Note 7

Vectors and matrices

7.1 Vectors in $\mathbb{F}^n$

As in the last chapter, we will denote by $\mathbb{F}$ a field. What we will explain works over any field, but the reader can just think of $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. When describing solutions to systems of linear equations, we already worked with $\mathbb{F}^n$, the set of all n-tuples with entries in $\mathbb{F}$. Also, we already explained that such an n-tuple is for convenience often identified with an $n \times 1$ matrix. This just means that:

$$(v_1, \dots, v_n) \text{ can also be written as } \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

When an n-tuple is written as an $n \times 1$ matrix, we say that the n-tuple is written in *vector* form. Elements in $\mathbb{F}^n$ are therefore called *vectors* with n entries from $\mathbb{F}$. If all entries of such a vector are zero, we call that vector the *zero vector* of $\mathbb{F}^n$.

Remark 7.1

Elements in $\mathbb{F}^{n\times 1}$ are sometimes called *column vectors*, while likewise elements from $\mathbb{F}^{1\times n}$ are called *row vectors*.

We have already used in the previous chapter that there is a natural way to add two vectors $\mathbf{v}$ and $\mathbf{w}$ from $\mathbb{F}^n$, and also that one can multiply a vector from $\mathbb{F}^n$ with an element $c \in \mathbb{F}$, often called a *scalar* in this context, since multiplying a vector by a constant can be thought of as scaling the vector. More precisely, addition of vectors is defined as:

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{bmatrix} \tag{7-1}$$

and the product of a scalar with a vector as:

$$c \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} c \cdot v_1 \\ \vdots \\ c \cdot v_n \end{bmatrix}. \tag{7-2}$$

As in the case for matrices, we will often use boldface fonts for vectors and typically use letters such as $\mathbf{u}, \mathbf{v}, \mathbf{w}$. For future reference, we state the following theorem, which collects a number of properties of the vector addition and scalar multiplication:

|||| Theorem 7.2

Let $\mathbb{F}$ be a field, $c, d \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{F}^n$. Then

1. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. $c \cdot (d \cdot \mathbf{u}) = (c \cdot d) \cdot \mathbf{u}$
4. $c \cdot (\mathbf{u} + \mathbf{v}) = c \cdot \mathbf{u} + c \cdot \mathbf{v}$
5. $(c + d) \cdot \mathbf{u} = c \cdot \mathbf{u} + d \cdot \mathbf{u}$

We leave the proof of this theorem out.

Now that we have vectors at out disposal, we will be able to discuss further properties they have. We start with an example.

|||| Example 7.3

Consider the vectors

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathbb{R}^2.$$

1. Compute $4 \cdot \mathbf{u} + 3 \cdot \mathbf{v}$.
2. Find c and d such that $c \cdot \mathbf{u} + d \cdot \mathbf{v} = \mathbf{0}$, where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^2$.

Answer:

1. Using the definition of scalar multiplication and vector addition, we find
$$4 \cdot \mathbf{u} + 3 \cdot \mathbf{v} = 4 \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix} + \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 11 \end{bmatrix}.$$
2. We have
$$c \cdot \mathbf{u} + d \cdot \mathbf{v} = c \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} + d \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} c \\ 2c \end{bmatrix} + \begin{bmatrix} 2d \\ d \end{bmatrix} = \begin{bmatrix} c+2d \\ 2c+d \end{bmatrix}.$$
If we want the outcome to be the zero vector, this means that we need to solve the homogeneous system of linear equations:
$$\begin{cases} c + 2d = 0 \\ 2c + d = 0 \end{cases}.$$
Now subtracting the first equation twice from the second equation, in other words performing the elementary row operation $R_2 \leftarrow R_2 - 2 \cdot R_1$, we obtain the system
$$\begin{cases} c + 2d = 0 \\ 0c - 3d = 0 \end{cases}.$$
We could continue and bring the system in reduced row echelon form, but it is already clear now that the only solution is $c = d = 0$.

An expression like $4 \cdot \mathbf{u} + 3 \cdot \mathbf{v}$ is called a *linear combination* of the vectors $\mathbf{u}$ and $\mathbf{v}$. More general, given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ and scalars $c_1, \ldots, c_n \in \mathbb{F}$, an expression of the form
$$c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$$
is called a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The second part of the example implies that apparently the only linear combination of the vectors $\mathbf{u}$ and $\mathbf{v}$ given there that is equal to the zero vector, is the linear combination $0 \cdot \mathbf{u} + 0 \cdot \mathbf{v}$. In general, a sequence of vectors can have this property. This is captured in the following:

Definition 7.4

A sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ is called *linearly independent* if and only if the equation $c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n = \mathbf{0}$ can only hold if $c_1 = \cdots = c_n = 0$.
If the sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ is not linearly independent, one says that it is *linearly dependent*.

In other words, a sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ is linearly independent if and only if the only linear combination of the vectors that is equal to the zero vector, occurs for $c_1 = \cdots = c_n = 0$. Using some logical expressions, linear independence of a sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ can be phrased as follows:

$$\text{for all } c_1, \ldots, c_n \in \mathbb{F} \text{ one has: } c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n = \mathbf{0} \Rightarrow c_1 = \cdots = c_n = 0. \tag{7-3}$$

Similarly, linear dependence of the sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ can be phrased in the following way:

$$\text{there exist } c_1, \ldots, c_n \in \mathbb{F} \text{ such that: } c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n = \mathbf{0} \ \wedge \ \text{ not all } c_i \text{ are zero.} \tag{7-4}$$

Instead of saying that a sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly (in)dependent, it is also quite common to simply say that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly (in)dependent. We will use this way of phrasing things quite often.

