have proven that $v + w \in A$. This completes the proof of the fact that A is closed under addition.

Proof that A is closed under scaling: Let $\lambda \in \mathbb{R}$ and $v \in A$. We must prove that $\lambda v \in A$.

Write the vector $v \in A \subseteq \mathbb{R}^3$ in the form $v = (v_1, v_2, v_3)^T$ (for three real numbers $v_1, v_2, v_3 \in \mathbb{R}$). Since $(v_1, v_2, v_3)^T = v \in A$, we have $v_1 - v_2 + 2v_3 = 0$.

From $v = (v_1, v_2, v_3)^T$, we obtain

$$\lambda v = \lambda (v_1, v_2, v_3)^T = (\lambda v_1, \lambda v_2, \lambda v_3)^T$$

(since column vectors are scaled entry by entry). Thus, in order to prove that $\lambda v \in A$, we must show that $\lambda v_1 - \lambda v_2 + 2\lambda v_3 = 0$ (by the definition of A). But showing this is easy: Just notice that

$$\lambda v_1 - \lambda v_2 + 2\lambda v_3 = \lambda \underbrace{(v_1 - v_2 + 2v_3)}_{=0} = \lambda 0 = 0.$$

Thus, we have proven that $\lambda v \in A$. This completes the proof of the fact that A is closed under scaling.

We now have shown that A contains the zero vector, is closed under addition, and is closed under scaling. Hence, A is a subspace of $\mathbb{R}^3$. Example 4.25 **(a)** is proven.

(b) The chain of equations $x_1 = 2x_2 = 3x_3$ in the definition of B is equivalent to "$x_1 = 2x_2$ and $2x_2 = 3x_3$". The proof of Example 4.25 **(b)** proceeds similarly to the above proof of Example 4.25 **(a)**, with the only difference that instead of the

single equation $x_1 - x_2 + 2x_3 = 0$, we now have to keep track of the two equations $x_1 = 2x_2$ and $2x_2 = 3x_3$. Let me only show one part of the proof, namely the verification that B is closed under addition:

Proof that B is closed under addition: Let $v \in B$ and $w \in B$. We must prove that $v + w \in B$.

Write the vector $v \in B \subseteq \mathbb{R}^3$ in the form $v = (v_1, v_2, v_3)^T$ (for three real numbers $v_1, v_2, v_3 \in \mathbb{R}$). Since $(v_1, v_2, v_3)^T = v \in B$, we have $v_1 = 2v_2$ and $2v_2 = 3v_3$.

Write the vector $w \in B \subseteq \mathbb{R}^3$ in the form $w = (w_1, w_2, w_3)^T$ (for three real numbers $w_1, w_2, w_3 \in \mathbb{R}$). Since $(w_1, w_2, w_3)^T = w \in B$, we have $w_1 = 2w_2$ and $2w_2 = 3w_3$.

From $v = (v_1, v_2, v_3)^T$ and $w = (w_1, w_2, w_3)^T$, we obtain

$$v + w = (v_1, v_2, v_3)^T + (w_1, w_2, w_3)^T = (v_1 + w_1, v_2 + w_2, v_3 + w_3)^T$$

(since column vectors are added entry by entry). Thus, in order to prove that $v + w \in B$, we must show that $v_1 + w_1 = 2(v_2 + w_2)$ and $2(v_2 + w_2) = 3(v_3 + w_3)$ (by the definition of B). But showing this is easy: Just notice that

$$\underbrace{v_1}_{=2v_2} + \underbrace{w_1}_{=2w_2} = 2v_2 + 2w_2 = 2(v_2 + w_2) \qquad \text{and}$$
$$2(v_2 + w_2) = \underbrace{2v_2}_{=3v_3} + \underbrace{2w_2}_{=3w_3} = 3v_3 + 3w_3 = 3(v_3 + w_3).$$

Thus, we have proven that $v + w \in B$. This completes the proof of the fact that B is closed under addition.

Proving that B contains the zero vector and is closed under scaling is left to the reader. (As I have said, the proofs are similar to the corresponding proofs for A, and the changes that need to be done are essentially the same as we did in the proof that B is closed under addition.) Thus, Example 4.25 **(b)** is proven.

(c) The zero vector $\overrightarrow{0} = 0_{3\times 1} = (0, 0, 0)^T$ of $\mathbb{R}^3$ does not belong to C (since it does not satisfy $0 - 0 + 2 \cdot 0 = 1$). Hence, C does not contain the zero vector. As a consequence, C cannot be a subspace of $\mathbb{R}^3$. Thus, Example 4.25 **(c)** is proven.

(Notice that C furthermore fails to be closed under addition, and also fails to be closed under scaling. Each of these failures gives another reason why C is not a subspace of $\mathbb{R}^3$. We just chose to use the simplest reason instead, namely that it fails to contain the zero vector.)

(d) We want to show that D is a subspace of $\mathbb{R}^3$. Thus, we need to show that D contains the zero vector, is closed under addition, and is closed under scaling. Before we do this, let us rewrite the definition of D as follows:

$$\begin{aligned} D &= \left\{ (u, 0, 2u + v)^T \mid u, v \in \mathbb{R} \right\} \\ &= \left\{ (x, 0, 2x + y)^T \mid x, y \in \mathbb{R} \right\} \end{aligned} \tag{157}$$

(here, we have renamed the indices u and v as x and y). (The reason for rewriting D this way was to get rid of the letters u and v; indeed, we are going to reuse these letters for vectors.)

Proof that D contains the zero vector: Recall that the zero vector of $\mathbb{R}^3$ is $0_{3\times 1} = (0,0,0)^T$. This vector $0_{3\times 1}$ has the form $(u, 0, 2u+v)^T$ for some $u, v \in \mathbb{R}$ (namely, for $u = 0$ and $v = 0$). Thus, it belongs to D (by the definition of D). In other words, D contains the zero vector.

Proof that D is closed under addition: Let $v \in D$ and $w \in D$. We must prove that $v + w \in D$.

We have $v \in D = \left\{ (x, 0, 2x+y)^T \mid x, y \in \mathbb{R} \right\}$ (by (157)). In other words, v has the form $(x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$. Fix two such x, y, and denote them by x_v, y_v. (I do not want to simply call them x, y, because I will introduce two other such x, y shortly.) Thus, x_v, y_v are real numbers and satisfy $v = (x_v, 0, 2x_v + y_v)^T$.

We have $w \in D = \left\{ (x, 0, 2x+y)^T \mid x, y \in \mathbb{R} \right\}$ (by (157)). In other words, w has the form $(x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$. Fix two such x, y, and denote them by x_w, y_w. Thus, x_w, y_w are real numbers and satisfy $w = (x_w, 0, 2x_w + y_w)^T$.

From $v = (x_v, 0, 2x_v + y_v)^T$ and $w = (x_w, 0, 2x_w + y_w)^T$, we obtain

$$\begin{aligned} v + w &= (x_v, 0, 2x_v + y_v)^T + (x_w, 0, 2x_w + y_w)^T \\ &= \left(x_v + x_w, \underbrace{0+0}_{=0}, \underbrace{(2x_v + y_v) + (2x_w + y_w)}_{=2(x_v+x_w)+(y_v+y_w)} \right)^T \\ &\qquad \left(\text{since column vectors are added entry by entry}\right) \\ &= \left(x_v + x_w, 0, 2\left(x_v + x_w\right) + \left(y_v + y_w\right) \right)^T. \end{aligned}$$

Thus, the vector $v + w$ has the form $(x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$ (namely, for $x = x_v + x_w$ and $y = y_v + y_w$). In other words, $v + w \in \left\{ (x, 0, 2x+y)^T \mid x, y \in \mathbb{R} \right\}$. In view of (157), this rewrites as $v + w \in D$. This completes the proof of the fact that D is closed under addition.

Proof that D is closed under scaling: Let $\lambda \in \mathbb{R}$ and $v \in D$. We must prove that $\lambda v \in D$.

We have $v \in D = \left\{ (x, 0, 2x+y)^T \mid x, y \in \mathbb{R} \right\}$ (by (157)). In other words, v has the form $(x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$. Fix two such x, y, and denote them by x_v, y_v. Thus, x_v, y_v are real numbers and satisfy $v = (x_v, 0, 2x_v + y_v)^T$.

Hence,

$$\lambda v = \lambda \left(x_v, 0, 2x_v + y_v\right)^T = \left(\lambda x_v, \underbrace{\lambda 0}_{=0}, \underbrace{\lambda\left(2x_v + y_v\right)}_{=2\cdot\lambda x_v + \lambda y_v}\right)^T$$

$$\left(\text{since column vectors are scaled entry by entry}\right)$$

$$= \left(\lambda x_v, 0, 2\cdot\lambda x_v + \lambda y_v\right)^T.$$

Thus, the vector λv has the form $\left(x, 0, 2x + y\right)^T$ for some $x, y \in \mathbb{R}$ (namely, for $x = \lambda x_v$ and $y = \lambda y_v$). In other words, $\lambda v \in \left\{\left(x, 0, 2x + y\right)^T \mid x, y \in \mathbb{R}\right\}$. In view of (157), this rewrites as $\lambda v \in D$. This completes the proof of the fact that D is closed under scaling.

We now have shown that D contains the zero vector, is closed under addition, and is closed under scaling. Hence, D is a subspace of $\mathbb{R}^3$. Example 4.25 **(d)** is proven.

(e) There exists no $u \in \mathbb{R}$ satisfying $\left(0, 0, 0\right)^T = \left(u, 0, u + 1\right)^T$ (because any such u would have to satisfy $0 = u$ and $0 = u + 1$ at the same time; but this is clearly impossible). In other words, $\left(0, 0, 0\right)^T$ is **not** a vector of the form $\left(u, 0, u + 1\right)^T$ with $u \in \mathbb{R}$. In other words, $\left(0, 0, 0\right)^T \notin \left\{\left(u, 0, u + 1\right)^T \mid u \in \mathbb{R}\right\}$. Since $E = \left\{\left(u, 0, u + 1\right)^T \mid u \in \mathbb{R}\right\}$, this rewrites as $\left(0, 0, 0\right)^T \notin E$.

But recall that the zero vector of the vector space $\mathbb{R}^3$ is $\left(0, 0, 0\right)^T$. Hence, this zero vector does not lie in E (since $\left(0, 0, 0\right)^T \notin E$). In other words, E does not contain the zero vector. Consequently, E cannot be a subspace of $\mathbb{R}^3$. Thus, Example 4.25 **(e)** is proven.

(f) The subset F is not closed under addition. In fact, if we set $v = \left(1, 0, 0\right)^T$ and $w = \left(0, 1, 1\right)^T$, then $v \in F$ (since $1 \cdot 0 \cdot 0 = 0$) and $w \in F$ (since $0 \cdot 1 \cdot 1 = 0$), but $v + w = \left(1, 1, 1\right)^T \notin F$ (since $1 \cdot 1 \cdot 1 \neq 0$).

Since F is not closed under addition, F is not a subspace of $\mathbb{R}^3$ (even though, as the reader can easily check, F contains the zero vector and is closed under scaling). This proves Example 4.25 **(f)**.

(g) The subset G is not closed under scaling. In fact, if we set $\lambda = \sqrt{2}$ and $v = \left(1, 0, 0\right)^T$, then $v \in G$ (since $1, 0, 0 \in \mathbb{Q}$), but $\lambda v = \left(\sqrt{2}, 0, 0\right)^T \notin G$ (since not all of $\sqrt{2}, 0, 0$ belong to $\mathbb{Q}$).

Since G is not closed under scaling, G is not a subspace of $\mathbb{R}^3$ (even though, as the reader can easily check, G contains the zero vector and is closed under addition). This proves Example 4.25 **(g)**.

(h) I claim that $H = D$.

In order to prove this, it suffices to show that $H \subseteq D$ and $D \subseteq H$.

Proof of $H \subseteq D$: Let $h \in H$. Thus, $h \in H = \left\{\left(u, 0, 2u + v + 1\right)^T \mid u, v \in \mathbb{R}\right\}$. In

other words, h has the form $(u, 0, 2u+v+1)^T$ for some $u, v \in \mathbb{R}$. Consider these u, v.

