Course 7

Chapter 3 MATRICES AND LINEAR SYSTEMS

In this chapter we present the essential connection between linear maps and matrices, developing a more computational apparatus. Using elementary operations, we are able to give practical methods for computing the rank or the inverse of a matrix as well as for solving linear systems of equations. We also study eigenvalues and eigenvectors, which among many applications, offer tools for computing powers of a matrix.

Throughout the present chapter K will always denote a field, and $m, n \in \mathbb{N}^*$.

3.1 Elementary operations

Definition 3.1.1 By an *elementary operation* on a list of vectors in a vector space we understand one of the following three processes:
(1) To interchange any two vectors of the list.
(2) To multiply a vector of the list by a non-zero scalar.
(3) To multiply a vector of the list by a scalar and add the result to another vector of the list.

Let us rewrite the things more formally.

Definition 3.1.2 Let V be a vector space over K. Then an *elementary operation* is one of the functions ε_{ij}, $\varepsilon_{i\alpha}$, $\varepsilon_{ij\alpha} : V^n \to V^n$ defined for every $(v_1, \ldots, v_n) \in V^n$ by:

$$\varepsilon_{ij}(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = (v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n), \quad (1)$$

$$\varepsilon_{i\alpha}(v_1, \ldots, v_i, \ldots, v_n) = (v_1, \ldots, \alpha v_i, \ldots, v_n), \quad \alpha \in K^*, \quad (2)$$

$$\varepsilon_{ij\alpha}(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = (v_1, \ldots, v_i + \alpha v_j, \ldots, v_j, \ldots, v_n), \quad \alpha \in K. \quad (3)$$

Theorem 3.1.3 *Using the previous notation, we have* $\varepsilon_{ij}, \varepsilon_{i\alpha}, \varepsilon_{ij\alpha} \in \mathrm{Aut}_K(V^n)$.

Proof. It is easy to show that V^n has a structure of a vector space over K, where the operations are defined by

$$(v_1, \ldots, v_n) + (v_1', \ldots, v_n') = (v_1 + v_1', \ldots, v_n + v_n'),$$
$$k(v_1, \ldots, v_n) = (kv_1, \ldots, kv_n),$$

$\forall k \in K$ and $\forall (v_1, \ldots, v_n), (v_1', \ldots, v_n') \in V^n$. Also, it is easy to check that ε_{ij}, $\varepsilon_{i\alpha}, \varepsilon_{ij\alpha}$ are K-linear maps. They are also bijections, having the inverses

$$(\varepsilon_{ij})^{-1} = \varepsilon_{ji}, \quad (\varepsilon_{i\alpha})^{-1} = \varepsilon_{i\alpha^{-1}}, \quad (\varepsilon_{ij\alpha})^{-1} = \varepsilon_{ij(-\alpha)},$$

which imply that they are automorphisms. □

Definition 3.1.4 Let V be a vector space over K. Then two lists X and X' of vectors in the vector space V^n over K are called *equivalent* if one of them can be obtained from the other by applying a finite number of elementary operations, that is, there exists a finite composition $\varphi : V^n \to V^n$ of elementary operations such that

$$\varphi(X) = X' \quad \text{or} \quad \varphi(X') = X.$$

Remark 3.1.5 (1) The composition $\varphi : V^n \to V^n$ of elementary operations is obviously bijective, hence by Theorem 3.1.3, if one of the lists can be obtained from the other by applying a finite number of elementary operations, then the other one also can.

(2) The name "equivalent" is justified by the fact that the previously defined relation is an equivalence relation (reflexive, transitive and symmetric).

Theorem 3.1.6 *Let V be a vector space over K and let X and X' be equivalent lists of vectors in the vector space V^n over K. Then:*
(i) X is linearly independent in $V^n \iff X'$ is linearly independent in V^n.
(ii) X is a system of generators for $V^n \iff X'$ is a system of generators for V^n.
(iii) X is a basis of $V^n \iff X'$ is a basis of V^n.