Example 7.5

The sequence of vectors consisting of

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \in \mathbb{R}^2$$

is linearly dependent. Indeed, since $\mathbf{v} = 2 \cdot \mathbf{u}$, we see that $(-2) \cdot \mathbf{u} + \mathbf{v} = \mathbf{0}$.

This example illustrates a more general principle: two vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent if and only if one is a scalar multiple of the other. Indeed, if for example $\mathbf{u} = c \cdot \mathbf{v}$, then $1 \cdot \mathbf{u} + (-c) \cdot \mathbf{v} = \mathbf{0}$, showing that the vectors are linearly dependent. Similarly, if $\mathbf{v} = c \cdot \mathbf{u}$, then $(-c) \cdot \mathbf{u} + 1 \cdot \mathbf{v} = \mathbf{0}$, again showing that the vectors are linearly dependent. Conversely if the vectors are linearly dependent, there exist $c, d \in \mathbb{F}$, not both zero, such that $c \cdot \mathbf{u} + d \cdot \mathbf{v} = \mathbf{0}$. If $c \neq 0$, then we obtain that $\mathbf{u} = (-d/c) \cdot \mathbf{v}$ so that $\mathbf{v}$ is a scalar multiple of $\mathbf{u}$. If $d \neq 0$, we similarly obtain that $\mathbf{v} = (-c/d) \cdot \mathbf{u}$ showing that in that case $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$. Hence intuitively, one can say that two vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent if and only if there is a line through the origin containing both $\mathbf{u}$ and $\mathbf{v}$.

Example 7.6

The sequence of vectors consisting of

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathbb{R}^2$$

is linearly independent. Indeed, we have seen in Example 7.3 that the equation $c \cdot \mathbf{u} + d \cdot \mathbf{v} = \mathbf{0}$ implies that $c = d = 0$.

This example suggests that the linear independence of a sequence of vectors can be investigated using the theory of systems of linear equations. This is indeed the case and the general result is the following:

Lemma 7.7

Let vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ be given and let $\mathbf{A} \in \mathbb{F}^{m \times n}$ be the $m \times n$ matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$, that is

$$\mathbf{A} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \ldots & \mathbf{v}_n \\ | & & | \end{bmatrix}.$$

The sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly independent if and only if the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$ only has the zero vector $\mathbf{0} \in \mathbb{F}^n$ as solution.

Proof. First suppose that the sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly independent and let $(c_1, \ldots, c_n) \in \mathbb{F}^n$ be a solution to the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$. This system can directly be rewritten as the equation $c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n = \mathbf{0}$. Using that we assumed that the sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly independent, we see that $(c_1, \ldots, c_n) = (0, \ldots, 0)$.

Now conversely, assume that the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$ only has the zero vector $\mathbf{0} \in \mathbb{F}^n$ as solution. If $(c_1, \ldots, c_n) \in \mathbb{F}^n$ satisfies $c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n = \mathbf{0}$, then we can immediately conclude that $(c_1, \ldots, c_n)$ is also a solution to the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$. But then by assumption, we may conclude that $(c_1, \ldots, c_n) = (0, \ldots, 0)$. □

This lemma leads to a short characterisation of linear independence:

|||| Theorem 7.8

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{F}^m$ be given and let $\mathbf{A} \in \mathbb{F}^{m \times n}$ be the matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The sequence of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is linearly independent if and only if the matrix $\mathbf{A}$ has rank n.

Proof. This follows from Corollary 6.30 and Lemma 7.7. □

|||| Example 7.9

Consider the following three vectors in $\mathbb{C}^3$:

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 1+i \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 0 \\ 1+i \\ 0 \end{bmatrix}, \text{ and } \mathbf{w} = \begin{bmatrix} 1+i \\ -1+5i \\ 2i \end{bmatrix}.$$

1. Are the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ linearly independent?
2. Are the vectors $\mathbf{u}$, $\mathbf{v}$ linearly independent?
3. Is the vector $\mathbf{u}$ linearly independent?

Answer: The general strategy for this type of questions is to use Theorem 7.8. Recall that in order to compute the rank of a matrix, it is by Definition 6.22, the definition of the rank of a matrix, enough to compute its reduced row echelon form. Now let us answer the three questions, one at the time.

1. Theorem 7.8 implies that to find the answer, we should determine the rank of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 1+i & -1+5i \\ 1+i & 0 & 2i \end{bmatrix}.$$

We have

$$\begin{bmatrix} 1 & 0 & 1+i \\ 0 & 1+i & -1+5i \\ 1+i & 0 & 2i \end{bmatrix} \xrightarrow[R_3 \leftarrow R_3-(1+i)\cdot R_1]{} \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 1+i & -1+5i \\ 0 & 0 & 0 \end{bmatrix}$$
$$\xrightarrow[R_2 \leftarrow (1+i)^{-1}\cdot R_2]{} \begin{bmatrix} 1 & 0 & 1+i \\ 0 & 1 & 2+3i \\ 0 & 0 & 0 \end{bmatrix}.$$

We can conclude that $\rho(\mathbf{A}) = 2$, which is less than three, the number of vectors we are considering. Hence the vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ are linearly dependent.

2. In this case, we should compute the rank of the matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1+i \\ 1+i & 0 \end{bmatrix}.$$

Using exactly the same elementary row operations as when solving the first questions, we find that the reduced row echelon form of $\mathbf{B}$ is the matrix

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

In particular, $\rho(\mathbf{B}) = 2$, which is equal to the number of vectors we are considering. Hence the vectors $\mathbf{u}$, $\mathbf{v}$ are linearly independent.