Now, $h = (u, 0, 2u+v+1)^T$. Hence, h has the form $(x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$ (namely, for $x = u$ and $y = v+1$). In other words, $h \in \left\{(x, 0, 2x+y)^T \mid x, y \in \mathbb{R}\right\}$. In view of (157), this rewrites as $h \in D$.

Now, we have proven that $h \in D$ for each $h \in H$. In other words, $H \subseteq D$.

Proof of $D \subseteq H$: Let $d \in D$. Then, $d \in D = \left\{(x, 0, 2x+y)^T \mid x, y \in \mathbb{R}\right\}$ (by (157)). In other words, d has the form $d = (x, 0, 2x+y)^T$ for some $x, y \in \mathbb{R}$. Consider these x, y.

Now, $d = \left(x, 0, 2x + \underbrace{y}_{=(y-1)+1}\right)^T = (x, 0, 2x+(y-1)+1)^T$. Hence, d has the form $(u, 0, 2u+v+1)^T$ for some $u, v \in \mathbb{R}$ (namely, for $u = x$ and $v = y-1$). In other words, $d \in \left\{(u, 0, 2u+v+1)^T \mid u, v \in \mathbb{R}\right\}$. In light of $\left\{(u, 0, 2u+v+1)^T \mid u, v \in \mathbb{R}\right\} = H$, this rewrites as $d \in H$.

We have now shown that $d \in H$ for each $d \in D$. In other words, $D \subseteq H$.

Combining the relations $H \subseteq D$ and $D \subseteq H$, we obtain $H = D$. Since D is a subspace of $\mathbb{R}^3$ (this was proven in Example 4.25 **(d)**), we thus know that H is a subspace of $\mathbb{R}^3$. Example 4.25 **(h)** is proven. □

4.5.3. The kernel of a matrix

You might have spotted some patterns in Example 4.25. For instance, you have noticed that the subsets A and B were "carved out" of $\mathbb{R}^3$ by systems of linear equations with no constant terms (i.e., linear equations of the form $a_1x_1 + a_2x_2 + a_3x_3 = 0$, or equivalent to such equations). This is a general pattern: Any subset of $\mathbb{R}^n$ "carved out" by a system of linear equations with no constant terms[105] is a subspace of $\mathbb{R}^n$. Let me state (and prove) this fact in a slightly cleaner form, replacing the system of linear equations by a matrix equation[106]:

Proposition 4.26. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $m \times n$-matrix. Let $\operatorname{Ker} A$ denote the subset

$$\{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\}$$

[105] Such systems are called *homogeneous systems* in [OlvSha06, §1.8].

[106] In fact, recall (from Proposition 3.2) that a system of linear equations in n unknowns $x_1, x_2, \ldots, x_n$ can be rewritten as a single equation $Ax = b$ for the column vector $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. When the system has no constant terms, the vector b is the zero vector $0_{m\times 1}$, and so the single equation $Ax = b$ becomes $Ax = 0_{m\times 1}$.

of $\mathbb{R}^n$. (Recall that $\mathbb{R}^n$ stands for the vector space $\mathbb{R}^{n\times 1}$ of all column vectors of size n.) Then, $\operatorname{Ker} A$ is a subspace of $\mathbb{R}^n$.

Proposition 4.26 generalizes both Example 4.25 **(a)** and Example 4.25 **(b)**.[107] The proof of Proposition 4.26 is essentially like the proofs given for those examples, but shorter because working with matrices is simpler than dealing with lots of single entries:

Proof of Proposition 4.26. We want to show that $\operatorname{Ker} A$ is a subspace of $\mathbb{R}^n$. Thus, we need to show that $\operatorname{Ker} A$ contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that $\operatorname{Ker} A$ *contains the zero vector:* We have $A0_{n\times 1} = 0_{m\times 1}$. Hence, $0_{n\times 1} \in \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\} = \operatorname{Ker} A$. In other words, $\operatorname{Ker} A$ contains the zero vector (since $0_{n\times 1}$ is the zero vector of the vector space $\mathbb{R}^n$).

Proof that $\operatorname{Ker} A$ *is closed under addition:* Let $v \in \operatorname{Ker} A$ and $w \in \operatorname{Ker} A$. We must prove that $v + w \in \operatorname{Ker} A$.

We have $v \in \operatorname{Ker} A = \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\}$. In other words, $Av = 0_{m\times 1}$. Similarly, $Aw = 0_{m\times 1}$. Now, $A(v+w) = \underbrace{Av}_{=0_{m\times 1}} + \underbrace{Aw}_{=0_{m\times 1}} = 0_{m\times 1} + 0_{m\times 1} = 0_{m\times 1}$. Hence, $v + w \in \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\} = \operatorname{Ker} A$. This completes the proof of the fact that $\operatorname{Ker} A$ is closed under addition.

Proof that $\operatorname{Ker} A$ *is closed under scaling:* Let $\lambda \in \mathbb{R}$ and $v \in \operatorname{Ker} A$. We must prove that $\lambda v \in \operatorname{Ker} A$.

We have $v \in \operatorname{Ker} A = \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\}$. In other words, $Av = 0_{m\times 1}$. Now, $A(\lambda v) = \lambda \underbrace{Av}_{=0_{m\times 1}} = \lambda 0_{m\times 1} = 0_{m\times 1}$. Hence, $\lambda v \in \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\} = \operatorname{Ker} A$. This completes the proof of the fact that $\operatorname{Ker} A$ is closed under scaling.

We now have shown that $\operatorname{Ker} A$ contains the zero vector, is closed under addition, and is closed under scaling. Hence, $\operatorname{Ker} A$ is a subspace of $\mathbb{R}^n$. This proves Proposition 4.26. □

Definition 4.27. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let A be an $m \times n$-matrix. The subspace $\operatorname{Ker} A$ of $\mathbb{R}^n$ introduced in Proposition 4.26 is called the *kernel* (or the *nullspace*) of A. (Some authors also denote it by $\ker A$.)

We will see some examples of kernels later; for now, let me only show one particularly simple example, namely the kernel of a zero matrix:

[107] Indeed, you can easily check that the subset A of $\mathbb{R}^3$ in Example 4.25 **(a)** is $\operatorname{Ker}\begin{pmatrix} 1 & -1 & 2 \end{pmatrix}$, whereas the subset B of $\mathbb{R}^3$ in Example 4.25 **(b)** is $\operatorname{Ker}\begin{pmatrix} 1 & -2 & 0 \\ 0 & 2 & -3 \end{pmatrix}$ (since the condition $x_1 = 2x_2 = 3x_3$ defining B is equivalent to the system $\begin{cases} x_1 - 2x_2 = 0; \\ 2x_2 - 3x_3 = 0 \end{cases}$).

Proposition 4.28. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, $\operatorname{Ker}(0_{m\times n}) = \mathbb{R}^n$.

Proof of Proposition 4.28. The definition of $\operatorname{Ker}(0_{m\times n})$ yields $\operatorname{Ker}(0_{m\times n}) = \{x \in \mathbb{R}^n \mid 0_{m\times n}x = 0_{m\times 1}\}$. But **every** $x \in \mathbb{R}^n$ satisfies $0_{m\times n}x = 0_{m\times 1}$. Hence, $\{x \in \mathbb{R}^n \mid 0_{m\times n}x = 0_{m\times 1}\} = \{x \in \mathbb{R}^n\} = \mathbb{R}^n$. Thus, $\operatorname{Ker}(0_{m\times n}) = \{x \in \mathbb{R}^n \mid 0_{m\times n}x = 0_{m\times 1}\} = \mathbb{R}^n$. This proves Proposition 4.28. □

4.5.4. The span of k vectors

Here is another fairly general construction of a subspace:

Definition 4.29. Let V be a vector space. Let $v_1, v_2, \ldots, v_k$ be finitely many vectors in V.

(a) A *linear combination* of $v_1, v_2, \ldots, v_k$ means a vector of the form

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \qquad \text{with } \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}.$$

(In words: A *linear combination* of $v_1, v_2, \ldots, v_k$ means a vector obtained by first scaling each of $v_1, v_2, \ldots, v_k$ by some real number, and then adding the results.)

(b) The *span* of $v_1, v_2, \ldots, v_k$ is the set of all linear combinations of $v_1, v_2, \ldots, v_k$. This set is denoted by $\operatorname{span}(v_1, v_2, \ldots, v_k)$. Many authors also denote it by $\operatorname{span}(v_1, v_2, \ldots, v_k)$, but this notation risks confusion with the inner product of two vectors v and w (which will be introduced later, and which will be denoted by $\langle v, w\rangle$).

Note that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ is defined as the set of all linear combinations of $v_1, v_2, \ldots, v_k$; but the latter linear combinations are the vectors of the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ with $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. Hence, $\operatorname{span}(v_1, v_2, \ldots, v_k)$ is the set of all vectors of the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ with $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$. In other words,

$$\operatorname{span}(v_1, v_2, \ldots, v_k) = \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}. \quad (158)$$

Remark 4.30. The span $\operatorname{span}()$ (that is, the span of no vectors) is the subspace $\left\{\overrightarrow{0}\right\}$ of V (not the empty set!). This is because $\overrightarrow{0}$ is a linear combination of an empty list of vectors (being the empty sum). This might appear somewhat counterintuitive, but is natural and important.

Proposition 4.31. Let V be a vector space. Let $v_1, v_2, \ldots, v_k$ be finitely many vectors in V. Then, $\operatorname{span}(v_1, v_2, \ldots, v_k)$ (that is, the span of $v_1, v_2, \ldots, v_k$) is a subspace of V.

Proof of Proposition 4.31. We want to show that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ is a subspace of V. Thus, we need to show that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ *contains the zero vector:* We have $0v_1 + 0v_2 + \cdots + 0v_k = \overrightarrow{0} + \overrightarrow{0} + \cdots + \overrightarrow{0} = \overrightarrow{0}$. Thus, $\overrightarrow{0} = 0v_1 + 0v_2 + \cdots + 0v_k$. Hence, the vector $\overrightarrow{0}$ has the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ for some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ (namely, for $\lambda_i = 0$). In other words, $\overrightarrow{0}$ is a linear combination of $v_1, v_2, \ldots, v_k$. In other words, $\overrightarrow{0} \in \operatorname{span}(v_1, v_2, \ldots, v_k)$ (since $\operatorname{span}(v_1, v_2, \ldots, v_k)$ is the set of all linear combinations of $v_1, v_2, \ldots, v_k$). In other words, $\operatorname{span}(v_1, v_2, \ldots, v_k)$ contains the zero vector (since $\overrightarrow{0}$ is the zero vector of the vector space V).

Proof that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ *is closed under addition:* Let $v \in \operatorname{span}(v_1, v_2, \ldots, v_k)$ and $w \in \operatorname{span}(v_1, v_2, \ldots, v_k)$. We must prove that $v + w \in \operatorname{span}(v_1, v_2, \ldots, v_k)$.

We have

$$v \in \operatorname{span}(v_1, v_2, \ldots, v_k) = \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}$$

(by (158)). In other words, there exist some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ satisfying $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$. Fix these $\lambda_1, \lambda_2, \ldots, \lambda_k$, and denote them by $\mu_1, \mu_2, \ldots, \mu_k$. Thus, $\mu_1, \mu_2, \ldots, \mu_k$ are elements of $\mathbb{R}$ and satisfy $v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k$.

We have

$$w \in \operatorname{span}(v_1, v_2, \ldots, v_k) = \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}$$

(by (158)). In other words, there exist some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ satisfying $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$. Fix these $\lambda_1, \lambda_2, \ldots, \lambda_k$, and denote them by $\kappa_1, \kappa_2, \ldots, \kappa_k$. Thus, $\kappa_1, \kappa_2, \ldots, \kappa_k$ are elements of $\mathbb{R}$ and satisfy $w = \kappa_1 v_1 + \kappa_2 v_2 + \cdots + \kappa_k v_k$.