Proof. Since the lists X and X' are equivalent, there exists a finite composition $\varphi : V^n \to V^n$ of elementary operations such that $\varphi(X) = X'$. Since by Theorem 3.1.3, each elementary operation is an isomorphism, it follows that φ is an isomorphism. But we know that isomorphisms preserve linearly independent lists, systems of generators, and bases. □

Remark 3.1.7 In what follows, let us apply the previously presented theory of elementary operations in the case of the vector space $M_{m,n}(K)$ of $m \times n$-matrices over K. In order to do that, we will see a matrix $A = (a_{ij}) \in M_{m,n}(K)$ as a list of vectors $(a^1, \ldots, a^n)$, each of them being a column

$$a^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Then we get the following well-known elementary operations for a matrix:

(1) Interchanging any two columns of the matrix.

(2) Multiplying a column of the matrix by a non-zero scalar.

(3) Multiplying a column of the matrix by a scalar and add the result to another column of the matrix.

Applying Definition 3.1.4 in this case, we say that two matrices are *equivalent* if one of them can be obtained from the other by a finite number of the previous elementary operations on columns.

Theorem 3.1.8 *The value of an elementary operation applied on a matrix $A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of column-vectors $(a^1, \ldots, a^n)$, is equal to A multiplied on the right hand side by the matrix obtained from the identity matrix I_n, also seen as a list of columns, by applying the same elementary operation.*

Proof. For simplicity of writing we are going to prove the theorem for the first two columns involved in the elementary operations. We have

$$\varepsilon_{12}(A) = \varepsilon_{12}(a^1, a^2, a^3, \ldots, a^n) = (a^2, a^1, a^3, \ldots, a^n) = \begin{pmatrix} a_{12} & a_{11} & a_{13} & \ldots & a_{1n} \\ a_{22} & a_{21} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m2} & a_{m1} & a_{m3} & \ldots & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} = A \cdot E_{12},$$

where E_{12} is the matrix obtained from the identity matrix I_n by interchanging the first two columns.

Furthermore, $\forall \alpha \in K^*$,

$$\varepsilon_{1\alpha}(A) = \varepsilon_{1\alpha}(a^1, a^2, \ldots, a^n) = (\alpha a^1, a^2, \ldots, a^n) = \begin{pmatrix} \alpha a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \alpha a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} \alpha & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & & \vdots & \ldots \\ 0 & 0 & \ldots & 1 \end{pmatrix} = A \cdot E_{1\alpha},$$

where $E_{1\alpha}$ is the matrix obtained from the identity matrix I_n by multiplying the first column by α.

Finally, $\forall \alpha \in K$,

$$\begin{aligned}\varepsilon_{12\alpha}(A) &= \varepsilon_{12\alpha}(a^1, a^2, a^3, \ldots, a^n) = (a^1 + \alpha a^2, a^2, a^3, \ldots, a^n) \\ &= \begin{pmatrix} a_{11} + \alpha a_{12} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} + \alpha a_{22} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} + \alpha a_{m2} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ \alpha & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} = A \cdot E_{12\alpha}\,,\end{aligned}$$

where $E_{12\alpha}$ is the matrix obtained from the identity matrix I_n by multiplying the second column by α and adding it to the first column. □

Definition 3.1.9 The matrices of the form E_{ij}, $E_{i\alpha}$, $E_{ij\alpha}$, obtained by applying elementary operations on the identity matrix I_n, are called *elementary matrices*.

Theorem 3.1.10 *The elementary matrices are invertible.*

Proof. We have $\det I_n = 1 \neq 0$. But the determinant remains non-zero by applying elementary operations on I_n. Hence the elementary matrices are invertible. □

Remark 3.1.11 (1) It is also possible to see a matrix $A = (a_{ij}) \in M_{m,n}(K)$ as a list of vectors $(a_1, \ldots, a_m)$, each of them being a row $a_i = (a_{i1} \ldots a_{in})$. In this case, the elementary operations are made on rows and the value of an elementary operation applied on A is equal to A multiplied on the left hand side by the matrix obtained from the identity matrix I_m by applying the same elementary operation.