3. If we only consider the vector $\mathbf{u}$, we need to determine the rank of the matrix

$$\mathbf{C} = \begin{bmatrix} 1 \\ 0 \\ 1+i \end{bmatrix}.$$

This matrix has rank one, since the one column this matrix has, is not the zero column. We can conclude that the sequence consisting of the vector $\mathbf{u}$ is linearly independent. In general, a sequence consisting of only one vector $\mathbf{u} \in \mathbb{F}^m$ is linearly independent if and only if $\mathbf{u} \neq \mathbf{0}$.

7.2 Matrices and vectors

When studying systems of linear equations, we introduced the notion of a matrix. A matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ was introduced as a rectangular scheme containing $m \times n$ elements from a given field $\mathbb{F}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

Sometimes one just writes $\mathbf{A} = [a_{ij}]_{1 \le i \le m, 1 \le j \le m}$ for brevity. When a matrix is given in this form, the element a_{ij}, sometimes also written as $a_{i,j}$, is the entry in row i and column j of the matrix $\mathbf{A}$. It is also common to denote this entry by $\mathbf{A}_{ij}$ or $\mathbf{A}_{i,j}$. The matrix $\mathbf{A}$ given above has m rows: $[a_{i1} \ \dots \ a_{in}]$ for $i = 1, \dots, m$ and n columns:

$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix} \text{ for } j = 1, \dots, n.$$

We will call rows of a matrix *row vectors* and similarly columns of a matrix *column vectors*.

It turns out to be extremely useful to be able to multiply a matrix with a vector. We define the following:

|||| Definition 7.10

Let $\mathbf{A} = (a_{ij})_{1 \le i \le m, 1 \le j \le m} \in \mathbb{F}^{m \times n}$ be a matrix and $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{F}^n$ a vector. Then we define $\mathbf{A} \cdot \mathbf{v} \in \mathbb{F}^m$ as follows:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} a_{11} \cdot v_1 + \cdots a_{1n} \cdot v_n \\ \vdots \\ a_{m1} \cdot v_1 + \cdots a_{mn} \cdot v_n \end{bmatrix}$$

Note that we can not multiply any matrix with any vector. Their sizes have to "fit": the number of columns of the matrix has to be the same as the number of entries in the vector. If this is not the case, the corresponding matrix-vector multiplication is not defined.

|||| Example 7.11

Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 7 \\ -1 \\ -2 \end{bmatrix}.$$

Compute $\mathbf{A} \cdot \mathbf{v}$.

Answer: Using Definition 7.10, we find that:

$$\mathbf{A} \cdot \mathbf{v} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \cdot 7 + 2 \cdot (-1) + 3 \cdot (-2) \\ 4 \cdot 7 + 5 \cdot (-1) + 6 \cdot (-2) \end{bmatrix} = \begin{bmatrix} -1 \\ 11 \end{bmatrix}.$$

Note that the matrix vector product occurs very naturally when considering a system of linear equations. A system of linear equations

$$\begin{cases} a_{11} \cdot x_1 & + & \cdots & + & a_{1n} \cdot x_n & = & b_1 \\ & & \vdots & & & & \vdots \\ a_{m1} \cdot x_1 & + & \cdots & + & a_{mn} \cdot x_n & = & b_m \end{cases}$$

can be expressed as

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \tag{7-5}$$

Now that we have defined a matrix vector product, one may wonder if more generally, matrices can be multiplied with each other as well. The answer turns out to be yes, provided again that their sizes fit. More precisely, we can do the following:

|||| Definition 7.12

Let $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{n \times \ell}$. Suppose that the columns of $\mathbf{B}$ are given by $\mathbf{b}_1, \ldots, \mathbf{b}_\ell \in \mathbb{F}^n$, that is to say, suppose that

$$\mathbf{A} = \begin{bmatrix} | & & | \\ \mathbf{b}_1 & \ldots & \mathbf{b}_\ell \\ | & & | \end{bmatrix}.$$

Then we define

$$\mathbf{A} \cdot \mathbf{B} = \begin{bmatrix} | & & | \\ \mathbf{A} \cdot \mathbf{b}_1 & \ldots & \mathbf{A} \cdot \mathbf{b}_\ell \\ | & & | \end{bmatrix}.$$

Note that the matrix product $\mathbf{A} \cdot \mathbf{B}$ is defined only if the number of columns of $\mathbf{A}$ is the same as the number of rows of $\mathbf{B}$. If these numbers match, then $\mathbf{A} \cdot \mathbf{B}$ is a matrix with m rows and ℓ columns. In other words, if $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{n \times \ell}$, then $\mathbf{A} \cdot \mathbf{B} \in \mathbb{F}^{m \times \ell}$.

Another way to look at the definition of the matrix product is to give a formula for the entries of the product $\mathbf{A} \cdot \mathbf{B}$ one at the time. Let us say, that we want to find a formula for the (i, j)-th entry of the product, $(\mathbf{A} \cdot \mathbf{B})_{i,j}$, that is to say, the entry in row i and column j. This amounts to determining the i-th entry of the product $\mathbf{A} \cdot \mathbf{b}_j$, where $\mathbf{b}_j$ is the j-th column of $\mathbf{B}$. This in turn is exactly the same as the outcome of multiplying the i-th row of the matrix $\mathbf{A}$ with the j-th column of the matrix $\mathbf{B}$. Since the i-th row of $\mathbf{A}$ can be written as equals $[a_{i1} \ \ldots \ a_{in}]$ and the j column of $\mathbf{B}$ as

$$\mathbf{b}_j = \begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix},$$

we see that

$$(\mathbf{A} \cdot \mathbf{B})_{i,j} = [a_{i1} \ \ldots \ a_{in}] \cdot \begin{bmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{bmatrix} = a_{i1} \cdot b_{1j} + \cdots + a_{in} \cdot b_{nj}.$$

Using the summation symbol from Section 5.3, we can rewrite this formula as follows:

$$(\mathbf{A} \cdot \mathbf{B})_{i,j} = \sum_{r=1}^{n} a_{ir} \cdot b_{rj}. \tag{7-6}$$

|||| Example 7.13

In this example, let $\mathbb{F} = \mathbb{R}$ and write

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 7 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}.$$

Compute, if possible, the matrix products $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{B} \cdot \mathbf{A}$.

