2 Matrices

In the previous section we considered the system of linear equations

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 1, \\ 4x_1 + x_2 &= -2, \\ -2x_1 + 2x_2 + x_3 &= 7, \end{aligned} \tag{2.1}$$

for which we found that in the process of Gaussian elimination it is convenient to put all the constants in the form of the following table

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 4 & 1 & 0 & -2 \\ -2 & 2 & 1 & 7 \end{array}\right]. \tag{2.2}$$

This table is actually much more commonly called an (*augmented*) *matrix* of system (2.1), and hence I am motivated to introduce the following definition.

Definition 2.1. *A rectangular array of numbers with m rows and n columns is called an $m \times n$ matrix (plural, matrices).*

I will usually denote matrices with bold capital letters $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \ldots$ The notation $\boldsymbol{A}_{m\times n}$ means an $m \times n$ matrix. The constants in the matrix are called the *matrix entries*, and I will denote them with small letters with two indexes, a_{ij}, where the first index refers to the i-th row and the second index refers to the j-th column, and hence the entry a_{ij} is at the intersection of row i and column j. In general I will use the square brackets to denote matrices:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \ddots & & & \vdots \\ a_{i1} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & & & \ddots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{bmatrix},$$

an equivalent and equally frequent notation is to use the usual round brackets

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \ddots & & & \vdots \\ a_{i1} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & & & \ddots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{pmatrix}.$$

I will use a short notation $\boldsymbol{A} = [a_{ij}]_{m\times n}$ to denote the same matrix.

A matrix is called *square* if the number of rows is equal to the number of columns, i.e., if $m = n$. A matrix that has only one column is called *column vector*, in this case usually one of the indexes is

Math 329: *Intermediate Linear Algebra* by Artem Novozhilov©
e-mail: artem.novozhilov@ndsu.edu. Spring 2017

omitted:

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix},$$

note that I use a small bold letter in this case to denote my column vector. A matrix that has only one row is called *row vector*:

$$\boldsymbol{b} = [b_1 \;\; \dots \;\; b_n].$$

It is much more common in mathematics to denote the row vectors as $\boldsymbol{b} = (b_1, \dots, b_n)$, but I prefer to keep this notation for *collections* or *system* of vectors.

The matrix that consists of all zeroes is called a *zero matrix* and will be denoted $\mathbf{0}$ (note that in many books, and, of course, on the blackboard, the notation is simply 0 and it is the reader who must understand what this zero actually means, whether it is a simple number zero or, say, 5 by 2 matrix, whose all entries are zeroes).

On the collection of matrices the natural operations of addition and multiplication by a scalar are defined. Here, and in what follows, I use the terms *scalar* and *constant* interchangeably and consider them to be synonyms.

Definition 2.2. *Let $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B} = [b_{ij}]$ be two $m \times n$ matrices. Then their sum, $\boldsymbol{A} + \boldsymbol{B}$, is, by definition, an $m \times n$ matrix $\boldsymbol{C} = [c_{ij}]$, such that*

$$c_{ij} = a_{ij} + b_{ij}, \quad i = 1, \dots, m,\ j = 1, \dots, n.$$

Multiplication $\alpha\boldsymbol{A}$ of matrix $\boldsymbol{A} = [a_{ij}]$ by a scalar α is, by definition, matrix $\boldsymbol{D} = [d_{ij}]$, such that

$$d_{ij} = \alpha a_{ij}$$

for all i, j.

Any time someone introduces new operations, the properties of these operations must be checked. It is a great danger to transfer our familiar properties of, say, addition of two numbers, onto some operation (which we confusingly also called addition). Hence our first proof in this lecture.

Proposition 2.3. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \mathbf{0}$ be matrices of the same size. Let α, β be two scalars. Then*

$$\begin{aligned}
&\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}, \qquad \text{(matrix addition is commutative)} \\
&(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}), \qquad \text{(matrix addition is associative)}, \\
&\boldsymbol{A} + \mathbf{0} = \boldsymbol{A}, \\
&\boldsymbol{A} + (-\boldsymbol{A}) = \mathbf{0}, \\
&\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}, \\
&(\alpha + \beta)\boldsymbol{A} = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}, \\
&\alpha(\beta\boldsymbol{A}) = (\alpha\beta)\boldsymbol{A}, \\
&1\boldsymbol{A} = \boldsymbol{A}.
\end{aligned}$$

Remark 2.4. Any mathematical proof relies on some known facts and properties that are certainly true. However, I just started talking about operations on matrices and do not have that many facts to use. In all such cases one should use the definition. I will prove only the first property and leave the rest as (almost trivial) exercises.