(2) By the reason of methods of computing the rank of a matrix and solving linear systems of equations, from now on we will apply the elementary operations on the rows of a matrix.

Now we would like to find for a given matrix a "better" equivalent matrix. This better form is described in the following definition.

Definition 3.1.12 We say that a matrix $A \in M_{m,n}(K)$ is in *(row) echelon form* with $r \geq 1$ non-zero rows if:
(1) the rows $1, \ldots, r$ are non-zero and the rows $r+1, \ldots, m$ are zero;
(2)

$$0 \leq N(1) < N(2) < \cdots < N(r)\,,$$

where $N(i)$ denotes the number of zero elements from the beginning of the row i, $\forall i \in \{1, \ldots, r\}$.

Theorem 3.1.13 *Every non-zero matrix $A \in M_{m,n}(K)$ is equivalent to a matrix in echelon form.*

Proof. As we have mentioned, we are going to use elementary operations on the rows of a matrix. Let $A = (a_{ij}) \in M_{m,n}(K)$. Look for a row i with the least $N(i)$, that is, the least number of zero elements from the beginning of the row. Then interchange it with the first row. Let $j_1 = N(1) + 1$, that is, a_{1j_1} is the first non-zero element of the first row. Such an element is sometimes called a *pivot*. Then make zeros on the column j_1 below the element a_{1j_1} by applying elementary operations on rows. In order to do that, multiply the first row by $-a_{kj_1}a_{1j_1}^{-1}$ and add it to the row k, $\forall k \in \{2, \ldots, m\}$. Now look for a row i with the least $N(i)$, where $i \in \{2, \ldots, n\}$. Then interchange it with the second row. Let $j_2 = N(2) + 1$,

that is, a_{2j_2} is the first non-zero element of the second row. Then make zeros on the column j_2 below the element a_{1j_2} by applying elementary operations on rows. Repeating this procedure, we get a matrix in echelon form. □

Example 3.1.14 Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \in M_{3,4}(\mathbb{R}) .$$

Then by applying elementary operations only on rows, we have the following succession of equivalent matrices (we denote by $\sim$ their equivalence):

$$A = \begin{pmatrix} \boxed{1} & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & \boxed{-1} & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix} .$$

We have pointed out the pivots by putting them in a box. In the first step, we have multiplied the first row by -3 and then by 1 and add it to the second row and to the third row respectively. In the second step, we have multiplied the second row by 2 and add it to the third row. The last matrix is in (trapezoidal) echelon form and it is equivalent to A.

3.2 Applications of elementary operations

Let us now see how to use elementary operations to compute easier the rank of a matrix and the inverse of a square matrix. Recall that we are going to apply elementary operations on the rows of a matrix.

Lemma 3.2.1 *Let $A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of row-vectors $(a_1, \dots, a_m)$. Then a sublist of $(a_1, \dots, a_m)$ consisting of r vectors is linearly independent in K^n if and only if there exists a non-zero minor of order r of the matrix A.*

Proof. For the sake of simplicity, let us consider the sublist $X = (a_1, \dots, a_r)$ of $(a_1, \dots, a_m)$. Then X is linearly independent if and only if

$$k_1a_1 + \dots + k_ra_r = 0 \Longrightarrow k_1 = \dots = k_r = 0 \quad (k_1, \dots, k_r \in K) .$$

But this is equivalent to the fact that the linear homogeneous system of equations

$$\begin{cases} k_1a_{11} + k_2a_{21} + \dots + k_ra_{r1} = 0 \\ k_1a_{12} + k_2a_{22} + \dots + k_ra_{r2} = 0 \\ \dots\dots\dots\dots\dots \\ k_1a_{1n} + k_2a_{2n} + \dots + k_ra_{rn} = 0 . \end{cases}$$

has only the zero solution. And this happens if and only if there exists a non-zero minor of order r of A (see a forthcoming theorem). □