Answer: First consider the matrix product $\mathbf{A} \cdot \mathbf{B}$. Since $\mathbf{A} \in \mathbb{R}^{2 \times 3}$ and $\mathbf{B} \in \mathbb{R}^{3 \times 3}$, the product $\mathbf{A} \cdot \mathbf{B}$ is defined. We have already computed the product of $\mathbf{A}$ and the first column of $\mathbf{B}$ in Example 7.11, so we will not repeat those computations. Taking that into account, we obtain that:

$$\mathbf{A}\cdot\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \cdot \begin{bmatrix} 7 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1\cdot 0+2\cdot 1+3\cdot 0 & 1\cdot 0+2\cdot 0+3\cdot 1 \\ 11 & 4\cdot 0+5\cdot 1+6\cdot 0 & 4\cdot 0+5\cdot 0+6\cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 2 & 3 \\ 11 & 5 & 6 \end{bmatrix}.$$

Now let us consider the matrix product $\mathbf{B}\cdot\mathbf{A}$. Since $\mathbf{B}$ has three columns and $\mathbf{A}$ has two rows, the matrix product $\mathbf{B}\cdot\mathbf{A}$ is not defined.

This example shows that in general $\mathbf{A}\cdot\mathbf{B} \neq \mathbf{B}\cdot\mathbf{A}$. In other words, matrix multiplication is not commutative. In fact, as we have just seen, it may even happen that one of the products is not defined. Even if both products are defined, the order of the matrices still matters and $\mathbf{A}\cdot\mathbf{B} \neq \mathbf{B}\cdot\mathbf{A}$ in general. Consider for example $\mathbf{A} = [1\,0]$ and $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then

$$\mathbf{A}\cdot\mathbf{B} = [1\,0]\cdot\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 1\cdot 0 + 0\cdot 1 = 0 \text{ and } \mathbf{B}\cdot\mathbf{A} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\cdot[1\,0] = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Let us define addition of matrices as well.

|||| Definition 7.14

Let $\mathbf{A}, \mathbf{A}' \in \mathbb{F}^{m\times n}$ be given, say

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \text{ and } \mathbf{A}' = \begin{bmatrix} a'_{11} & \dots & a'_{1n} \\ \vdots & & \vdots \\ a'_{m1} & \dots & a'_{mn} \end{bmatrix}.$$

Then we define $\mathbf{A} + \mathbf{A}'$ as follows:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} a'_{11} & \dots & a'_{1n} \\ \vdots & & \vdots \\ a'_{m1} & \dots & a'_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+a'_{11} & \dots & a_{1n}+a'_{1n} \\ \vdots & & \vdots \\ a_{m1}+a'_{m1} & \dots & a_{mn}+a'_{mn} \end{bmatrix}.$$

Addition of matrices is only defined if they have the same sizes. On the level of entries, we can see that $(\mathbf{A}+\mathbf{A}')_{i,j} = a_{ij} + a'_{ij}$. Addition and multiplication of matrices satisfy many similar rules as summation and multiplication of real or complex numbers. We collect some in the following theorem. The main exception, as already mentioned before, is that matrix multiplication is not commutative in general.

Theorem 7.15

Let $\mathbb{F}$ be a field. Then

1. $\mathbf{A}+\mathbf{A}' = \mathbf{A}'+\mathbf{A}$ for all $\mathbf{A}, \mathbf{A}' \in \mathbb{F}^{m\times n}$.
2. $(\mathbf{A}+\mathbf{A}')+\mathbf{A}'' = \mathbf{A}+(\mathbf{A}'+\mathbf{A}'')$ for all $\mathbf{A}, \mathbf{A}', \mathbf{A}'' \in \mathbb{F}^{m\times n}$.
3. $\mathbf{A}\cdot(\mathbf{B}\cdot\mathbf{C}) = (\mathbf{A}\cdot\mathbf{B})\cdot\mathbf{C}$ for all $\mathbf{A} \in \mathbb{F}^{m\times n}$, $\mathbf{B} \in \mathbb{F}^{n\times \ell}$, and $\mathbf{C} \in \mathbb{F}^{\ell\times k}$.
4. $\mathbf{A}\cdot(\mathbf{B}+\mathbf{B}') = \mathbf{A}\cdot\mathbf{B}+\mathbf{A}\cdot\mathbf{B}'$ for all $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\mathbf{B}, \mathbf{B}' \in \mathbb{F}^{n\times \ell}$.
5. $(\mathbf{A}+\mathbf{A}')\cdot\mathbf{B} = \mathbf{A}\cdot\mathbf{B}+\mathbf{A}'\cdot\mathbf{B}$ for all $\mathbf{A}, \mathbf{A}' \in \mathbb{F}^{m\times n}$ and $\mathbf{B} \in \mathbb{F}^{n\times \ell}$.