Proof. By definition $\boldsymbol{A}+\boldsymbol{B}$ is the matrix with the entries $a_{ij}+b_{ij}$. By the same definition, $\boldsymbol{B}+\boldsymbol{A}$ is the matrix with the entries $b_{ij}+a_{ij}$. a_{ij} and b_{ij} are constants for which I know that $a_{ij}+b_{ij}=b_{ij}+a_{ij}$. Hence I conclude that

$$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}.$$

■

Exercise 1. Prove the rest of the properties of matrix addition and multiplication by scalars.

Remark 2.5. Using the introduced notations I can write my proof more concisely as

$$\boldsymbol{A}+\boldsymbol{B}=[a_{ij}]+[b_{ij}]=[a_{ij}+b_{ij}]=[b_{ij}+a_{ij}]=[b_{ij}]+[a_{ij}]=\boldsymbol{B}+\boldsymbol{A}.$$

A more complicated matrix operation is the *multiplication* of two matrices. We will see the actual reason for such definition slightly later in the course. First, I will define the product of an $1\times p$ row vector $\boldsymbol{a}=[a_i]_{1\times p}$ by an $p\times 1$ column vector $\boldsymbol{b}=[b_i]_{p\times 1}$:

$$\boldsymbol{ab}=[a_1\ \dots\ a_p]\begin{bmatrix}b_1\\ \vdots\\ b_p\end{bmatrix}=a_1b_1+\ldots+a_pb_p=\sum_{l=1}^{p}a_lb_l=\boldsymbol{a}\cdot\boldsymbol{b}.$$

You should recognize the familiar *dot product* of two vectors, which is quite often denoted as $\boldsymbol{a}\cdot\boldsymbol{b}$.

Now, let me take two matrices, $\boldsymbol{A}_{m\times p}$ and $\boldsymbol{B}_{p\times n}$, that is, the number of columns of the first one is equal to the number of rows of the second one; I will also use the notations $\boldsymbol{r}_1,\ldots,\boldsymbol{r}_m$ for the *rows* of the matrix $\boldsymbol{A}$ and $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_n$ for the columns of $\boldsymbol{B}$. Below the vertical lines between $\boldsymbol{c}_j$ emphasize that they are columns. Then the product $\boldsymbol{AB}$ is *defined* as

$$\boldsymbol{AB}=\begin{bmatrix}\boldsymbol{r}_1\\ \vdots\\ \boldsymbol{r}_n\end{bmatrix}[\boldsymbol{c}_1|\ldots|\boldsymbol{c}_n]=\begin{bmatrix}\boldsymbol{r}_1\cdot\boldsymbol{c}_1 & \ldots & \boldsymbol{r}_1\cdot\boldsymbol{c}_n\\ \vdots & \ddots & \vdots\\ \boldsymbol{r}_m\cdot\boldsymbol{c}_1 & \ldots & \boldsymbol{r}_m\cdot\boldsymbol{c}_n\end{bmatrix},$$

i.e., as an $m\times n$ matrix $\boldsymbol{C}=[c_{ij}]$, where each entry is the dot product of the i-th row of the first matrix and the j-th column of the second one. Note that all the dot products are defined since all the rows and columns have the same number (namely, p) of elements. Now I can reword this definition in terms of entries of $\boldsymbol{A}$ and $\boldsymbol{B}$:

Definition 2.6. *Let $\boldsymbol{A}_{m\times p}$ and $\boldsymbol{B}_{p\times n}$ be two matrices. Their product, $\boldsymbol{AB}$, is an $m\times n$ matrix $\boldsymbol{C}=[c_{ij}]_{m\times n}$, such that*

$$c_{ij}=\sum_{l=1}^{p}a_{il}b_{lj}.$$

Example 2.7.

$$\begin{bmatrix}2 & 3\\ 4 & 0\end{bmatrix}\begin{bmatrix}1 & 2 & 0\\ 5 & -1 & 0\end{bmatrix}=\begin{bmatrix}17 & 1 & 0\\ 4 & 8 & 0\end{bmatrix}.$$

I already mentioned that the actual motivation for such (rather strange, you should admit) definition of matrix multiplication, will be given later in the course. The utility of it, however, can be immediately seen from, for example, a shorthand notation for the systems of linear equations. In general, a system of m linear equations with n unknowns has the form

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_n, \\ &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m. \end{aligned}$$

If I introduce the notations

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

for the matrix of coefficients, the vector of unknowns, and the vector of the right hand side respectively, then, using the matrix multiplication, this system can be written (check this) as

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}.$$

From the definition it should be already clear that even if $\boldsymbol{AB}$ is defined, $\boldsymbol{BA}$ may not exist (give an example of such matrices). $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are always defined if both $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of the same size. In this case, however, we have that $\boldsymbol{AB} \neq \boldsymbol{BA}$, in words, matrix multiplication is *not commutative.* Let me show that this is indeed true.

