Chapter 1
Matrices

**Definition 1.1.** An \( n \times k \) matrix over a field \( F \) is a function \( A : \{1, 2, \ldots, n\} \times \{1, 2, \ldots, k\} \rightarrow F. \)

A matrix is usually represented by (and identified with) an \( n \times k \) (read as “\( n \) by \( k \)”) array of elements of the field (usually numbers). The horizontal lines of a matrix are referred to as **rows** and the vertical ones as **columns**. The individual elements are called **entries** of the matrix. Thus an \( n \times k \) matrix has \( n \) rows, \( k \) columns and \( nk \) entries. Matrices will be denoted by capital letters and their entries by the corresponding small letters. Thus, in case of a matrix \( A \) we will write \( A(i,j)=a_{i,j} \) and will refer to \( a_{i,j} \) as the element of the \( i \)-th row and \( j \)-th column of \( A \). On the other hand we will use the symbol \([a_{i,j}]\) to denote the matrix \( A \) with entries \( a_{i,j} \).

**Definition 1.2.** Let \( A \) be an \( n \times k \) matrix. Then the **transpose** of \( A \) is the \( k \times n \) matrix \( A^T \), such that for each \( i \) and \( j \), \( A^T(i,j)=A(j,i) \).

In other words, \( A^T \) is what we get when we replace rows of \( A \) with columns and vice versa.

**Proposition 1.1.** For any \( n, k \in \mathbb{N} \) the set \( Mtr(n,k) \) consisting of all \( n \times k \) matrices over \( F \) with ordinary function addition and multiplication of functions by constants is a vector space over \( F \), \( \dim(Mtr(n,k))=nk. \)

**Proof.** \( Mtr(n,k) \) is a vector space (see Example 4.4??). As a basis for \( Mtr(n,k) \) we can use the set consisting of matrices \( A_{p,q} \), where \( A_{p,q}(i, j) = \begin{cases} 1 & \text{if } (i, j) = (p, q) \\ 0 & \text{otherwise} \end{cases} \)

The operation of matrix multiplication is something quite different from the abovementioned operations of matrix scaling and addition. It is not inherited from function theory, it is a specifically matrix operation. The reason for this will soon become clear.
**Definition 1.3.** Suppose A is an \(n \times k\) matrix and B is a \(k \times s\) matrix. Then the **product** of A by B is the matrix C, where \(c_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \ldots + a_{i,k}b_{k,j} = \sum_{t=1}^{k} a_{i,t}b_{t,j}\)

In other words, \(c_{i,j}\) is the sum of the products of consecutive elements of the \(i\)-th row of A by the corresponding elements of the \(j\)-th column of B. Strictly speaking multiplication of matrices in general is not an operation on matrices as it maps \(\text{Mtr}(n,k) \times \text{Mtr}(k,s)\) into \(\text{Mtr}(n,s)\). It is an operation though, when \(n=k=s\). Matrices with the same number of rows and columns are called **square** matrices.

Notice that \(C=AB\) is an \(n \times s\) matrix. Notice also that for the definition to work, the number of columns in A (the first factor) must equal the number of rows in B (the second factor). In other cases the product is not defined. This suggest that matrix multiplication need not be commutative.

**Example 1.1.** Let \(A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & -3 \end{bmatrix}\) and \(B = \begin{bmatrix} 2 & 0 & 2 \\ -1 & 3 & 1 \\ 1 & -2 & 2 \end{bmatrix}\). Then
\[
AB = \begin{bmatrix} 2 + 1 + 2 & 0 - 3 - 4 & 2 - 1 + 4 \\ 4 + 0 - 3 & 0 + 0 + 6 & 4 + 0 - 6 \end{bmatrix} = \begin{bmatrix} 5 & -7 & 5 \\ 1 & 6 & -2 \end{bmatrix}
\]

and \(BA\) is not defined since the number of columns of B is not equal to the number of rows of A.

**Example 1.2.** Let \(A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\) and \(B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\). Then \(AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\) while \(BA = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}\). This example proves that \(AB\) may differ from \(BA\) even when both products exist and have the same size.

**Example 1.3.** The \(n \times n\) matrix I defined as \(I(s,t) = \begin{cases} \text{if } s = t & 1 \\ \text{if } s \neq t & 0 \end{cases}\) is the identity element for matrix multiplication on \(\text{Mtr}(n,n)\). Indeed, for every matrix A, \(AI(i,j) = \sum_{t=1}^{n} A(i,t)I(t,j) = A(i,j)\) because \(A(i,t)I(t,j)\) is equal to \(A(i,j)\) when \(t=j\) and is equal to 0 otherwise. Hence \(AI=A\). A similar argument proves that \(IA=A\). All entries of the identity matrix I are equal to 0 except for the **diagonal entries** which are all equal to 1. The term **diagonal entries** of a matrix A refers to the diagonal of a square \(n \times n\) matrix, i.e. the line connecting the top-left with the
bottom-right corner of the matrix. The line consists of all elements of the form \(A(i,i)\), \(i=1,2,\ldots,n\).