Adding the equalities $v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k$ and $w = \kappa_1 v_1 + \kappa_2 v_2 + \cdots + \kappa_k v_k$, we obtain

$$\begin{aligned} v + w &= (\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k) + (\kappa_1 v_1 + \kappa_2 v_2 + \cdots + \kappa_k v_k) \\ &= (\mu_1 v_1 + \kappa_1 v_1) + (\mu_2 v_2 + \kappa_2 v_2) + \cdots + (\mu_k v_k + \kappa_k v_k) \\ &= (\mu_1 + \kappa_1) v_1 + (\mu_2 + \kappa_2) v_2 + \cdots + (\mu_k + \kappa_k) v_k. \end{aligned}$$

Hence, the vector $v + w$ has the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ for some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ (namely, for $\lambda_i = \mu_i + \kappa_i$). In other words,

$$v + w \in \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}.$$

In view of (158), this rewrites as $v + w \in \operatorname{span}(v_1, v_2, \ldots, v_k)$. This completes the proof of the fact that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ is closed under addition.

Proof that $\operatorname{span}(v_1, v_2, \ldots, v_k)$ *is closed under scaling:* Let $\lambda \in \mathbb{R}$ and $v \in \operatorname{span}(v_1, v_2, \ldots, v_k)$. We must prove that $\lambda v \in \operatorname{span}(v_1, v_2, \ldots, v_k)$.

We have

$$v \in \operatorname{span}(v_1, v_2, \ldots, v_k) = \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}$$

(by (158)). In other words, there exist some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ satisfying $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$. Fix these $\lambda_1, \lambda_2, \ldots, \lambda_k$, and denote them by $\mu_1, \mu_2, \ldots, \mu_k$. Thus, $\mu_1, \mu_2, \ldots, \mu_k$ are elements of $\mathbb{R}$ and satisfy $v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k$.

Multiplying both sides of the equality $v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k$ by λ, we find

$$\lambda v = \lambda \left(\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k \right) = \lambda \mu_1 v_1 + \lambda \mu_2 v_2 + \cdots + \lambda \mu_k v_k.$$

Hence, the vector λv has the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ for some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ (namely, for $\lambda_i = \lambda \mu_i$). In other words,

$$\lambda v \in \left\{ \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \ \mid \ \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R} \right\}.$$

In view of (158), this rewrites as $\lambda v \in \operatorname{span}\left(v_1, v_2, \ldots, v_k\right)$. This completes the proof of the fact that $\operatorname{span}\left(v_1, v_2, \ldots, v_k\right)$ is closed under scaling.

We now have shown that $\operatorname{span}\left(v_1, v_2, \ldots, v_k\right)$ contains the zero vector, is closed under addition, and is closed under scaling. Hence, $\operatorname{span}\left(v_1, v_2, \ldots, v_k\right)$ is a subspace of V. This proves Proposition 4.31. □

Example 4.32. It is easy to construct examples of spans. Here is one:

Let $x = (1,0,2)^T$ and $y = (0,0,1)^T$. Then, x and y are two vectors in $\mathbb{R}^3$. What is their span $\operatorname{span}(x,y)$? The equality (158) (applied to $V = \mathbb{R}^3$, $k = 2$ and $(v_1, v_2, \ldots, v_k) = (x, y)$) yields

$$\operatorname{span}(x,y) = \left\{ \lambda_1 x + \lambda_2 y \ \mid \ \lambda_1, \lambda_2 \in \mathbb{R} \right\} = \left\{ u \underbrace{x}_{=(1,0,2)^T} + v \underbrace{y}_{=(0,0,1)^T} \ \mid \ u, v \in \mathbb{R} \right\}$$

(here, we have renamed the indices λ_1 and λ_2 as u and v)

$$= \left\{ \underbrace{u (1,0,2)^T + v (0,0,1)^T}_{\substack{=(u\cdot 1+v\cdot 0, u\cdot 0+v\cdot 0, u\cdot 2+v\cdot 1)^T \\ =(u,0,2u+v)^T}} \ \mid \ u, v \in \mathbb{R} \right\}$$

$$= \left\{ (u, 0, 2u+v)^T \ \mid \ u, v \in \mathbb{R} \right\}.$$

This is precisely the set D defined in Example 4.25 **(d)**. In particular, this shows once again that the latter set D is a subspace of $\mathbb{R}^3$ (since Proposition 4.31 shows that $\operatorname{span}(x,y)$ is a subspace of $\mathbb{R}^3$). Thus, we have found a new proof of the claim of Example 4.25 **(d)**.

The following property of spans is fundamental:

Proposition 4.33. Let V be a vector space. Let $v_1, v_2, \ldots, v_k$ be finitely many vectors in V.

(a) The vectors $v_1, v_2, \ldots, v_k$ belong to the span $\operatorname{span}\left(v_1, v_2, \ldots, v_k\right)$.

(b) If U is a subspace of V that contains the vectors $v_1, v_2, \ldots, v_k$, then $\operatorname{span}\left(v_1, v_2, \ldots, v_k\right) \subseteq U$.

Propositions 4.31 and 4.33 are often stated together in one single laconic sentence: "The span span $(v_1, v_2, \ldots, v_k)$ of k vectors $v_1, v_2, \ldots, v_k$ is the smallest subspace of V that contains $v_1, v_2, \ldots, v_k$" (where "smallest" means that any other subspace of V that contains $v_1, v_2, \ldots, v_k$ must contain span $(v_1, v_2, \ldots, v_k)$ as a subset). This sentence includes the claim of Proposition 4.31 (since it says that the span span $(v_1, v_2, \ldots, v_k)$ is a subspace of V), the claim of Proposition 4.33 **(a)** (since it says that this span contains $v_1, v_2, \ldots, v_k$), and the claim of Proposition 4.33 **(b)** (since it says that this span is the **smallest** subspace of V that contains $v_1, v_2, \ldots, v_k$).

We shall prove Proposition 4.33 later (in Section 4.6).

How to actually compute the kernel of a matrix? The next three examples illustrate a method:

Example 4.34. Let $n = 4$ and $m = 3$. Let A be the $m \times n$-matrix $\begin{pmatrix} 1 & 3 & 0 & 1 \\ 2 & 0 & 1 & 2 \\ 0 & 6 & -2 & 0 \end{pmatrix}$. Let us find the kernel $\operatorname{Ker} A$ of A. Proposition 4.26 tells us that $\operatorname{Ker} A$ will be a subspace of $\mathbb{R}^n = \mathbb{R}^4$; but it does not tell us which one. To find out, we will have to do some work.

The definition of $\operatorname{Ker} A$ yields

$$\operatorname{Ker} A = \{x \in \mathbb{R}^n \mid Ax = 0_{m\times 1}\} = \left\{x \in \mathbb{R}^4 \mid Ax = 0_{3\times 1}\right\}$$

(since $n = 4$ and $m = 3$). Hence, finding $\operatorname{Ker} A$ means finding all vectors $x \in \mathbb{R}^4$ satisfying $Ax = 0_{3\times 1}$. In other words, it means solving the equation $Ax = 0_{3\times 1}$. We know how to do this using Gaussian elimination:

Let $x = (x_1, x_2, x_3, x_4)^T \in \mathbb{R}^4$. Then, $Ax = 0_{3\times 1}$ rewrites as

$$\begin{pmatrix} 1 & 3 & 0 & 1 \\ 2 & 0 & 1 & 2 \\ 0 & 6 & -2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is equivalent to the system of equations

$$\begin{cases} 1x_1 + 3x_2 + 0x_3 + 1x_4 = 0; \\ 2x_1 + 0x_2 + 1x_3 + 2x_4 = 0; \\ 0x_1 + 6x_2 + (-2)\, x_3 + 0x_4 = 0 \end{cases}.$$

The solutions x of this system are precisely the vectors of the form

$$x = \begin{pmatrix} -r \\ 0 \\ 0 \\ r \end{pmatrix} \qquad \text{for } r \in \mathbb{R}.$$

Thus,

$$\begin{aligned}
\operatorname{Ker} A &= \left\{ x \in \mathbb{R}^4 \mid Ax = 0_{3\times 1} \right\} \\
&= \left\{ \begin{pmatrix} -r \\ 0 \\ 0 \\ r \end{pmatrix} \mid r \in \mathbb{R} \right\} \qquad (159) \\
&= \left\{ r \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \mid r \in \mathbb{R} \right\} \qquad (160) \\
&\qquad \left(\text{since } \begin{pmatrix} -r \\ 0 \\ 0 \\ r \end{pmatrix} = r \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ for each } r \in \mathbb{R} \right) \\
&= \operatorname{span} \left(\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right).
\end{aligned}$$

We have thus written $\operatorname{Ker} A$ as a span of vectors (in our case: one vector) in $\mathbb{R}^4$. This is not the only way to do so, but it is probably the most explicit and simplest way.

Example 4.35. Let $n = 3$ and $m = 3$. Let A be the $m \times n$-matrix $\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 3 & 6 & 6 \end{pmatrix}$. Let us find the kernel $\operatorname{Ker} A$ of A.

The method is the same as in Example 4.34, but the last few steps are a bit more complicated. Again, finding $\operatorname{Ker} A$ means solving the equation $Ax = 0_{3\times 1}$ (but this time, x is to lie in $\mathbb{R}^3$).

Let $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$. Then, $Ax = 0_{3\times 1}$ rewrites as

$$\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 3 & 6 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is equivalent to the system of equations

$$\begin{cases} 1x_1 + 2x_2 + 2x_3 = 0; \\ 2x_1 + 4x_2 + 4x_3 = 0; \\ 3x_1 + 6x_2 + 6x_3 = 0 \end{cases}.$$

The solutions x of this system are precisely the vectors of the form

$$x = \begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix} \qquad \text{for } r,s \in \mathbb{R}.$$

Thus,

$$\begin{aligned}
\operatorname{Ker} A &= \left\{ x \in \mathbb{R}^3 \mid Ax = 0_{3\times 1} \right\} \\
&= \left\{ \begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix} \mid r,s \in \mathbb{R} \right\} \qquad (161) \\
&= \left\{ r\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \mid r,s \in \mathbb{R} \right\} \qquad (162) \\
&\qquad \left(\text{since } \begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix} = r\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \text{ for each } r,s \in \mathbb{R} \right) \\
&= \operatorname{span}\left(\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \right).
\end{aligned}$$

We have thus written $\operatorname{Ker} A$ as a span of vectors (in our case: two vectors) in $\mathbb{R}^3$. Again, there are various other ways to do so.

Let me explain how I got from (161) to (162). I have rewritten $\begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix}$ as $r\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$. The purpose of this step was to "push the variables r and s out of the column vector", so that the result would be of the form "r times a **constant** column vector plus s times a **constant** column vector". And the ultimate purpose of this was to write $\operatorname{Ker} A$ as a span of two column vectors.

You should be able to do such transformations yourself! To illustrate the method in a bit more detail, let me state it as a two-step procedure:

1. First, decompose the vector $\begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix}$ into a sum of two vectors by isolating all multiples of r into one vector and all multiples of s into another. Thus,
$$\begin{pmatrix} -2r-2s \\ r \\ s \end{pmatrix} = \begin{pmatrix} -2r \\ r \\ 0 \end{pmatrix} + \begin{pmatrix} -2s \\ 0 \\ s \end{pmatrix}.$$

2. Now, push the factor r out of the first vector (i.e., rewrite $\begin{pmatrix} -2r \\ r \\ 0 \end{pmatrix}$ as $r\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$), and push the factor s out of the second vector (i.e., rewrite $\begin{pmatrix} -2s \\ 0 \\ s \end{pmatrix}$ as $s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$). The result is $r\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$.

In hindsight, the passage from (159) to (160) has been another instance of such a transformation. Namely, I have rewritten $\begin{pmatrix} -r \\ 0 \\ 0 \\ r \end{pmatrix}$ as $r\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ in order to "push the r out of the column vector". Of course, this was simpler because we had only one variable (hence, no need to isolate) and therefore only one constant vector.