Proposition 2.8. *Let $\boldsymbol{A}, \boldsymbol{B}$ be two $n \times n$ matrices. Then $\boldsymbol{AB} \neq \boldsymbol{BA}$ in general.*

Remark 2.9. How to prove such a statement? If you were asked to show that $\boldsymbol{AB} = \boldsymbol{BA}$, this would mean that this equality holds for *all* possible matrices. Therefore, it would be reasonable to start with something like "Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two arbitrary square matrices of the same size..." and then work on, e.g., the definition of matrix multiplication. You are asked, however, to show that the equality does *not* hold in general. That in, it *may* be true for some matrices, but... In other, more precise, words, you are asked to prove that *there exist two such matrices* that... Maybe you can just simply take two random matrices and check?

Proof. I will leave it as an exercise to come up with a single example that shows that $\boldsymbol{AB} \neq \boldsymbol{BA}$. ■

Are there matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA}$? Actually, yes, plenty of them. The simplest such example is probably provided by an *identity matrix.* Let me start with a definition of a diagonal matrix. Let $\boldsymbol{A} = [a_{ij}]$ be a square matrix, the entries a_{ii} are said to be on the main diagonal. If the only nonzero elements are on the main diagonal then we call such matrix *diagonal.* (A common notation is $\boldsymbol{A} = \text{diag}(a_{11}, \ldots, a_{nn})$).

A diagonal matrix with all entries on the main diagonal equal to 1 is called the *identity matrix*:

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}.$$

Identity matrix is usually denoted $\boldsymbol{I}_n$ or simply $\boldsymbol{I}$ if the size of the matrix is clear from the context. Note that here I for the first time used the convention that the entries, which I do not explicitly display, are all zero.

Proposition 2.10. *Let $\boldsymbol{A} = [a_{ij}]_{n\times n}$. Then*

$$\boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}.$$

Proof. Let me prove that $\boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$. By the definition of matrix multiplication the i, j entry of $\boldsymbol{I}_n\boldsymbol{A}$ is the dot product of the i-th row of $\boldsymbol{I}_n$ and the j-th column of $\boldsymbol{A}$. Since the only nonzero element of the i-th row of $\boldsymbol{I}_n$ is at the i-th position and equal to 1, hence this dot product is equal to a_{ij}, which completes the proof that $\boldsymbol{I}_n\boldsymbol{A} = \boldsymbol{A}$.

I will leave the proof of the second equality as an (almost trivial) exercise. ■

While the commutativity of matrix multiplication clearly does not hold, there are a lot of identities, similar to the ones that are true for familiar numbers, that are still satisfied for matrices.

Proposition 2.11. *Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be the matrices such that all the operation below make sense and let α be a scalar. Then*

$$\begin{aligned}
\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) &= \boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C},\\
(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} &= \boldsymbol{A}\boldsymbol{C}+\boldsymbol{B}\boldsymbol{C},\\
(\boldsymbol{A}\boldsymbol{B})\boldsymbol{C} &= \boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}),\\
\alpha(\boldsymbol{A}\boldsymbol{B}) &= (\alpha\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(\alpha\boldsymbol{B}).
\end{aligned}$$

Remark 2.12. The most important property here is the third one, which says that the matrix multiplication is associative. To prove these identities I just need to carefully use the definition of the corresponding operations. I will prove only the first one.

Proof. Let $\boldsymbol{A} = [a_{ij}]_{m\times p}, \boldsymbol{B} = [b_{ij}]_{p\times n}, \boldsymbol{C} = [c_{ij}]_{p\times n}$, where I chose the dimensions such that all the operations in the first identity are defined. For the left hand side I have

$$\boldsymbol{B}+\boldsymbol{C} = [b_{ij}]_{p\times n} + [c_{ij}]_{p\times n} = [b_{ij}+c_{ij}]_{p\times n}$$

by the definition of matrix addition. Further, by the definition of matrix multiplication,

$$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \left[\sum_{l=1}^{p} a_{il}(b_{lj}+c_{lj})\right]_{m\times n} = \left[\sum_{l=1}^{p} a_{il}b_{lj}\right]_{m\times n} + \left[\sum_{l=1}^{p} a_{il}c_{lj}\right]_{m\times n} = \boldsymbol{A}\boldsymbol{B}+\boldsymbol{A}\boldsymbol{C},$$

where the final two equalities follow from the definition of matrix addition. ■

Exercise 2. Prove the rest of the properties of matrix multiplication in Proposition 2.11.