Proof. We will prove the third item only and leave the other parts to the reader. Using equation (7-6) for the product $(\mathbf{B}\cdot\mathbf{C})$, we obtain that $(\mathbf{B}\cdot\mathbf{C})_{s,j} = \sum_{r=1}^{\ell} b_{sr}\cdot c_{rj}$. Using this and equation (7-6) for the product $\mathbf{A}\cdot(\mathbf{B}\cdot\mathbf{C})$ and rewriting the resulting expression, we see that:

$$
\begin{aligned}
(\mathbf{A}\cdot(\mathbf{B}\cdot\mathbf{C}))_{i,j} &= \sum_{s=1}^{n} a_{is}\cdot(\mathbf{B}\cdot\mathbf{C})_{s,j}\\
&= \sum_{s=1}^{n} a_{is}\cdot\sum_{r=1}^{\ell} b_{sr}\cdot c_{rj}\\
&= \sum_{s=1}^{n}\sum_{r=1}^{\ell} a_{is}\cdot(b_{sr}\cdot c_{rj})\\
&= \sum_{s=1}^{n}\sum_{r=1}^{\ell} (a_{is}\cdot b_{sr})\cdot c_{rj}\\
&= \sum_{r=1}^{\ell}\sum_{s=1}^{n} (a_{is}\cdot b_{sr})\cdot c_{rj}\\
&= \sum_{r=1}^{\ell}\left(\sum_{s=1}^{n} a_{is}\cdot b_{sr}\right)\cdot c_{rj}\\
&= \sum_{r=1}^{\ell}(\mathbf{B}\cdot\mathbf{C}))_{i,r}\cdot c_{rj}\\
&= ((\mathbf{A}\cdot\mathbf{B})\cdot\mathbf{C})_{i,j}.
\end{aligned}
$$

□

We finish this section by explaining two more operations on matrices. We have already seen that vectors can be multiplied with a scalar. The generalisation to matrices is immediate: for $c \in \mathbb{F}$ and

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{F}^{m\times n}, \text{ we define } c\cdot\mathbf{A} = \begin{bmatrix} c\cdot a_{11} & \cdots & c\cdot a_{1n} \\ \vdots & & \vdots \\ c\cdot a_{m1} & \cdots & c\cdot a_{mn} \end{bmatrix}. \tag{7-7}$$

Finally, there is a way to reverse the roles of rows and columns in a matrix $\mathbf{A}$. This is simply done by taking the *transpose* of a matrix, which is denoted by $\mathbf{A}^T$. More precisely, given

$$\mathbf{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \in \mathbb{F}^{m\times n}, \text{ we define } \mathbf{A}^T = \begin{bmatrix} a_{11} & \dots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \dots & a_{mn} \end{bmatrix}. \tag{7-8}$$

|||| Example 7.16

Let the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

be given. Compute $\mathbf{A}^T$.

Answer:

We have

$$\mathbf{A}^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Note that if $\mathbf{A} \in \mathbb{F}^{m\times n}$, then $\mathbf{A}^T \in \mathbb{F}^{n\times m}$. On the level of entries, we simply have that the (i,j)-th entry of $\mathbf{A}^T$ is equal to the (j,i)-th entry of $\mathbf{A}$.

The transpose behaves well with respect to matrix additions and matrix products. More precisely, we have the following theorem.

|||| Theorem 7.17

Let $\mathbb{F}$ be a field. Then

1. $(\mathbf{A}^T)^T = \mathbf{A}$ for all $\mathbf{A} \in \mathbb{F}^{m\times n}$.
2. $(\mathbf{A} + \mathbf{A}')^T = \mathbf{A}^T + (\mathbf{A}')^T$ for all $\mathbf{A}, \mathbf{A}' \in \mathbb{F}^{m\times n}$.
3. $(\mathbf{A} \cdot \mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$ for all $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\mathbf{B} \in \mathbb{F}^{n\times \ell}$.

Proof. We only show the first item. In general, the (i,j)-th entry of $\mathbf{B}^T$ is equal to the (j,i)-th entry of $\mathbf{B}$ for any matrix $\mathbf{B}$. Applying this first for the matrix $\mathbf{A}^T$, then for the matrix $\mathbf{A}$, we obtain that $((\mathbf{A}^T)^T)_{i,j} = (\mathbf{A}^T)_{j,i} = (\mathbf{A})_{i,j}$. This shows that the matrices $\mathbf{A}^T$ and $\mathbf{A}$ have exactly the same entries and hence that they are equal. □

It is important to remember the order of multiplication in item 3 before and after transposing. In some sense, transposing reverses the order of the terms in a product. There is a good reason for this. Given matrices $\mathbf{A} \in \mathbb{F}^{m \times n}$ and $\mathbf{B} \in \mathbb{F}^{n \times \ell}$, the product $\mathbf{A}^T \cdot \mathbf{B}^T$ is in general not even defined! Indeed, the number of columns in $\mathbf{A}^T$ is m, while the number of rows in $\mathbf{B}^T$ is ℓ. However, the product $\mathbf{B}^T \cdot \mathbf{A}^T$ makes perfect sense, since the number of columns in $\mathbf{B}^T$ is n, which is the same as the number of rows in $\mathbf{A}^T$. Though these observations do not prove item three from Theorem 7.17, they do explain why it is quite natural that the multiplication order is given as it is.