Here is another example of such a transformation:

$$\begin{pmatrix} 5r+2s+3t \\ t \\ -2r+3s \\ s \\ -3r \\ r \\ 0 \end{pmatrix} = r\begin{pmatrix} 5 \\ 0 \\ -2 \\ 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 2 \\ 0 \\ 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Example 4.36. Let $n = 2$ and $m = 2$. Let A be the $m \times n$-matrix $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Let us find the kernel $\operatorname{Ker} A$ of A.

The method is the same as in Example 4.34, and this time we will actually have it a lot easier, as long as we do not get confused by empty lists and zero vectors. Finding $\operatorname{Ker} A$ means solving the equation $Ax = 0_{2\times 1}$ for $x \in \mathbb{R}^2$.

Let $x = (x_1, x_2)^T \in \mathbb{R}^2$. Then, $Ax = 0_{2\times 1}$ rewrites as

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This is equivalent to the system of equations

$$\begin{cases} 1x_1 + 2x_2 = 0; \\ 2x_1 + 3x_2 = 0 \end{cases}.$$

The solutions x of this system are precisely the vectors of the form

$$x = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus,

$$\operatorname{Ker} A = \left\{ x \in \mathbb{R}^2 \mid Ax = 0_{2\times 1} \right\} = \left\{ \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{=0_{2\times 1}} \right\} = \left\{ 0_{2\times 1} \right\}.$$

This is as simple a result as we can get, but if we actually want to rewrite $\operatorname{Ker} A$ as a span, we can achieve this by proceeding as in Examples 4.34 and 4.35:

$$\begin{aligned} \operatorname{Ker} A &= \left\{ x \in \mathbb{R}^2 \mid Ax = 0_{2\times 1} \right\} = \left\{ \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{=(\text{empty sum})} \right\} \\ &= \{(\text{empty sum})\} = \operatorname{span}() \qquad (\text{the span of no vectors}). \end{aligned}$$

Here, we have "decomposed" the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ into an empty sum, and realized that the set consisting of the empty sum is precisely the set of all linear combinations of no vectors, i.e., the span of no vectors. Thus, $\operatorname{Ker} A$ is the span of no vectors.

4.5.5. The image of a matrix

A further way to construct subspaces is the following:

Definition 4.37. Let A be an $n \times m$-matrix. We define $A\mathbb{R}^m$ to be the subset $\{Ax \mid x \in \mathbb{R}^m\}$ of $\mathbb{R}^n$. This subset $A\mathbb{R}^m$ is called the *column space* (or the *image*, or the *range*) of A.

The name "column space" for $A\mathbb{R}^m$ in Definition 4.37 might appear strange at first, but we will soon see why it is justified: In Proposition 4.38 **(a)** below, we will show that it is the span of the columns of A. The name "image" is due to the fact that $A\mathbb{R}^m$ is the image of the map $\mathbb{R}^m \to \mathbb{R}^n$, $x \mapsto Ax$.

Various authors have different notations for the column space $A\mathbb{R}^m$ of a matrix A. It is called $\operatorname{rng} A$ in [OlvSha06]; other notations for it include $\operatorname{Range} A$, $\operatorname{Im} A$, and $A(\mathbb{R}^m)$.

Proposition 4.38. Let A be an $n \times m$-matrix.
(a) We have $A\mathbb{R}^m = \operatorname{span}(\operatorname{col}_1 A, \operatorname{col}_2 A, \ldots, \operatorname{col}_m A)$.
(b) The set $A\mathbb{R}^m$ is a subspace of $\mathbb{R}^n$.

Before we prove this, let us show a simple lemma:

Lemma 4.39. Let A be an $n \times m$-matrix. Let $(x_1, x_2, \ldots, x_m)^T \in \mathbb{R}^m$. Then,

$$A\,(x_1, x_2, \ldots, x_m)^T = x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A.$$

Proof of Lemma 4.39. For each $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$, we have

$$\left(\operatorname{col}_j A\right)_{i,1} = A_{i,j} \tag{163}$$

(because $\operatorname{col}_j A$ is the j-th column of A, and thus its i-th entry is the (i,j)-th entry of A). Now, for each $i \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned}
&(x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A)_{i,1} \\
&= \underbrace{(x_1 \operatorname{col}_1 A)_{i,1}}_{\substack{=x_1(\operatorname{col}_1 A)_{i,1} \\ \text{(since matrices are} \\ \text{scaled entry by entry)}}} + \underbrace{(x_2 \operatorname{col}_2 A)_{i,1}}_{\substack{=x_2(\operatorname{col}_2 A)_{i,1} \\ \text{(since matrices are} \\ \text{scaled entry by entry)}}} + \cdots + \underbrace{(x_m \operatorname{col}_m A)_{i,1}}_{\substack{=x_m(\operatorname{col}_m A)_{i,1} \\ \text{(since matrices are} \\ \text{scaled entry by entry)}}} \\
&\qquad \left(\begin{array}{c} \text{since matrices are added entry by entry} \\ \text{(more precisely, we are using Proposition 2.45 here)} \end{array} \right) \\
&= x_1 \underbrace{(\operatorname{col}_1 A)_{i,1}}_{\substack{=A_{i,1} \\ \text{(by (163),} \\ \text{applied to } j=1)}} + x_2 \underbrace{(\operatorname{col}_2 A)_{i,1}}_{\substack{=A_{i,2} \\ \text{(by (163),} \\ \text{applied to } j=2)}} + \cdots + x_m \underbrace{(\operatorname{col}_m A)_{i,1}}_{\substack{=A_{i,m} \\ \text{(by (163),} \\ \text{applied to } j=m)}} \\
&= x_1 A_{i,1} + x_2 A_{i,2} + \cdots + x_m A_{i,m}.
\end{aligned} \tag{164}$$

Set $B = (x_1, x_2, \ldots, x_m)^T$. Then, B is an $m \times 1$-matrix, and its entries are

$$B_{k,1} = x_k \qquad \text{for each } k \in \{1, 2, \ldots, m\}. \tag{165}$$

Now, AB is an $n \times 1$-matrix (since A is an $n \times m$-matrix and B is an $m \times 1$-matrix), i.e., a column vector of size n. For each $i \in \{1, 2, \ldots, n\}$, we have

$$\begin{aligned}
(AB)_{i,1} &= A_{i,1} \underbrace{B_{1,1}}_{\substack{=x_1 \\ \text{(by (165),} \\ \text{applied to } k=1)}} + A_{i,2} \underbrace{B_{2,1}}_{\substack{=x_2 \\ \text{(by (165),} \\ \text{applied to } k=2)}} + \cdots + A_{i,m} \underbrace{B_{m,1}}_{\substack{=x_m \\ \text{(by (165),} \\ \text{applied to } k=m)}} \\
&\qquad \text{(by Proposition 2.19 \textbf{(a)}, applied to } p = 1 \text{ and } j = 1) \\
&= A_{i,1} x_1 + A_{i,2} x_2 + \cdots + A_{i,m} x_m \\
&= x_1 A_{i,1} + x_2 A_{i,2} + \cdots + x_m A_{i,m} \\
&= (x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A)_{i,1}
\end{aligned}$$

(by (164)). In other words, for each $i \in \{1, 2, \ldots, n\}$, the i-th entry of the column vector AB equals the i-th entry of the column vector $x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots +$

$x_m \operatorname{col}_m A$ (because the i-th entry of any column vector v is $v_{i,1}$). Thus, the column vector AB equals $x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A$. In other words,

$$AB = x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A.$$

Since $B = (x_1, x_2, \ldots, x_m)^T$, this rewrites as

$$A (x_1, x_2, \ldots, x_m)^T = x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A.$$

Lemma 4.39 is proven. □

Proof of Proposition 4.38. **(a)** Applying (158) to $V = \mathbb{R}^n$, $k = m$ and $v_i = \operatorname{col}_i A$, we obtain

$$\begin{aligned} &\operatorname{span}(\operatorname{col}_1 A, \operatorname{col}_2 A, \ldots, \operatorname{col}_m A) \\ &= \{\lambda_1 \operatorname{col}_1 A + \lambda_2 \operatorname{col}_2 A + \cdots + \lambda_m \operatorname{col}_m A \mid \lambda_1, \lambda_2, \ldots, \lambda_m \in \mathbb{R}\} \\ &= \{x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A \mid x_1, x_2, \ldots, x_m \in \mathbb{R}\} \end{aligned} \tag{166}$$

(here, we have renamed the summation indices $\lambda_1, \lambda_2, \ldots, \lambda_m$ as $x_1, x_2, \ldots, x_m$). But

$$\begin{aligned} A\mathbb{R}^m &= \{Ax \mid x \in \mathbb{R}^m\} \qquad \text{(by the definition of } A\mathbb{R}^m\text{)} \\ &= \left\{ \underbrace{A (x_1, x_2, \ldots, x_m)^T}_{\substack{= x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A \\ \text{(by Lemma 4.39)}}} \mid x_1, x_2, \ldots, x_m \in \mathbb{R} \right\} \\ &\qquad \left(\begin{array}{c} \text{here, we have substituted } (x_1, x_2, \ldots, x_m)^T \text{ for } x, \\ \text{since the elements } x \text{ of } \mathbb{R}^m \text{ are exactly the column vectors} \\ \text{of the form } (x_1, x_2, \ldots, x_m)^T \text{ with } x_1, x_2, \ldots, x_m \in \mathbb{R} \end{array} \right) \\ &= \{x_1 \operatorname{col}_1 A + x_2 \operatorname{col}_2 A + \cdots + x_m \operatorname{col}_m A \mid x_1, x_2, \ldots, x_m \in \mathbb{R}\}. \end{aligned}$$

Comparing this with (166), we obtain $A\mathbb{R}^m = \operatorname{span}(\operatorname{col}_1 A, \operatorname{col}_2 A, \ldots, \operatorname{col}_m A)$. This proves Proposition 4.38 **(a)**.

(b) Proposition 4.31 (applied to $V = \mathbb{R}^n$, $k = m$ and $v_i = \operatorname{col}_i A$) shows that $\operatorname{span}(\operatorname{col}_1 A, \operatorname{col}_2 A, \ldots, \operatorname{col}_m A)$ is a subspace of $\mathbb{R}^n$. In other words, $A\mathbb{R}^m$ is a subspace of $\mathbb{R}^n$ (since Proposition 4.38 **(a)** yields $A\mathbb{R}^m = \operatorname{span}(\operatorname{col}_1 A, \operatorname{col}_2 A, \ldots, \operatorname{col}_m A)$). This proves Proposition 4.38 **(b)**. □

The probably simplest example of an image is the following:

Proposition 4.40. Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Then, $0_{n \times m} \mathbb{R}^m = \{0_{n \times 1}\}$.

Proof of Proposition 4.40. The definition of $0_{n \times m} \mathbb{R}^m$ yields

$$0_{n \times m} \mathbb{R}^m = \left\{ \underbrace{0_{n \times m} x}_{= 0_{n \times 1}} \mid x \in \mathbb{R}^m \right\} = \{0_{n \times 1} \mid x \in \mathbb{R}^m\} \subseteq \{0_{n \times 1}\}. \tag{167}$$

On the other hand, $0_{n\times 1} = 0_{n\times m} 0_{m\times 1} \in 0_{n\times m}\mathbb{R}^m$ (since $0_{m\times 1} \in \mathbb{R}^m$), and thus $\{0_{n\times 1}\} \subseteq 0_{n\times m}\mathbb{R}^m$. Combining this with (167), we obtain $0_{n\times m}\mathbb{R}^m = \{0_{n\times 1}\}$. This proves Proposition 4.40. □

4.5.6. Subspaces from subspaces

Next, let us see two ways to construct subspaces from other subspaces:

Proposition 4.41. Let V be a vector space. Let U_1 and U_2 be two subspaces of V. Then, $U_1 \cap U_2$ is a subspace of V.

Proof of Proposition 4.41. Clearly, $U_1 \cap U_2$ is a subset of V (since $U_1 \cap U_2 \subseteq U_1 \subseteq V$). We want to show that $U_1 \cap U_2$ is a subspace of V. Thus, we need to show that $U_1 \cap U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that $U_1 \cap U_2$ contains the zero vector: The set U_1 is a subspace of V, and thus contains the zero vector. In other words, $\overrightarrow{0} \in U_1$. Similarly, $\overrightarrow{0} \in U_2$. Combining $\overrightarrow{0} \in U_1$ with $\overrightarrow{0} \in U_2$, we obtain $\overrightarrow{0} \in U_1 \cap U_2$. In other words, $U_1 \cap U_2$ contains the zero vector (since $\overrightarrow{0}$ is the zero vector of the vector space V).