Remark 2.13. A similar proof of associativity of the matrix multiplication requires knowledge how to manipulate expression containing more than one symbol $\sum$. I invite the students to fill in the details for this proof, but remark that at some point, after introducing necessary definitions and correct context, we will see that this property becomes almost obvious.

Finally, let me introduce an *inverse* matrix.

Definition 2.14. *Let $\boldsymbol{A}$ be a square matrix. If there is a matrix $\boldsymbol{B}$ such that*

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I},$$

then $\boldsymbol{B}$ is called an inverse of $\boldsymbol{A}$ and denoted $\boldsymbol{A}^{-1}$. A matrix $\boldsymbol{A}$ that has an inverse is called an invertible matrix.

Every time a new definition is introduced it is important to make sure that it makes sense, i.e., it actually defines something which exists. Here is an example.

Example 2.15. Matrix

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

is invertible. Its inverse is

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

which can be checked directly.

Note that the definition does not say anything about uniqueness of an inverse matrix. Actually, if it exists it is unique. Here is a simple proof that uses already introduced properties of matrices.

Proposition 2.16. *The inverse, if it exists, is unique.*

Remark 2.17. How to prove that something is unique? Actually a very common trick that works quite often is to assume initially that there are two such objects and then show that they coincide. Here is our first example of this kind.

Proof. Let $\boldsymbol{B}$ and $\boldsymbol{B}'$ be two different inverses for $\boldsymbol{A}$. Then, by the properties of matrix multiplication and definition of the inverse,

$$\boldsymbol{B} = \boldsymbol{BI} = \boldsymbol{B}(\boldsymbol{AB}') = (\boldsymbol{BA})\boldsymbol{B}' = \boldsymbol{IB}' = \boldsymbol{B}'.$$

■

Remark 2.18. The proof of the previous proposition goes without any words. However, it should be clear which properties I used. To be (excessively) detailed: First I use that multiplication by the identity matrix does not change anything, then I use the definition of inverse, then associativity of matrix multiplication, then once again the definition of the inverse, and finally once more, the fact that multiplication by the identity matrix does not change the result. Make sure that in any proof you read or write you are capable to properly justify every small step.

Remark 2.19. To give you a little headache, I note that the definition requires that *both* $\boldsymbol{AB} = \boldsymbol{I}$ and $\boldsymbol{BA} = \boldsymbol{I}$ be satisfied for the inverse matrix. Since matrix multiplication is not commutative, at this point we cannot conclude that if $\boldsymbol{AB} = \boldsymbol{I}$ then $\boldsymbol{BA} = \boldsymbol{I}$. Is it actually possible to have such a matrix $\boldsymbol{B}$ for which $\boldsymbol{AB} = \boldsymbol{I}$ but $\boldsymbol{BA} \neq \boldsymbol{I}$?

Proposition 2.20. *Let $\boldsymbol{A}, \boldsymbol{B}$ be two invertible square matrices of the same size. Then their product is invertible and has the inverse $\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.*

Proof. We are given that there are $\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I},$$

and a similar statement is true for $\boldsymbol{B}$. We need to show that if $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ then $\boldsymbol{C}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$. This follows from the associativity of matrix multiplication:

$$(\boldsymbol{A}\boldsymbol{B})(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{A}(\boldsymbol{B}\boldsymbol{B}^{-1})\boldsymbol{A}^{-1} = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A}^{-1} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I},$$

and the student is invited to check the second necessary equality. ■

Remark 2.21. Very often proofs for some important claims have, as a side product, immediate and interesting consequences. They are usually called *corollaries*. Here is my first example of a corollary in this course. This is, of course, a corollary of the previous proposition.

Corollary 2.22. *Let $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_p$ be invertible square matrices of the same size. Then $\boldsymbol{A}_1 \ldots \boldsymbol{A}_p$ is invertible, with the inverse*

$$\boldsymbol{A}_p^{-1}\boldsymbol{A}_{p-1}^{-1} \ldots \boldsymbol{A}_2^{-1}\boldsymbol{A}_1^{-1}.$$

I do hope that the student can provide a convincing arguments to prove this corollary.

To finish this section, I record one more fact about invertible matrices that I will use several times in what follows.

Proposition 2.23. *If a matrix has a row of zeroes it cannot have an inverse.*

Exercise 3. Prove this proposition.