Theorem 3.2.2 *Let $0 \neq A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of row-vectors $(a_1, \dots, a_m)$ or as a list of column-vectors $(a^1, \dots, a^n)$. Then:*
(i) $\text{rank}(A) = \dim < a_1, \dots, a_m >= \dim < a^1, \dots, a^n >$.
(ii) $\text{rank}(A) = \text{rank}(C) = r$, *where C is a matrix in echelon form with r non-zero rows equivalent to A.*

Proof. (i) Denote $r = \dim < a_1, \dots, a_m >$. Then the maximum number of linearly independent vectors in the list $(a_1, \dots, a_m)$ is r. Then by Lemma 3.2.1, the maximum order of the non-zero minors of A is r. Therefore, $\text{rank}(A) = r$.

Similarly, $\text{rank}(A) = \dim < a^1, \dots, a^n >$.

(ii) Let us see the matrices A and C as lists of row-vectors $(a_1, \ldots, a_m)$ and $(c_1, \ldots, c_r, c_{r+1}, \ldots, c_m)$. Clearly, $c_{r+1} = \cdots = c_m = 0$ and the list $(c_1, \ldots, c_r)$ is linearly independent. By Theorems 3.1.6 and 3.2.2, the elementary operations preserve the rank of a matrix. It follows that

$$\text{rank}(A) = \dim < a_1, \ldots, a_m >= \dim < c_1, \ldots, c_m >= \dim < c_1, \ldots, c_r >= r\,.$$

Clearly, $\text{rank}(C) = r$. □

Example 3.2.3 Consider the matrix

$$A = \begin{pmatrix} -3 & 5 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -3 \end{pmatrix} \in M_{3,4}(\mathbb{R})\,.$$

Then

$$A \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ -1 & 1 & 0 & 1 \\ -3 & 5 & -1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & 2 & -1 & -2 \\ 0 & 8 & -4 & -8 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}\,,$$

hence $\text{rank}(A) = 2$.

Theorem 3.2.4 *Let $A \in M_n(K)$ with $\det(A) \neq 0$. Then A is equivalent to the identity matrix I_n and the inverse matrix A^{-1} of A is obtained from the identity matrix I_n by applying the same elementary operations as one does to obtain I_n from A.*

Proof. Since $\det(A) \neq 0$, $A \in M_n(K)$ is invertible, hence we have $\text{rank}(A) = n$. Then by applying elementary operations we get to a matrix C in echelon form, having n non-zero rows. The matrix C has all the elements below the principal diagonal zero. Then one can make zeros above the principal diagonal, by applying elementary operations on rows starting with the last row. Thus, A is equivalent to a matrix in diagonal form. But since its rank is n, all the elements on the diagonal are non-zero, hence we may multiply by their inverses to get I_n. Therefore, A is equivalent to I_n.

But by Theorem 3.1.8 this means that there exist some elementary matrices $E_1, \ldots, E_k$ such that $E_k \ldots E_1 A = I_n$. It follows that

$$A^{-1} = E_k \ldots E_1 I_n\,,$$

that is, A^{-1} is obtained from the identity matrix I_n by applying the same elementary operations as one does to obtain I_n from A. □

Example 3.2.5 Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R})\,.$$

Then $\det(A) = 1 \neq 0$, hence A is invertible. Let us determine its inverse with the above described method. For simplicity of writing, we will put in the same matrix both A and the identity matrix I_3 and we will apply in parallel elementary operations on rows. We have

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right)\,.$$

We read the inverse matrix A^{-1} from the right hand half of the last matrix, hence we have

$$A^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Now consider the same matrix with entries in $\mathbb{Z}_2$, that is,

$$A = \begin{pmatrix} \widehat{0} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{0} & \widehat{1} \end{pmatrix} \in M_3(\mathbb{Z}_2).$$

The above computations, considered now in $\mathbb{Z}_2$ (where $\widehat{1} + \widehat{1} = \widehat{0}$, hence $\widehat{-1} = \widehat{1}$), give

$$A^{-1} = \begin{pmatrix} \widehat{1} & \widehat{1} & \widehat{0} \\ \widehat{0} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{1} & \widehat{1} \end{pmatrix}.$$

Extra: LU decomposition

We present a matrix decomposition which offers more efficient ways for computing the inverse or the determinant of a matrix or for solving square linear systems of equations.