Proof that $U_1 \cap U_2$ is closed under addition: Let $v \in U_1 \cap U_2$ and $w \in U_1 \cap U_2$. We must prove that $v + w \in U_1 \cap U_2$.

We have $v \in U_1 \cap U_2 \subseteq U_1$ and $w \in U_1 \cap U_2 \subseteq U_1$. But U_1 is a subspace of V, and thus is closed under addition. Hence, from $v \in U_1$ and $w \in U_1$, we can conclude that $v + w \in U_1$. The same argument (but with the roles of U_1 and U_2 interchanged) shows that $v + w \in U_2$.

We now know that the element $v + w$ lies in both sets U_1 and U_2; therefore, $v + w$ lies in their intersection $U_1 \cap U_2$. In other words, $v + w \in U_1 \cap U_2$. This completes the proof of the fact that $U_1 \cap U_2$ is closed under addition.

Proof that $U_1 \cap U_2$ is closed under scaling: This proof is left to the reader. (It is similar to the above proof that $U_1 \cap U_2$ is closed under addition.)

We now have shown that $U_1 \cap U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Hence, $U_1 \cap U_2$ is a subspace of V. This proves Proposition 4.41. □

Definition 4.42. Let V be a vector space. Let U_1 and U_2 be two subspaces of V. Then, $U_1 + U_2$ denotes the set

$$\{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\}.$$

This set $U_1 + U_2$ is called the *sum* of the subspaces U_1 and U_2.

Proposition 4.43. Let V be a vector space. Let U_1 and U_2 be two subspaces of V. Then, $U_1 + U_2$ is a subspace of V.

Proof of Proposition 4.43. We have

$$U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\} \tag{168}$$

(by the definition of $U_1 + U_2$). Hence, the elements of $U_1 + U_2$ are vectors of the form $u_1 + u_2$ with $u_1 \in U_1$ and $u_2 \in U_2$; clearly, these vectors all belong to V. Hence, $U_1 + U_2$ is a subset of V. We want to show that $U_1 + U_2$ is a subspace of V. Thus, we need to show that $U_1 + U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that $U_1 + U_2$ contains the zero vector: The set U_1 is a subspace of V, and thus contains the zero vector. In other words, $\overrightarrow{0} \in U_1$. Similarly, $\overrightarrow{0} \in U_2$. Now, $\overrightarrow{0} = \overrightarrow{0} + \overrightarrow{0}$. Hence, $\overrightarrow{0}$ is a vector of the form $u_1 + u_2$ with $u_1 \in U_1$ and $u_2 \in U_2$ (namely, $u_1 = \overrightarrow{0}$ and $u_2 = \overrightarrow{0}$). In other words,

$$\overrightarrow{0} \in \{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\}.$$

In light of (168), this rewrites as $\overrightarrow{0} \in U_1 + U_2$. In other words, $U_1 + U_2$ contains the zero vector (since $\overrightarrow{0}$ is the zero vector of the vector space V).

Proof that $U_1 + U_2$ is closed under addition: Let $v \in U_1 + U_2$ and $w \in U_1 + U_2$. We must prove that $v + w \in U_1 + U_2$.

We have $v \in U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\}$. In other words, v has the form $v = u_1 + u_2$ for some $u_1 \in U_1$ and $u_2 \in U_2$. Denote these u_1 and u_2 by v_1 and v_2. Thus, $v_1 \in U_1$ and $v_2 \in U_2$ and $v = v_1 + v_2$.

We have $w \in U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\}$. In other words, w has the form $w = u_1 + u_2$ for some $u_1 \in U_1$ and $u_2 \in U_2$. Denote these u_1 and u_2 by w_1 and w_2. Thus, $w_1 \in U_1$ and $w_2 \in U_2$ and $w = w_1 + w_2$.

Now, the set U_1 is a subspace of V, and thus is closed under addition. Hence, from $v_1 \in U_1$ and $w_1 \in U_1$, we can conclude that $v_1 + w_1 \in U_1$. Similarly, $v_2 + w_2 \in U_2$.

Adding the equalities $v = v_1 + v_2$ and $w = w_1 + w_2$, we obtain

$$v + w = (v_1 + v_2) + (w_1 + w_2) = (v_1 + w_1) + (v_2 + w_2).$$

Hence, $v + w$ has the form $v + w = u_1 + u_2$ for some $u_1 \in U_1$ and $u_2 \in U_2$ (namely, for $u_1 = v_1 + w_1$ and $u_2 = v_2 + w_2$) [108]. In other words,

$$v + w \in \{u_1 + u_2 \mid u_1 \in U_1 \text{ and } u_2 \in U_2\}.$$

In light of (168), this rewrites as $v + w \in U_1 + U_2$. This completes the proof of the fact that $U_1 + U_2$ is closed under addition.

Proof that $U_1 + U_2$ is closed under scaling: We leave this proof to the reader. (It is similar to the above proof that $U_1 + U_2$ is closed under addition.)

We now have shown that $U_1 + U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Hence, $U_1 + U_2$ is a subspace of V. This proves Proposition 4.43. □

[108] Here, we are using the observations (made above) that $v_1 + w_1 \in U_1$ and $v_2 + w_2 \in U_2$.

Generalizing Definition 4.42, we can define the sum $U_1 + U_2 + \cdots + U_n$ of any (finite) number of subspaces $U_1, U_2, \ldots, U_n$ of V. This, too, will be a subspace of V. The proof of this is analogous to our above proof of Proposition 4.43.

Proposition 4.44. Let V_1 and V_2 be two vector spaces. Let U_1 be a subspace of V_1. Let U_2 be a subspace of V_2. Then, $U_1 \times U_2$ is a subspace of $V_1 \times V_2$. (See Definition 4.15 for the definition of the vector space $V_1 \times V_2$.)

Proof of Proposition 4.44. Every element of $U_1 \times U_2$ is a pair $(u_1, u_2) \in U_1 \times U_2$. All such pairs (u_1, u_2) belong to $V_1 \times V_2$ [109]. Hence, $U_1 \times U_2$ is a subset of $V_1 \times V_2$. We want to show that $U_1 \times U_2$ is a subspace of $V_1 \times V_2$. Thus, we need to show that $U_1 \times U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that $U_1 \times U_2$ contains the zero vector: In the following, we will denote the zero vector of a vector space W by $\overrightarrow{0}_W$. (We are using this notation instead of the usual notation $\overrightarrow{0}$, because we want to be able to distinguish between the zero vectors of different vector spaces.)

The set U_1 is a subspace of V_1, and thus contains the zero vector. In other words, $\overrightarrow{0}_{V_1} \in U_1$. Similarly, $\overrightarrow{0}_{V_2} \in U_2$. Now, the definition of $\overrightarrow{0}_{V_1 \times V_2}$ yields $\overrightarrow{0}_{V_1 \times V_2} = \left(\overrightarrow{0}_{V_1}, \overrightarrow{0}_{V_2}\right) \in U_1 \times U_2$ (since $\overrightarrow{0}_{V_1} \in U_1$ and $\overrightarrow{0}_{V_2} \in U_2$). In other words, $U_1 \times U_2$ contains the zero vector (of the vector space $V_1 \times V_2$).

Proof that $U_1 \times U_2$ is closed under addition: Let $v \in U_1 \times U_2$ and $w \in U_1 \times U_2$. We must prove that $v + w \in U_1 \times U_2$.

We have $v \in U_1 \times U_2$. In other words, v is a pair of the form $v = (u_1, u_2)$ for some $u_1 \in U_1$ and $u_2 \in U_2$. Denote these u_1 and u_2 by v_1 and v_2. Thus, $v_1 \in U_1$ and $v_2 \in U_2$ and $v = (v_1, v_2)$.

We have $w \in U_1 \times U_2$. In other words, w is a pair of the form $w = (u_1, u_2)$ for some $u_1 \in U_1$ and $u_2 \in U_2$. Denote these u_1 and u_2 by w_1 and w_2. Thus, $w_1 \in U_1$ and $w_2 \in U_2$ and $w = (w_1, w_2)$.

Now, the set U_1 is a subspace of V_1, and thus is closed under addition. Hence, from $v_1 \in U_1$ and $w_1 \in U_1$, we can conclude that $v_1 + w_1 \in U_1$. Similarly, $v_2 + w_2 \in U_2$.

Adding the equalities $v = (v_1, v_2)$ and $w = (w_1, w_2)$, we obtain

$$v + w = (v_1, v_2) + (w_1, w_2) = (v_1 + w_1, v_2 + w_2).$$

Hence, $v + w$ has the form $v + w = (u_1, u_2)$ for some $u_1 \in U_1$ and $u_2 \in U_2$ (namely, for $u_1 = v_1 + w_1$ and $u_2 = v_2 + w_2$) [110]. In other words, $v + w \in U_1 \times U_2$. This completes the proof of the fact that $U_1 \times U_2$ is closed under addition.

Proof that $U_1 \times U_2$ is closed under scaling: We leave this proof to the reader. (It is similar to the above proof that $U_1 \times U_2$ is closed under addition.)

[109] *Proof.* Let $(u_1, u_2) \in U_1 \times U_2$. Then, $u_1 \in U_1$ and $u_2 \in U_2$. Hence, $u_1 \in U_1 \subseteq V_1$ and $u_2 \in U_2 \subseteq V_2$. Thus, $(u_1, u_2) \in V_1 \times V_2$, qed.

[110] Here, we are using the observations (made above) that $v_1 + w_1 \in U_1$ and $v_2 + w_2 \in U_2$.

We now have shown that $U_1 \times U_2$ contains the zero vector, is closed under addition, and is closed under scaling. Hence, $U_1 \times U_2$ is a subspace of $V_1 \times V_2$. This proves Proposition 4.44. □

4.5.7. Matrix spaces

As we have seen in Example 4.6, the $n \times m$-matrices (with real entries, for given n and m) form a vector space, called $\mathbb{R}^{n \times m}$. How do subspaces of this space look like? Let us see some examples of subspaces of $\mathbb{R}^{2\times 2}$:

Example 4.45. (a) Let A be the subset

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} \mid a+b+c+d = 0 \right\}$$

of $\mathbb{R}^{2\times 2}$. Then, A is a subspace of $\mathbb{R}^{2\times 2}$.

(b) Let B be the subset

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} \mid ab+cd = 0 \right\}$$

of $\mathbb{R}^{2\times 2}$. Then, B is **not** a subspace of $\mathbb{R}^{2\times 2}$.

(c) Let C be the set of all upper-triangular 2×2-matrices. Then, C is a subspace of $\mathbb{R}^{2\times 2}$.

(d) Let D be the set of all strictly upper-triangular 2×2-matrices. Then, D is a subspace of $\mathbb{R}^{2\times 2}$.

(e) Let E be the set of all upper-unitriangular 2×2-matrices. Then, E is **not** a subspace of $\mathbb{R}^{2\times 2}$.

(f) Let F be the set of all invertible 2×2-matrices. Then, F is **not** a subspace of $\mathbb{R}^{2\times 2}$.

(g) Fix a vector $p \in \mathbb{R}^2$. Let G_p be the set of all 2×2-matrices $X \in \mathbb{R}^{2\times 2}$ satisfying $Xp = 0_{2\times 1}$. Then, G_p is a subspace of $\mathbb{R}^{2\times 2}$.

Proof of Example 4.45. This proof is similar to that of Example 4.25, except that the vectors are now 2×2-matrices instead of being column vectors of size 3. Again, in order to prove that some subset U of $\mathbb{R}^{2\times 2}$ is a subspace of $\mathbb{R}^{2\times 2}$, we must prove that it satisfies the three conditions **(a)**, **(b)** and **(c)** of Definition 4.21; but in order to prove that some subset U of $\mathbb{R}^{2\times 2}$ is **not** a subspace of $\mathbb{R}^{2\times 2}$, it suffices to show that at least one of these three conditions is violated.