7.3 Square matrices

If the number of rows and columns of a matrix are the same, it is called a *square* matrix. In other words, a matrix $\mathbf{A}$ is a square matrix, if $\mathbf{A} \in \mathbb{F}^{n \times n}$ for some positive integer n. Given n, the $n \times n$ matrix $\mathbf{I}_n$, called the *identity* matrix, is the matrix having 1's on its main diagonal, and 0's everywhere else. So for $n = 4$, we have for example

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This matrix is called the identity matrix because it has no effect on a vector when multiplied (from the left) with that vector. More precisely, a direct calculation shows that $\mathbf{I}_n \cdot \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. Hence the function $L : \mathbb{F}^n \to \mathbb{F}^n$ defined by $L(\mathbf{v}) = \mathbf{I}_n \cdot \mathbf{v}$ is just the identity function. With this matrix in place, the following definition makes sense:

|||| Definition 7.18

A square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ is called *invertible* if there exists a matrix $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n.$$

The matrix $\mathbf{B}$, if it exists, is called the inverse matrix of $\mathbf{A}$ and denoted by $\mathbf{A}^{-1}$.

Inverse matrices will appear in many situations later on, but already when solving some systems of linear equations, they can be handy. Suppose for example, that one wants to solve the system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$, with a square coefficient matrix $\mathbf{A}$, vector of variables $\mathbf{x} = (x_1, \ldots, x_n)$ and righthand-side $\mathbf{b} = (b_1, \ldots, b_n)$. If the coefficient matrix $\mathbf{A}$ has an inverse, we can multiply from the left with $\mathbf{A}^{-1}$ and simplify: $\mathbf{A}^{-1} \cdot (\mathbf{A} \cdot \mathbf{x}) = (\mathbf{A}^{-1} \cdot \mathbf{A}) \cdot \mathbf{x} = \mathbf{I}_n \cdot \mathbf{x}$. But this means that the equation $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ implies that

$\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}$. Conversely, if $\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}$, then by multiplying with $\mathbf{A}$ from the left, we obtain that $\mathbf{A} \cdot \mathbf{x} = \mathbf{A} \cdot (\mathbf{A}^{-1} \cdot \mathbf{b}) = (\mathbf{A} \cdot \mathbf{A}^{-1}) \cdot \mathbf{b} = \mathbf{I}_n \cdot \mathbf{b} = \mathbf{b}$. Hence we have shown that:

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} \text{ if and only if } \mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{b}, \text{ provided } \mathbf{A}^{-1} \text{ exists.} \tag{7-9}$$

This observation actually has a nice consequence about the rank of invertible matrices:

Lemma 7.19

Let $\mathbf{A} \in \mathbb{F}^{n \times n}$ be given and suppose that its inverse matrix exists. Then $\rho(\mathbf{A}) = n$.

Proof. Equation (7-9) implies that the homogeneous system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{0}$ only has the solution $\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{0} = \mathbf{0}$. But then by Corollary 6.30, the rank of $\mathbf{A}$ is equal to n. □

More is true, but we will return to that later. The question is now how to figure out when a matrix has an inverse and if it does, how to compute it. We will first find an algorithmic answer and after that describe a theoretical characterisation of invertible matrices.

What we will do first, is to find an algorithm that for a given $n \times n$ matrix $\mathbf{A}$, computes an $n \times n$ matrix $\mathbf{B}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ if it exists. Hence the outcome of the algorithm will either be that such a $\mathbf{B}$ does not exist, or it will return such a $\mathbf{B}$. Note that according to Definition 7.18, the inverse of $\mathbf{A}$, here denoted by $\mathbf{B}$, should satisfy $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ *and* $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$. Fortunately, it turns out that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ *implies* $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$, so that the algorithm we are about to describe indeed will compute the inverse matrix $\mathbf{B} = \mathbf{A}^{-1}$, provided it exists.

Let us denote the i-th column of the identity matrix $\mathbf{I}_n$ by $\mathbf{e}_i$ for $i = 1, \ldots, n$. So for example for $n = 4$, we have

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

The idea of the algorithm to find inverse matrices is the following: we are trying to find a matrix $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ for a given $\mathbf{A} \in \mathbb{F}^{n \times n}$. Now let us denote the columns of $\mathbf{B}$ as $\mathbf{b}_1, \ldots, \mathbf{b}_n$. The i-th column of $\mathbf{A} \cdot \mathbf{B}$ is by definition of the matrix product equal to $\mathbf{A} \cdot \mathbf{b}_i$, while the i-th column of $\mathbf{I}_n$ is equal to $\mathbf{e}_i$. Hence $\mathbf{A} \cdot \mathbf{b}_i = \mathbf{e}_i$ for all

i between 1 and n. Conversely, if $\mathbf{A} \cdot \mathbf{b}_i = \mathbf{e}_i$ for all i between 1 and n, then the matrices $\mathbf{A} \cdot \mathbf{B}$ and $\mathbf{I}_n$ have the same columns, whence $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$. Therefore we see that

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n \text{ if and only if } \mathbf{A} \cdot \mathbf{b}_i = \mathbf{e}_i \text{ for all } i \text{ between 1 and } n.$$

Therefore, we can find $\mathbf{b}_i$, by solving the inhomogeneous system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$.

From the theory from the previous chapter, we see that to figure out if the system of equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$, it is enough to compute the reduced row echelon form of the augmented matrix $[\mathbf{A}|\mathbf{e}_i]$. If $\rho(\mathbf{A}) = \rho([\mathbf{A}|\mathbf{e}_i])$, then according to Corollary 6.27, there exists a solution and otherwise not. Hence precisely if for all i between 1 and n it holds that $\rho(\mathbf{A}) = \rho([\mathbf{A}|\mathbf{e}_i])$, we will be able to find a matrix $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$.