Definition 3.2.6 A matrix $A \in M_n(K)$ has an *LU decomposition* if it may be written as $A = L \cdot U$ for some lower triangular matrix L (that is, a matrix all of whose elements above its principal diagonal are zero) and upper triangular matrix U (that is, a matrix all of whose elements under its principal diagonal are zero).

Theorem 3.2.7 *If $A \in M_n(K)$ can be reduced to an echelon form without interchanging any rows, then A has an LU decomposition, not necessarily unique. More generally, for every $A \in M_n(K)$ there is a permutation matrix P (that is, a matrix obtained by repeatedly interchanging the rows and columns of an identity matrix) such that $P \cdot A$ has an LU decomposition.*

Remark 3.2.8 If A has an LU decomposition, then $\det(A) = \det(L) \cdot \det(U)$. Moreover, if A is also invertible, then $A^{-1} = U^{-1} \cdot L^{-1}$. The determinants and the inverses of L and U are computed much easier than the determinant and the inverse of A.

Example 3.2.9 We have already seen in Example 3.2.5 that the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R})$$

is invertible and $\det(A) = 1$. Alternatively, note that there is the permutation matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{R})$$

(corresponding to interchanging the first two rows) such that

$$P \cdot A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} = L \cdot U$$

as a product of a lower triangular matrix L and an upper triangular matrix U. Then

$$\det(A) = \det(P)^{-1} \cdot \det(L) \cdot \det(U) = 1,$$

and

$$A^{-1} = U^{-1} \cdot L^{-1} \cdot P = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Course 8

3.3 The matrix of a list of vectors

In the previous section, we have seen a matrix as a list of row-vectors. Now we discuss a converse, namely we define the matrix associated to a list of vectors, with respect to a basis.

Definition 3.3.1 Let V be a vector space over K, $B = (v_1, \dots, v_n)$ a basis of V and $X = (u_1, \dots, u_m)$ a list of vectors in V. Let

$$\begin{cases} u_1 = a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ u_2 = a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \cdots\cdots\cdots\cdots\cdots \\ u_m = a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{cases}$$

be the unique writings of the vectors in X as linear combinations of vectors of the basis B, for some $a_{ij} \in K$. The *matrix of the list of vectors* X in the basis B is the matrix having as its rows the coordinates of the vectors in X in the basis B, that is,

$$[X]_B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Example 3.3.2 Consider the canonical basis $B = (e_1, e_2, e_3, e_4)$ and the list $X = (u_1, u_2, u_3)$ in the canonical real vector space $\mathbb{R}^4$, where

$$\begin{cases} u_1 = (1, 2, 3, 4) \\ u_2 = (5, 6, 7, 8) \\ u_3 = (9, 10, 11, 12) \end{cases}.$$

Since the coordinates of a vector in the canonical basis are just its components, we get

$$[X]_B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}.$$

Now we give a theorem which allows one to determine the dimension of the subspace generated by a list of vectors.

Theorem 3.3.3 *Let V be a vector space over K, $B = (v_1, \dots, v_n)$ a basis of V and $X = (u_1, \dots, u_m)$ a list of vectors in V having the matrix A in the basis B. Then:*
(i) $\dim\langle X\rangle = \operatorname{rank}(A)$.
(ii) A basis of $\langle X\rangle$ is the list of non-zero row-vectors $(c_1, \dots, c_r)$ of an echelon form C equivalent to A.