Keep in mind that the zero vector of the vector space $\mathbb{R}^{2\times 2}$ is the zero matrix $0_{2\times 2}$. We shall sometimes keep using the notation $\overrightarrow{0}$ for it, just because it is shorter.

(a) We want to show that A is a subspace of $\mathbb{R}^{2\times 2}$. Thus, we need to show that A contains the zero vector, is closed under addition, and is closed under scaling. Let us do this:

Proof that A contains the zero vector: Recall that the zero vector of $\mathbb{R}^{2\times 2}$ is $0_{2\times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. This vector $0_{2\times 2}$ lies in A if and only if $0+0+0+0 = 0$ (by the definition of A). Thus, $0_{2\times 2}$ lies in A (since $0+0+0+0=0$ holds). In other words, A contains the zero vector.

Proof that A is closed under addition: Let $v \in A$ and $w \in A$. We must prove that $v+w \in A$.

Write the vector[111] $v \in A \subseteq \mathbb{R}^{2\times 2}$ in the form $v = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix}$ (for four real numbers $v_{1,1}, v_{1,2}, v_{2,1}, v_{2,2} \in \mathbb{R}$). Since $\begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix} = v \in A$, we have $v_{1,1}+v_{1,2}+v_{2,1}+v_{2,2}=0$.

Write the vector $w \in A \subseteq \mathbb{R}^{2\times 2}$ in the form $w = \begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{pmatrix}$ (for four real numbers $w_{1,1}, w_{1,2}, w_{2,1}, w_{2,2} \in \mathbb{R}$). Since $\begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{pmatrix} = w \in A$, we have $w_{1,1}+w_{1,2}+w_{2,1}+w_{2,2}=0$.

From $v = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix}$ and $w = \begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{pmatrix}$, we obtain

$$v+w = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix} + \begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \end{pmatrix} = \begin{pmatrix} v_{1,1}+w_{1,1} & v_{1,2}+w_{1,2} \\ v_{2,1}+w_{2,1} & v_{2,2}+w_{2,2} \end{pmatrix}$$

(since matrices are added entry by entry). Thus, in order to prove that $v+w \in A$, we must show that

$$(v_{1,1}+w_{1,1}) + (v_{1,2}+w_{1,2}) + (v_{2,1}+w_{2,1}) + (v_{2,2}+w_{2,2}) = 0$$

(by the definition of A). But showing this is easy: Just notice that

$$\begin{aligned} &(v_{1,1}+w_{1,1}) + (v_{1,2}+w_{1,2}) + (v_{2,1}+w_{2,1}) + (v_{2,2}+w_{2,2}) \\ &= \underbrace{(v_{1,1}+v_{1,2}+v_{2,1}+v_{2,2})}_{=0} + \underbrace{(w_{1,1}+w_{1,2}+w_{2,1}+w_{2,2})}_{=0} = 0+0=0. \end{aligned}$$

Thus, we have proven that $v+w \in A$. This completes the proof of the fact that A is closed under addition.

Proof that A is closed under scaling: We leave this proof to the reader again. (It is similar to the proof of the fact that A is closed under addition, except that now we must reason about λv instead of $v+w$.)

We now have shown that A contains the zero vector, is closed under addition, and is closed under scaling. Hence, A is a subspace of $\mathbb{R}^{2\times 2}$. Example 4.45 **(a)** is proven.

[111] It might appear to refer to the elements of A as vectors (after all, they are 2×2-matrices), and to denote them by lowercase letters. This should remind you once again that vectors (in the general meaning of this word – i.e., elements of a vector space) can be almost anything.

(b) The vector $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ belongs to B (since $1 \cdot 0 + 0 \cdot 1 = 0$). So does the vector $\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ (since $1 \cdot 0 + 1 \cdot 0 = 0$). But the sum of these vectors does not (because it is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$, which does not satisfy $2 \cdot 0 + 1 \cdot 1 = 0$). Hence, B is not closed under addition. Thus, B is not a subspace of $\mathbb{R}^{2\times 2}$. Example 4.45 **(b)** is proven.

(c) We want to show that C is a subspace of $\mathbb{R}^{2\times 2}$. Thus, we need to show that C contains the zero vector, is closed under addition, and is closed under scaling. Let me show only one of these three steps:

Proof that C is closed under addition: Let $v \in C$ and $w \in C$. We must prove that $v + w \in C$.

The matrix v belongs to C. In other words, v is upper-triangular (since C is the set of all upper-triangular 2×2-matrices). In other words,

$$v_{i,j} = 0 \qquad \text{whenever } i > j \tag{169}$$

(because this is how "upper-triangular" was defined). The same argument (but for w instead of v) yields

$$w_{i,j} = 0 \qquad \text{whenever } i > j. \tag{170}$$

Now, every $(i,j) \in \{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$ satisfying $i > j$ must satisfy

$$\begin{aligned} (v+w)_{i,j} &= \underbrace{v_{i,j}}_{\substack{=0 \\ \text{(by (169))}}} + \underbrace{w_{i,j}}_{\substack{=0 \\ \text{(by (170))}}} \qquad \text{(since matrices are added entry by entry)} \\ &= 0 + 0 = 0. \end{aligned}$$

In other words,

$$(v+w)_{i,j} = 0 \qquad \text{whenever } i > j.$$

In other words, the matrix $v + w$ is upper-triangular (since this is how "upper-triangular" was defined). In other words, $v + w \in C$ (since C is the set of all upper-triangular 2×2-matrices). This completes the proof of the fact that C is closed under addition.

As I have promised, I am omitting the proofs of the facts that C contains the zero vector and is closed under scaling. (They are similar to the above proof, and the modifications necessary to obtain them should be obvious by now.) Thus, Example 4.45 **(c)** is proven.

(d) The proof of Example 4.45 **(d)** is analogous to the above proof of Example 4.45 **(c)**, except that the word "upper-triangular" must now be replaced by "strictly upper-triangular", and that the condition "$i > j$" must be replaced by "$i \geq j$". (And, of course, the symbol "C" must be replaced by "D".)

(e) The vector $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ belongs to E (since it is an upper-unitriangular 2×2-matrix). But the sum of this vector with itself does not (since it is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, which is not upper-unitriangular). Thus, E is not closed under addition. Thus, E is not a subspace of $\mathbb{R}^{2\times 2}$. Example 4.45 **(e)** is proven.

(f) The vectors $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ both belong to F (since they both are invertible matrices[112]). But their sum does not (because it is $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, which is not invertible). Hence, F is not closed under addition. Thus, F is not a subspace of $\mathbb{R}^{2\times 2}$. Example 4.45 **(f)** is proven.

(g) We want to show that G_p is a subspace of $\mathbb{R}^{2\times 2}$. Thus, we need to show that G_p contains the zero vector, is closed under addition, and is closed under scaling.

Proof that G_p contains the zero vector: Recall that the zero vector of $\mathbb{R}^{2\times 2}$ is $0_{2\times 2}$. This vector $0_{2\times 2}$ satisfies $0_{2\times 2}p = 0_{2\times 1}$, and thus belongs to G_p (by the definition of G_p). In other words, G_p contains the zero vector.

Proof that G_p is closed under addition: Let $v \in G_p$ and $w \in G_p$. We must prove that $v + w \in G_p$.

We have $v \in G_p$. In other words, $vp = 0_{2\times 1}$ (by the definition of G_p). (Keep in mind that v is a 2×2-matrix.)

We have $w \in G_p$. In other words, $wp = 0_{2\times 1}$ (by the definition of G_p).

Now, $(v+w)\,p = \underbrace{vp}_{=0_{2\times 1}} + \underbrace{wp}_{=0_{2\times 1}} = 0_{2\times 1} + 0_{2\times 1} = 0_{2\times 1}$. In other words, $v + w \in G_p$ (by the definition of G_p). This completes the proof of the fact that G_p is closed under addition.

Proof that G_p is closed under scaling: This proof is left to the reader.

We now have shown that G_p contains the zero vector, is closed under addition, and is closed under scaling. Hence, G_p is a subspace of $\mathbb{R}^{2\times 2}$. Example 4.45 **(g)** is proven. □

[...]

TODO 4.46. Polynomials p such that $p(3) = 0$.

TODO 4.47. Functions f such that f is continuous.

TODO 4.48. Are functions $[-1, 1] \to \mathbb{R}$ are a subspace of the functions $\mathbb{R} \to \mathbb{R}$? No, since they are not even a subset.

[112] Actually, each of them is its own inverse.

TODO 4.49. Are functions $\mathbb{R} \to \mathbb{R}$ with image in $[-1,1]$ a subspace of the functions $\mathbb{R} \to \mathbb{R}$? No (e.g., scaling fails).

TODO 4.50. Subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$ in geometric terms.

[...]

4.6. <DRAFT> More on subspaces

If U is a subspace of a vector space V, then the sum of any two elements of U must belong to U (by the definition of a subspace). The same holds for sums of any (finite) number of elements of U:

Proposition 4.51. Let V be a vector space. Let U be a subspace of V.
(a) If $u_1, u_2, \ldots, u_k$ are elements of U, then $u_1 + u_2 + \cdots + u_k \in U$.
(b) If $u_1, u_2, \ldots, u_k$ are elements of U, then every linear combination of $u_1, u_2, \ldots, u_k$ also lies in U.
(c) Let $u_1, u_2, \ldots, u_k$ be elements of U. Then,

$$\operatorname{span}(u_1, u_2, \ldots, u_k) \subseteq U.$$

The claim of Proposition 4.51 **(b)** is often expressed as the following short slogan: "A subspace U of a vector space V is closed under linear combination".

Proof of Proposition 4.51. **(a)** Roughly speaking, this is just a matter of applying the "closed under addition" axiom several times. But there is a subtlety involved (the sum of 0 elements of U is not obtained by addition, but rather defined as $\overrightarrow{0}$), and I want to illustrate the principle of mathematical induction once again, so I am going to present the proof in full detail.

Let $u_1, u_2, \ldots, u_k$ be elements of U. We must show that $u_1 + u_2 + \cdots + u_k \in U$.

We shall prove that

$$u_1 + u_2 + \cdots + u_i \in U \qquad \text{for every } i \in \{0, 1, \ldots, k\}. \tag{171}$$

We will prove (171) by induction over i. (See the proof of Proposition 2.45 given above for a brief explanation of what this principle means).

1. *Induction base:* For $i = 0$, the statement (171) claims that $u_1 + u_2 + \cdots + u_0 \in U$. In order to make sense of this, we must recall that empty sums of vectors are defined to mean $\overrightarrow{0}$. Thus,

$$u_1 + u_2 + \cdots + u_0 = (\text{empty sum of vectors}) = \overrightarrow{0}.$$

But U is a subspace of V, and thus contains $\overrightarrow{0}$ (this is one of the axioms for a subspace). Thus, $\overrightarrow{0} \in U$, so that $u_1 + u_2 + \cdots + u_0 = \overrightarrow{0} \in U$. In other words, (171) holds for $i = 0$. This completes the induction base.

2. *Induction step:* Let $j \in \{0, 1, \ldots, k-1\}$ be such that (171) holds for $i = j$. (The statement that (171) holds for $i = j$ is called the "induction hypothesis".) We must show that (171) also holds for $i = j + 1$.

   Since (171) holds for $i = j$, we have $u_1 + u_2 + \cdots + u_j \in U$. Now,

   $$u_1 + u_2 + \cdots + u_{j+1} = \underbrace{\left(u_1 + u_2 + \cdots + u_j\right)}_{\in U} + \underbrace{u_{j+1}}_{\in U}.$$

   This is a sum of two vectors in U, and thus belongs to U (since U is closed under addition). In other words, $u_1 + u_2 + \cdots + u_{j+1} \in U$. Thus, (171) also holds for $i = j + 1$. This completes the induction step.

Now, the proof of (171) is complete (since both the induction base and the induction step are complete).