Now, one could deal with the system of equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$ for one i at the time and in that way compute one column of the matrix $\mathbf{B}$ at the time, if it exists. However, the first part of the corresponding augmented matrices is always the same, namely $\mathbf{A}$. Therefore, it is faster to deal with all n systems at the same time by computing the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{e}_1|\mathbf{e}_2|\dots|\mathbf{e}_n] = [\mathbf{A}|\mathbf{I}_n]$.

Hence the algorithm of how to determine if a square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ has an inverse, and if yes how to compute it, is the following:

1. Compute the reduced row echelon form of the $n \times 2n$ matrix $[\mathbf{A}|\mathbf{I}_n]$. This can be done using elementary row operations, just as we did in Section 6.3
2. If the resulting reduced row echelon form is not of the form $[\mathbf{I}_n|\mathbf{B}]$, conclude that $\mathbf{A}$ does not have an inverse.
3. If it is of the form $[\mathbf{I}_n|\mathbf{B}]$, conclude that $\mathbf{A}$ does have an inverse, namely $\mathbf{A}^{-1} = \mathbf{B}$.

To see how this works in practice, let us consider two examples.

|||| Example 7.20

Let $\mathbb{F} = \mathbb{R}$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Determine whether or not this matrix has an inverse and if yes, compute it.

Answer: First we determine the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{I}_2]$. We obtain:

$$[\mathbf{A}|\mathbf{I}_2] = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{bmatrix} \underset{R_2 \leftarrow R_2 - 3 \cdot R_1}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{bmatrix}$$

$$\underset{R_2 \leftarrow (-1/2) \cdot R_2}{\longrightarrow} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 3/2 & -1/2 \end{bmatrix} \underset{R_1 \leftarrow R_1 - 2 \cdot R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{bmatrix}.$$

Hence we conclude that $\mathbf{A}$ has an inverse, namely

$$\mathbf{A}^{-1} = \begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix}.$$

|||| Example 7.21

Let $\mathbb{F} = \mathbb{R}$ and

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{bmatrix}.$$

Determine whether or not this matrix has an inverse and if yes, compute it.

Answer: We start determining the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{I}_3]$. We obtain:

$$\begin{aligned}
[\mathbf{A}|\mathbf{I}_3] &= \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 5 & 7 & 9 & 0 & 0 & 1 \end{bmatrix} \\
\underset{R_2 \leftarrow R_2 - 4 \cdot R_1}{\longrightarrow} & \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 5 & 7 & 9 & 0 & 0 & 1 \end{bmatrix} \\
\underset{R_3 \leftarrow R_3 - 5 \cdot R_1}{\longrightarrow} & \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -3 & -6 & -5 & 0 & 1 \end{bmatrix} \\
\underset{R_3 \leftarrow R_3 - R_2}{\longrightarrow} & \begin{bmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{bmatrix}.
\end{aligned}$$

Even though we have not found the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_3]$ yet, we already found an echelon form of it. The pivots can already be read off and are contained in the first, second, and fourth columns of the matrix. When proceeding to find the reduced row echelon form, the first three entries of the third row will remain zero. The reader is encouraged to compute the reduced echelon form and see that this indeed is true. Hence the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_3]$ will not be of the form $[\mathbf{I}_3|\mathbf{B}]$. We conclude that the matrix $\mathbf{A}$ does not have an inverse.

In principle, we now have an algorithm that can determine if a square matrix has an inverse and if yes, computes it. However, we have not shown that the algorithm is correct. In other words, if we follow the steps of the algorithm, will the outcome always be what it should be? First of all, we should make sure that if the reduced row echelon form of the $n \times 2n$ matrix $[\mathbf{A}|\mathbf{I}_n]$ is not of the form $[\mathbf{I}_n|\mathbf{B}]$, then $\mathbf{A}$ indeed has no inverse. And second of all, we should make sure that if a matrix $\mathbf{B} \in \mathbb{F}^{n\times n}$ satisfies $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$, then also $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$, so that we indeed can conclude that $\mathbf{B}$ is the inverse of $\mathbf{A}$. We will address these issues in the rest of this section. It turns out that everything is as it should be and one can show that:

$$\begin{aligned}\mathbf{A}^{-1} \text{ exists} \quad &\Leftrightarrow \quad \text{the reduced row echelon form of } [\mathbf{A}|\mathbf{I}_n] \text{ is of the form } [\mathbf{I}_n|\mathbf{B}] \\ &\Leftrightarrow \quad \rho(\mathbf{A}) = n \text{ (that is: the rank of } \mathbf{A} \text{ is } n). \end{aligned} \tag{7-10}$$

A reader willing to accept this without proof can skip the remainder of this section, but for the other readers we will give a proof below.

|||| Theorem 7.22

Let $\mathbf{A} \in \mathbb{F}^{n\times n}$ be a square matrix. Then the following statements are logically equivalent:

1. The reduced row echelon form of the $n \times 2n$ matrix $[\mathbf{A}|\mathbf{I}_n]$ is of the form $[\mathbf{I}_n|\mathbf{B}]$ for some square matrix $\mathbf{B} \in \mathbb{F}^{n\times n}$.

2. There exists a square matrix $\mathbf{B} \in \mathbb{F}^{n\times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$.