**Theorem 1.1.** For every \(n \times k\) matrix \(A\) and for every two \(k \times s\) matrices \(B\) and \(C\),
\[A(B+C) = AB + AC,\]
i.e. matrix multiplication is distributive with respect to matrix addition.

**Proof.**
\[
A(B+C)[i,j] = \sum_{p=1}^{k} A[i, p](B + C)[p, j] = \sum_{p=1}^{k} A[i, p][B[p, j] + C[p, j]] = \\
\sum_{p=1}^{k} A[i, p]B[p, j] + \sum_{p=1}^{k} A[i, p]C[p, j]) = (AB+AC)[i,j]. \square
\]

**Definition 1.4.** Let \(A\) be an \(n \times k\) matrix. We say that \(A\) is a row echelon matrix iff
(a) if \(r_i\) is a nonzero row of \(A\) then \(r_{i-1}\) is also a nonzero row, \(i=2,3,\ldots\)
(b) if \(a_{i,j}\) is the first nonzero entry in \(r_i\) and \(a_{i-1,p}\) is the first nonzero entry in \(r_{i-1}\) then \(p<j\)
If, in addition, 
(c) the first nonzero entry in each nonzero row is equal to 1
(d) the first nonzero entry in each nonzero row is the only nonzero entry in its column
then \(A\) is called a row canonical matrix.

**Definition 1.5.** The following transformations of a matrix are called elementary row operations
(a) \(r_i \leftrightarrow r_j\) - replacing row \(r_i\) with \(r_j\) and vice versa (row swapping)
(b) \(r_i \leftarrow cr_i\) - replacing row \(r_i\) with its multiple by a nonzero constant \(c\) (scaling of a row).

In practice we abbreviate the symbol to \(cr_i\)
(c) \(r_i \leftarrow r_i + r_j\) - replacing row \(r_i\) with the sum of \(r_i\) and \(r_j\) (adding of \(r_j\) to \(r_i\)). We usually write simply \(r_i+r_j\).
(d) \(r_i \leftarrow r_i + cr_j\) - replacing row \(r_i\) with the sum of \(r_i\) and a multiple of \(r_j\) by a constant \(c\).

Normally we just write \(r_i+cr_j\).

Notice that the operation (d) is a composition of (b) and (c). Namely, we can do \(cr_j\), then \(r_i+r_j\) (here \(r_j\) denotes the “new” row \(j\), after scaling) and finally \(c^{-1}r_j\) to return to the original row \(j\).
**Definition 1.6.** Matrices A and B are said to be *row-equivalent* iff A can be transformed into B by a sequence of elementary row operations. We denote row-equivalence by \( A \sim B \).

**Definition 1.7.** The *row rank* of a matrix A, \( r(A) \), is the dimension of the subspace of \( \mathbb{F}^k \) spanned by rows of A.

**Theorem 1.2.** For every two matrices A and B, if \( A \sim B \) then \( r(A) = r(B) \).

**Proof.** We prove this by showing that each elementary row operation preserves the very space spanned by rows of the matrix, hence they also preserve its dimension. \( \square \)

**Theorem 1.3.** For every matrix A, \( r(A) = r(A^T) \).

**Proof.** Skipped

Definitions 1.4 – 1.7 and Theorem 1.2 could just as well be phrased in terms of columns rather than rows, leading to the concept of the *column rank* of a matrix. However, Theorem 1.3 states that the two are the same.

Since the rank of any row echelon matrix is clearly the number of its nonzero rows, Theorem 1.2 provides a strategy for calculating the rank of a matrix - row reduce the matrix to a row echelon matrix and then count the nonzero rows.

**Definition 1.8.** The *determinant* is a function which assigns a number (an element of the underlying field) to every square (i.e. \( n \times n \)) matrix A. The function is defined inductively with respect to n:

1. if \( n=1 \) then \( \det(A) = A[1,1] \)

2. if \( n>1 \) then \( \det(A) = \sum_{i=1}^{n} (-1)^{i+1} A[i,1] \det(A_{i,1}) \), where \( A_{p,s} \) is the \( n-1 \times n-1 \) matrix obtained from A by the removal of p-th row and s-th column.

The sum appearing in part two of the definition is known as the Laplace expansion of the determinant with respect to the first column.