Now that (171) is proven, we can simply apply (171) to $i = k$, and conclude that $u_1 + u_2 + \cdots + u_k \in U$. This proves Proposition 4.51 **(a)**.

(b) Let $u_1, u_2, \ldots, u_k$ be elements of U. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be real numbers.

The set U is a subspace of V, and thus is closed under scaling. Hence, the vectors $\lambda_1 u_1, \lambda_2 u_2, \ldots, \lambda_k u_k$ all belong to U (since the vectors $u_1, u_2, \ldots, u_k$ belong to U). Thus, Proposition 4.51 **(a)** (applied to $\lambda_1 u_1, \lambda_2 u_2, \ldots, \lambda_k u_k$ instead of $u_1, u_2, \ldots, u_k$) shows that $\lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_k u_k \in U$.

Now, we have shown that $\lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_k u_k \in U$ whenever $\lambda_1, \lambda_2, \ldots, \lambda_k$ are real numbers. In other words: every linear combination of $u_1, u_2, \ldots, u_k$ also lies in U. This proves Proposition 4.51 **(b)**.

(c) Proposition 4.51 **(b)** shows that every linear combination of $u_1, u_2, \ldots, u_k$ lies in U. In other words, the set of all linear combinations of $u_1, u_2, \ldots, u_k$ is a subset of U. But since $\operatorname{span}(u_1, u_2, \ldots, u_k)$ is precisely the set of all linear combinations of $u_1, u_2, \ldots, u_k$ (in fact, this is how $\operatorname{span}(u_1, u_2, \ldots, u_k)$ was defined), this rewrites as follows: $\operatorname{span}(u_1, u_2, \ldots, u_k)$ is a subset of U. This proves Proposition 4.51 **(c)**. □

I owe you a proof of Proposition 4.33; it can now be done quite easily:

Proof of Proposition 4.33. **(a)** Let $i \in \{1, 2, \ldots, k\}$. We claim that $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_k)$.

[*Proof:* We have

$$\begin{aligned} &\underbrace{0v_1}_{=\overrightarrow{0}} + \underbrace{0v_2}_{=\overrightarrow{0}} + \cdots + \underbrace{0v_{i-1}}_{=\overrightarrow{0}} + \underbrace{1v_i}_{=v_i} + \underbrace{0v_{i+1}}_{=\overrightarrow{0}} + \underbrace{0v_{i+2}}_{=\overrightarrow{0}} + \cdots + \underbrace{0v_k}_{=\overrightarrow{0}} \\ &= \overrightarrow{0} + \overrightarrow{0} + \cdots + \overrightarrow{0} + v_i + \overrightarrow{0} + \overrightarrow{0} + \cdots + \overrightarrow{0} = v_i, \end{aligned}$$

so that $v_i = 0v_1 + 0v_2 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + 0v_{i+2} + \cdots + 0v_k$. Hence, the vector v_i has the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ for some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ (namely, for $\lambda_j = \begin{cases} 1, \text{ if } j = i; \\ 0, \text{ if } j \neq i \end{cases}$). In other words,

$$v_i \in \left\{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\right\}.$$

In view of (158), this rewrites as $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_k)$. [113]]

Now, forget that we fixed i. We thus have shown that $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_k)$ for each $i \in \{1, 2, \ldots, k\}$. In other words, the vectors $v_1, v_2, \ldots, v_k$ all belong to $\operatorname{span}(v_1, v_2, \ldots, v_k)$. This proves Proposition 4.33 **(a)**.

(b) Let U be a subspace of V that contains the vectors $v_1, v_2, \ldots, v_k$. We must prove that $\operatorname{span}(v_1, v_2, \ldots, v_k) \subseteq U$. But this follows immediately from Proposition 4.51 **(c)** (applied to $u_i = v_i$). Thus, Proposition 4.33 **(b)** is proven. □

[...]
[...]

4.7. <TODO> More on spans

References for vector spaces: [LaNaSc16, §5.1] (possibly the best one), [OlvSha06, §2.3] (focusses on the analysis-related and applied stuff) and [Heffer16, Two.I.2].

TODO 4.52. The terminology "$(v_1, v_2, \ldots, v_k)$ spans V" (or "$v_1, v_2, \ldots, v_k$ span V") means $V = \operatorname{span}(v_1, v_2, \ldots, v_k)$.

TODO 4.53. $E_{i,j}$ span $\mathbb{R}^{n \times m}$. (Use Proposition 3.50.)

TODO 4.54. $e_i = E_{i,1}$ span $\mathbb{R}^n$.

TODO 4.55. How to check that a given vector is in a given span:
$w \in \operatorname{span}(v_1, v_2, \ldots, v_k)$
$\iff$ the equation $w = \lambda_1 v_1 + \cdots + \lambda_k v_k$ has at least one solution $(\lambda_1, \lambda_2, \ldots, \lambda_k)^T \in \mathbb{R}^k$.
Can solve this using Gaussian elimination.

[113] Here is a more rigorous version of the same proof:
Proposition 4.20 (applied to $p = 1$, $q = k$ and $r = i$) yields

$$\sum_{g=1}^{k} \delta_{g,i} v_g = v_i.$$

(Here, we are using the letter g instead of k for the summation index, in order to avoid a clash with the already existing meaning of the letter k.) Thus,

$$v_i = \sum_{g=1}^{k} \delta_{g,i} v_g = \delta_{1,i} v_1 + \delta_{2,i} v_2 + \cdots + \delta_{k,i} v_k.$$

Hence, the vector v_i has the form $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$ for some $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$ (namely, for $\lambda_j = \delta_{j,i}$). In other words,

$$v_i \in \{\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}\}.$$

In view of (158), this rewrites as $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_k)$.

TODO 4.56. How to check that a given span is contained in another given span:

Checking that $\operatorname{span}(\alpha,\beta) \subseteq \operatorname{span}(\gamma,\delta,\varepsilon)$ is tantamount to checking that both α and β lie in $\operatorname{span}(\gamma,\delta,\varepsilon)$ (by Proposition 4.51 **(c)**).

TODO 4.57. How to check that two given spans are identical?

Check that each is contained in the other.

TODO 4.58. How to write the subspace carved out by equations (i.e., a kernel) as a span?

Solve the system.

For instance, write $\operatorname{Ker}\begin{pmatrix} 1 & 3 & 2 \\ 3 & 9 & 6 \\ 2 & 6 & 4 \end{pmatrix}$ as a span.

Solve the system $\begin{pmatrix} 1 & 3 & 2 \\ 3 & 9 & 6 \\ 2 & 6 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Solution: $x = \begin{pmatrix} -3s-2r \\ s \\ r \end{pmatrix}$. Rewrite this as $x = r\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} + s\begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$ (here I have just moved the free variables r, s outside of the vector, so that only constant numbers have remained in the vector).

Thus, the elements of $\operatorname{Ker}\begin{pmatrix} 1 & 3 & 2 \\ 3 & 9 & 6 \\ 2 & 6 & 4 \end{pmatrix}$ are precisely the vectors of the form $x = r\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} + s\begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$. In other words, they are the linear combinations of $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}$. In other words, $\operatorname{Ker}\begin{pmatrix} 1 & 3 & 2 \\ 3 & 9 & 6 \\ 2 & 6 & 4 \end{pmatrix} = \operatorname{span}\left(\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}\right)$.

TODO 4.59. Example: $\left((1,3,2)^T, (0,1,1)^T\right)$ spans $W := \left\{(x,y,z)^T \mid x+z=y\right\} \subseteq \mathbb{R}^3$.

Prove this in full: Let $\alpha = (1,3,2)^T$ and $\beta = (0,1,1)^T$.

Show $\alpha, \beta \in W$. Conclude that $\operatorname{span}(\alpha,\beta) \subseteq W$ by Proposition 4.51 **(c)**.

Now prove converse: Solve the system $\{x+z=y$ and write the solution as $(y-z,y,z)^T = y(1,1,0)^T + z(-1,0,1)^T$. So $W = \operatorname{span}\left((1,1,0)^T, (-1,0,1)^T\right)$. Now, need to check that two spans are equal. We know how this is done.

TODO 4.60. Example: $\left((1,-1,1)^T\right)$ spans $X := \left\{(x,y,z)^T \mid x+3y+z=0 \text{ and } y+z=0\right\}$.
Prove this too.

TODO 4.61. Exercise: what spans $\left\{(u-v,v-u)^T\right\} \subseteq \mathbb{R}^2$? Is there an easier description?

TODO 4.62. Exercise: what spans $\left\{(u,v,0)^T\right\}$?
$\left\{(u,0,0)^T\right\}$?
$\left\{(u,v,u)^T\right\}$?
$\left\{(x_1,x_2,x_3)^T \mid x_1^2+x_2^2=0\right\}$?

TODO 4.63. Redundant vectors in spans can be removed.

TODO 4.64. hw3: If $(w_1,w_2,\ldots,w_k)$ and $(v_1,v_2,\ldots,v_\ell)$ are two lists of vectors such that $\{w_1,w_2,\ldots,w_k\} = \{v_1,v_2,\ldots,v_\ell\}$, then $\operatorname{span}(w_1,w_2,\ldots,w_k) = \operatorname{span}(v_1,v_2,\ldots,v_\ell)$. In particular, if we reorder or duplicate the elements of a list, then its span does not change.

TODO 4.65. Rest of hw3.

TODO 4.66. The list $\left(1,x,x^2,\ldots,x^n\right)$ spans the vector space $\mathcal{P}_n$ of all polynomial functions $\mathbb{R} \to \mathbb{R}$ of degree $\leq n$. (This allows the constant-0 function, which is understood to have degree $-\infty$.)

TODO 4.67. hw3: The subspace $\left\{(x_1,x_2,\ldots,x_n)^T \in \mathbb{R}^n \mid x_1+x_2+\cdots+x_n=0\right\}$ of $\mathbb{R}^n$ is spanned by each of the two lists $(e_1-e_n,e_2-e_n,\ldots,e_{n-1}-e_n)$ and $(e_1-e_2,e_2-e_3,\ldots,e_{n-1}-e_n)$.

Definition 4.68. A vector space V is said to be *finite-dimensional* if there is a (finite) list $(v_1,v_2,\ldots,v_k)$ spanning V.

TODO 4.69. Examples: $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, $\mathbb{R}$, $\mathbb{C}$ and all their subspaces are finite-dimensional.
On the other hand, $\mathbb{R}^{\mathbb{R}}$, the set of all polynomial functionss, etc. are not finite-dimensional.

[...]

4.8. <TODO> Linear independence

References for vector spaces: [LaNaSc16, §5.2] (possibly the best one), [OlvSha06, §2.3] (focusses on the analysis-related and applied stuff) and [Heffer16, Two.II].

Definition 4.70. Let $v_1, v_2, \ldots, v_k$ be vectors in a vector space V. We say that the list $(v_1, v_2, \ldots, v_k)$ is *linearly independent* if the only k-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ of real numbers satisfying $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \overrightarrow{0}$ is $(0, 0, \ldots, 0)$.

Instead of saying "the list $(v_1, v_2, \ldots, v_k)$ is linearly independent", we can also say "the vectors $v_1, v_2, \ldots, v_k$ are linearly independent". (But keep in mind that linear independence is a property of the whole list $(v_1, v_2, \ldots, v_k)$, not of each single vector in this list. It often happens (for example) that two vectors a, b are linearly independent, and two further vectors c, d are linearly independent, but the four vectors a, b, c, d together are not linearly independent.)

TODO 4.71. Equivalent restatements of linear independence:

1. The list $(v_1, v_2, \ldots, v_k)$ is linearly independent.
2. The vectors $v_1, v_2, \ldots, v_k$ is linearly independent. (This is just another way of saying things.)
3. The only k-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ of real numbers satisfying $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \overrightarrow{0}$ is $(0, 0, \ldots, 0)$.
4. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are real numbers satisfying $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \overrightarrow{0}$, then $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$.
5. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are real numbers, **not all zero**, then $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \neq \overrightarrow{0}$.
6. For each vector $v \in V$, there exists **at most one** k-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ of real numbers satisfying $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = v$.
7. Each $i \in \{1, 2, \ldots, k\}$ satisfies $v_i \notin \operatorname{span}(v_1, v_2, \ldots, v_{i-1})$. (See Proposition below.)
8. Each $i \in \{1, 2, \ldots, k\}$ satisfies $v_i \notin \operatorname{span}(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, v_{i+2}, \ldots, v_k)$. (See Proposition below.)