Proof. Let us assume that the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{I}_n]$ is of the form $[\mathbf{I}_n|\mathbf{B}]$ for some square matrix $\mathbf{B} \in \mathbb{F}^{n\times n}$. Let us denote by $\mathbf{b}_i$ the i-th column of the matrix $\mathbf{B}$. Then using the same elementary row operations to transform the matrix $[\mathbf{A}|\mathbf{I}_n]$ into the form $[\mathbf{I}_n|\mathbf{B}]$ can be used to transform the matrix $[\mathbf{A}|\mathbf{e}_i]$ into $[\mathbf{I}_n|\mathbf{b}_i]$. Since $[\mathbf{I}_n|\mathbf{b}_i]$ is in reduced row echelon form, we can conclude that the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{e}_i]$ is equal to $[\mathbf{I}_n|\mathbf{b}_i]$. This implies that $\mathbf{b}_i$ is a solution to the system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$. But then $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$. In particular $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ for some square matrix $\mathbf{B} \in \mathbb{F}^{n\times n}$, namely the matrix occurring in the right part of the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$.

Now conversely, suppose that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ for some square matrix $\mathbf{B} \in \mathbb{F}^{n \times n}$. Then for all i from 1 to n, the system of linear equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$ has a solution, namely the i-th column of the matrix $\mathbf{B}$. We claim that the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ only contains pivots in its first n columns. We will proof this claim using a proof by contradiction. Assume therefore that the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ in fact has a pivot contained in a column with index $n+i$ for some $i > 0$. Then the reduced row echelon form of the matrix $[\mathbf{A}|\mathbf{e}_i]$ would contain a pivot in its $(n+1)$-th column. In particular, $\mathbf{A}$ and $[\mathbf{A}|\mathbf{e}_i]$ would not have the same rank. But then by Corollary 6.27, the system $\mathbf{A} \cdot \mathbf{x} = \mathbf{e}_i$ has no solution. Since we already observed that it does have a solution, we obtain a contradiction. This proves the claim that the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ only contains pivots in its first n columns. Next, we claim that the rank of $[\mathbf{A}|\mathbf{I}_n]$ is equal to n. To obtain a contradiction, suppose that the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ contains a zero row. Considering the second part of the matrix, $\mathbf{I}_n$, we can conclude that there exist a sequence of elementary row operations that can transform $\mathbf{I}_n$ into a matrix for a zero row. But $\mathbf{I}_n$ is a matrix with rank n, while an $n \times n$ matrix with a zero row can have rank at most $n-1$. This proves the second claim. Combining the two claims, we conclude that the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ contains a pivot in each of its first n columns. But then it is of the form $[\mathbf{I}_n|\mathbf{C}]$ for some square matrix $\mathbf{C} \in \mathbb{F}^{n \times n}$. □

|||| Corollary 7.23

Let $\mathbf{A} \in \mathbb{F}^{n \times n}$ be given. Then there exists $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ if and only if $\rho(\mathbf{A}) = n$.

Proof. If $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$ for some $\mathbf{B} \in \mathbb{F}^{n \times n}$, then by Theorem 7.22 the reduced row echelon form of the $n \times 2n$ matrix $[\mathbf{A}|\mathbf{I}_n]$ is of the form $[\mathbf{I}_n|\mathbf{C}]$ for some $\mathbf{C} \in \mathbb{F}^{n \times n}$. But then the reduced row echelon form of $\mathbf{A}$ itself is $\mathbf{I}_n$, implying that $\rho(\mathbf{A}) = n$.

Conversely, if $\rho(\mathbf{A}) = n$, the reduced row echelon form of $\mathbf{A}$ is equal to $\mathbf{I}_n$. But then the reduced row echelon form of $[\mathbf{A}|\mathbf{I}_n]$ is of the form $[\mathbf{I}_n|\mathbf{C}]$ for some square matrix $\mathbf{C} \in \mathbb{F}^{n \times n}$. By Theorem 7.22, we may conclude that there exists $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$. □

|||| Corollary 7.24

Let $\mathbf{A} \in \mathbb{F}^{n \times n}$ be a square matrix and suppose that there exists a square matrix $\mathbf{B} \in \mathbb{F}^{n \times n}$ such that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$. Then $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$ and therefore $\mathbf{B} = \mathbf{A}^{-1}$, the inverse of $\mathbf{A}$.

Proof. To conclude that $\mathbf{B}$ is the inverse of $\mathbf{A}$, we need to show that $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$. Since we are given that $\mathbf{A} \cdot \mathbf{B} = \mathbf{I}_n$, what is left to show, is that $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$.

Now note that $\mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{A}) = (\mathbf{A} \cdot \mathbf{B}) \cdot \mathbf{A} = \mathbf{I}_n \cdot \mathbf{A} = \mathbf{A} = \mathbf{A} \cdot \mathbf{I}_n$. Hence $\mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{A} - \mathbf{I}_n) = \mathbf{A} \cdot (\mathbf{B} \cdot \mathbf{A}) - \mathbf{A} \cdot \mathbf{I}_n = \mathbf{0}$, where here $\mathbf{0}$ denotes the $n \times n$ zero matrix.

Note that the previous equation implies that any column of $\mathbf{B} \cdot \mathbf{A} - \mathbf{I}_n$ is a solution to the homogeneous system of equations $\mathbf{A} \cdot \mathbf{x} = \mathbf{0}$. On the other hand, the previous corollary implies that the matrix $\mathbf{A}$ has rank n. Hence, we know from Corollary 6.30 that the system $\mathbf{A} \cdot \mathbf{x} = \mathbf{0}$ only has the solution $\mathbf{x} = \mathbf{0}$. Hence all columns of $\mathbf{B} \cdot \mathbf{A} - \mathbf{I}_n$ are zero, implying that $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}_n$. This is exactly what we needed to show. □

|||| Corollary 7.25

Let $\mathbf{A} \in \mathbb{F}^{n \times n}$ be given. Then its inverse matrix exists if and only if $\rho(\mathbf{A}) = n$.

Proof. This follows from the previous two corollaries. □