TODO 4.72. Example: The standard basis vectors $e_1, e_2, \ldots, e_n$ of $\mathbb{R}^n$ are linearly independent.

TODO 4.73. Non-example: The vectors $e_1 - e_2, e_2 - e_3, e_3 - e_1$ in $\mathbb{R}^3$ are **not** linearly independent. In fact,

$$1(e_1 - e_2) + 1(e_2 - e_3) + 1(e_3 - e_1) = \overrightarrow{0}.$$

TODO 4.74. We say "linearly dependent" for "not linearly independent".

Another non-example: The vectors $e_1, \overrightarrow{0}, e_2$ in $\mathbb{R}^3$ are linearly dependent. In fact,

$$0e_1 + 1 \cdot \overrightarrow{0} + 0e_2 = \overrightarrow{0}.$$

More generally, if $\overrightarrow{0}$ appears in a list of vectors, then said list must be linearly dependent.

TODO 4.75. Another non-example: If α and β are two vectors, then (α, β, α) is linearly dependent. Indeed,

$$1\alpha + 0\beta + (-1)\alpha = \overrightarrow{0}.$$

More generally, if a vector in a list appears more than once, then said list is linearly dependent.

TODO 4.76. How to check whether a list of vectors in $\mathbb{R}^n$ (or $\mathbb{R}^{n\times m}$) is linearly independent?

You have to check whether the only solution of $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \overrightarrow{0}$ is $(0, 0, \ldots, 0)$. This can be done by Gaussian elimination.

For example, is the list $\left(\begin{pmatrix}2\\1\\3\end{pmatrix}, \begin{pmatrix}1\\3\\6\end{pmatrix}, \begin{pmatrix}2\\5\\5\end{pmatrix}\right)$ of vectors in $\mathbb{R}^3$ linearly independent?

So we are asking for the solutions of $\lambda_1 \begin{pmatrix}2\\1\\3\end{pmatrix} + \lambda_2 \begin{pmatrix}1\\3\\6\end{pmatrix} + \lambda_3 \begin{pmatrix}2\\5\\5\end{pmatrix} = \overrightarrow{0}$. This rewrites as the system $\begin{cases} 2\lambda_1 + 1\lambda_2 + 2\lambda_3 = 0; \\ 1\lambda_1 + 3\lambda_2 + 5\lambda_5 = 0; \\ 3\lambda_1 + 6\lambda_2 + 5\lambda_3 = 0 \end{cases}$. The only solution is $(\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0)$. So, yes, it is linearly independent.

TODO 4.77. Visual meaning of linear independency in $\mathbb{R}^2$ and $\mathbb{R}^3$.

TODO 4.78. If $(v_1, v_2, \ldots, v_k)$ is linearly independent, then so is every rearrangement of $(v_1, v_2, \ldots, v_k)$, and every sublist of $(v_1, v_2, \ldots, v_k)$ too.

TODO 4.79. Proposition: Let V be a vector space. Let $v_1, v_2, \ldots, v_k$ be k vectors in V. Assume that the list $(v_1, v_2, \ldots, v_k)$ is linearly dependent. Then:

(a) There exists some $i \in \{1, 2, \ldots, k\}$ such that $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_{i-1})$. [Notice that when $i = 1$, the span $\operatorname{span}(v_1, v_2, \ldots, v_{i-1})$ has to be interpreted as the "empty span" $\operatorname{span}() = \left\{\overrightarrow{0}\right\}$.]

(b) This i satisfies $\operatorname{span}(v_1, v_2, \ldots, v_k) = \operatorname{span}(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$.

TODO 4.80. Corollary: Let V be a vector space. Let $v_1, v_2, \ldots, v_k$ be k vectors in V. Then, the following are equivalent:

(a) The list $(v_1, v_2, \ldots, v_k)$ is linearly independent.

(b) No i satisfies $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_{i-1})$.

(c) No i satisfies $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$.

TODO 4.81. Thus, if a list spans a subspace U of V, then we can keep shrinking this list (by throwing out redundant elements) until we obtain a linearly independent list that also spans U. For example, the list $(e_1 - e_2, e_1, e_2 - 2e_1, e_1 + e_3, e_2 + e_3)$ spanning $\mathbb{R}^3$ can be shrunk by first throwing out $e_2 - 2e_1$ (which is $\in \operatorname{span}(e_1 - e_2, e_1)$) and then throwing out $e_2 + e_3$ (which is $\in \operatorname{span}(e_1 - e_2, e_1, e_1 + e_3)$); the result is a linearly independent list $(e_1 - e_2, e_1, e_1 + e_3)$ that still spans $\mathbb{R}^3$.

How do we find redundant entries in our list? One way to get them is just to check, for each i, whether $v_i \in \operatorname{span}(v_1, v_2, \ldots, v_{i-1})$. Another (faster) way is to pick a not-all-zeroes solution $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ of the equation $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \overrightarrow{0}$, and take the highest i for which $\lambda_i \neq 0$; then, the corresponding v_i is redundant.

Exercise: Check that the list obtained at the end of this process is indeed redundant.

TODO 4.82. The **Steinitz exchange lemma** is the claim that if $(v_1, v_2, \ldots, v_k)$ is a list of vectors in a vector space V such that $V = \operatorname{span}(v_1, v_2, \ldots, v_k)$, and if $(w_1, w_2, \ldots, w_l)$ is a linearly independent list of vectors in the same vector space V, then

(a) we have $k \geq \ell$ (that is, the spanning list is at least as long as the linearly independent list), and

(b) we can obtain a basis of V by filling up the list $(w_1, w_2, \ldots, w_\ell)$ with some vectors from $(v_1, v_2, \ldots, v_k)$. ("Filling up" means that we pick some vectors from the list $(v_1, v_2, \ldots, v_k)$ and append them to the list $(w_1, w_2, \ldots, w_\ell)$. For example, filling up the list $(w_1, w_2, \ldots, w_\ell)$ with some vectors from $(v_1, v_2, \ldots, v_k)$ can result in the list $(w_1, w_2, \ldots, w_\ell, v_2, v_4, v_9)$. But "some" can also be none; thus, we might end up with $(w_1, w_2, \ldots, w_\ell)$.)

The Wikipedia has a proof of the Steinitz exchange lemma, and it appears to be among the most readable: `https://en.wikipedia.org/wiki/Steinitz_exchange_lemma#Proof` . (Note that the Wikipedia talks of sets instead of lists, but you can just read "list" for "set" there.)

Another source is Theorem 5.2.9 in L/N/S. Notice that L/N/S are trying to avoid talking about empty lists; this forces them to take some slightly clumsy precautions (for example, the "v_1 is nonzero" condition in Lemma 5.2.7 is just there because they don't want to allow j to be 1). If you feel at home with empty lists, you can simplify their proof of Theorem 5.2.9 as well (Step 1 becomes just a particular case of Step k for k = 1). That said, I still prefer the proof on Wikipedia.

Another proof is in Olver/Shakiban, using Gaussian elimination.

[...]

4.9. <TODO> Bases and dimension

References for vector spaces: [LaNaSc16, §5.3-5.4] (possibly the best one), [OlvSha06, §2.4] (focusses on the analysis-related and applied stuff) and [Heffer16, Two.III].

Definition 4.83. Let $v_1, v_2, \ldots, v_k$ be vectors in a vector space V. We say that the list $(v_1, v_2, \ldots, v_k)$ is a *basis* of V if and only if this list spans V (that is, it satisfies $V = \operatorname{span}(v_1, v_2, \ldots, v_k)$) and is linearly independent.

Theorem 4.84. Let V be a finite-dimensional vector space. Then, any two bases of V have the same size.

TODO 4.85. Prove this theorem using the Steinitz exchange lemma.

TODO 4.86. If $(v_1, v_2, \ldots, v_k)$ is a linearly independent list in a vector space V, then $(v_1, v_2, \ldots, v_k)$ is a basis of the subspace $\operatorname{span}(v_1, v_2, \ldots, v_k)$.

TODO 4.87. Each spanning list can be shrunk to a basis. (This was done above, just without saying the word "basis".)

TODO 4.88. .....

[...]
[to be continued]

References

[Artin10] Michael Artin, *Algebra*, 2nd edition, Pearson 2010.

[BarSch73] Hans Schneider, George Phillip Barker, *Matrices and Linear Algebra*, 2nd edition, Dover 1973.

[Camero08] Peter J. Cameron, *Notes on Linear Algebra*, version 5 Sep 2008. `http://www.maths.qmul.ac.uk/~pjc/notes/linalg.pdf`

[Chen08] William Chen, *Linear Algebra*, 2011.
https://rutherglen.science.mq.edu.au/wchen/lnlafolder/lnla.html

[deBoor] Carl de Boor, *An empty exercise*. ftp://ftp.cs.wisc.edu/Approx/empty.pdf .

[Drucker12] D. Drucker, *Annotated Bibliography of Linear Algebra Books*, February 2012.
http://www.math.wayne.edu/~drucker/linalgrefsW12.pdf

[DumFoo04] David S. Dummit, Richard M. Foote, *Abstract Algebra*, 3rd edition 2004.

[Goodma15] Frederick M. Goodman, *Algebra: Abstract and Concrete*, edition 2.6.
http://homepage.divms.uiowa.edu/~goodman/algebrabook.dir/download.htm

[Grcar10] Joseph F. Grcar, *How ordinary elimination became Gaussian elimination*, Historia Mathematica, Volume 38, Issue 2, May 2011, pp. 163–218.
http://www.sciencedirect.com/science/article/pii/S0315086010000376

[Gill12] Stanley Gill Williamson, *Matrix Canonical Forms: notational skills and proof techniques*, 15 July 2015.

[Grinbe16] Darij Grinberg, *Notes on the combinatorial fundamentals of algebra*, 10 January 2019.
http://www.cip.ifi.lmu.de/~grinberg/primes2015/sols.pdf

[OlvSha06] Peter J. Olver, Chehrzad Shakiban, *Applied Linear Algebra*, Prentice Hall, 2006.
See also http://www.math.umn.edu/~olver/ala.html for corrections.

[Heffer16] Jim Hefferon, *Linear Algebra*, 2016.
http://joshua.smcvt.edu/linearalgebra/ (Scroll down to "The next edition" to download the newest version.)

[Kowals16] Emmanuel Kowalski, *Linear Algebra*, version 15 Sep 2016.
https://people.math.ethz.ch/~kowalski/script-la.pdf

[LaNaSc16] Isaiah Lankham, Bruno Nachtergaele, Anne Schilling, *Linear Algebra As an Introduction to Abstract Mathematics*, 2016.
https://www.math.ucdavis.edu/~anne/linear_algebra/mat67_course_notes.pdf

[LeLeMe16] Eric Lehman, F. Thomson Leighton, Albert R. Meyer, *Mathematics for Computer Science*, revised Wednesday 6th June, 2018, https://courses.csail.mit.edu/6.042/spring18/mcs.pdf .

[m.se709196] Daniela Diaz and others, *Definition of General Associativity for binary operations*, math.stackexchange question #709196 . http://math.stackexchange.com/q/709196

[Stanle12] Richard P. Stanley, *Enumerative Combinatorics, Volume 1*, 2nd edition, CUP 2012. See http://math.mit.edu/~rstan/ec/ec1/ for a preliminary version.

[Treil15] Sergei Treil, *Linear Algebra Done Wrong*, 2017. https://www.math.brown.edu/~treil/papers/LADW/LADW.html

[Wildon16] Mark Wildon, *MT182 Matrix Algebra*, version 14 May 2016. http://www.ma.rhul.ac.uk/~uvah099/Maths/MatrixAlgebra15/MT1822015Notes.pdf

[Zuker14] M. Zuker, *Binary Operations & General Associativity*, http://snark.math.rpi.edu/Teaching/MATH-4010/Binary_Ops.pdf