Example 3.3.4 Let us determine the dimensions of the subspaces S, T, $S+T$ and $S\cap T$ of the canonical real vector space $\mathbb{R}^4$, where

$$S = \langle(-3, 5, -1, 1), (-1, 1, 0, 1), (1, 1, -1, -3)\rangle,$$
$$T = \langle(1, 0, 2, 0), (2, 1, -1, 2)\rangle.$$

One can easily show that the ranks of the matrices in the canonical basis corresponding to the vectors from S and from T respectively are both 2. Hence $\dim S = \dim T = 2$.

Furthermore, $S+T=\langle S\cup T\rangle$. We write the matrix of $S\cup T$ in the canonical basis and we have

$$\begin{pmatrix}-3&5&-1&1\\-1&1&0&1\\1&1&-1&-3\\1&0&2&0\\2&1&-1&2\end{pmatrix}\sim\begin{pmatrix}1&1&-1&-3\\-1&1&0&1\\-3&5&-1&1\\1&0&2&0\\2&1&-1&2\end{pmatrix}\sim\begin{pmatrix}1&1&-1&-3\\0&2&-1&-2\\0&8&-4&-8\\0&-1&3&3\\0&-1&1&8\end{pmatrix}\sim\begin{pmatrix}1&1&-1&-3\\0&-1&3&3\\0&2&-1&-2\\0&2&-1&-2\\0&-1&1&8\end{pmatrix}$$
$$\sim\begin{pmatrix}1&1&-1&-3\\0&-1&3&3\\0&0&5&4\\0&0&5&4\\0&0&-2&5\end{pmatrix}\sim\begin{pmatrix}1&1&-1&-3\\0&-1&3&3\\0&0&5&4\\0&0&0&\frac{33}{5}\\0&0&0&0\end{pmatrix}\sim\begin{pmatrix}1&1&-1&-3\\0&-1&3&3\\0&0&5&4\\0&0&0&33\\0&0&0&0\end{pmatrix}.$$

Then by Theorem 3.3.3, $\dim(S+T)=4$ and a basis of $S+T$ consists of the non-zero row-vectors from the echelon form, that is, $((1,1,-1,-3),(0,-1,3,3),(0,0,5,4),(0,0,0,33))$. Now by the Second Dimension Theorem, it follows that $\dim(S\cap T)=\dim S+\dim T-\dim(S+T)=2+2-4=0$.

Now we are going to define the matrix of a vector in a basis of a vector space. Even if one might expect to define it as a row-matrix, by considering a single vector list, it is more convenient to define it as a column-matrix for our purposes concerning linear maps in order to avoid formulas involving transposes.

Definition 3.3.5 Let V be a vector space over K, $v\in V$ and $B=(v_1,\dots,v_n)$ a basis of V. If $v=k_1v_1+\dots+k_nv_n$ $(k_1,\dots,k_n\in K)$ is the unique writing of v as a linear combination of the vectors of the basis B, then the *matrix of the vector* v in the basis B is

$$[v]_B=\begin{pmatrix}k_1\\\vdots\\k_n\end{pmatrix}.$$

Example 3.3.6 Consider the vector $v=(1,2,3)$ in the canonical real vector space $\mathbb{R}^3$, and let E be the canonical basis. Then $[v]_E=\begin{pmatrix}1\\2\\3\end{pmatrix}$.

3.4 The matrix of a linear map

Definition 3.4.1 Let $f:V\to V'$ be a K-linear map, $B=(v_1,\dots,v_n)$ a basis of V and $B'=(v_1',\dots,v_m')$ a basis of V'. Then we can uniquely write the vectors in $f(B)$ as linear combinations of the vectors of the basis B', say

$$\begin{cases}f(v_1)=a_{11}v_1'+a_{21}v_2'+\dots+a_{m1}v_m'\\f(v_2)=a_{12}v_1'+a_{22}v_2'+\dots+a_{m2}v_m'\\\dots\dots\dots\dots\dots\\f(v_n)=a_{1n}v_1'+a_{2n}v_2'+\dots+a_{mn}v_m'\end{cases}$$

for some $a_{ij}\in K$.
Then the *matrix of the K-linear map f* in the bases B and B' is the matrix having as its columns the coordinates of the vectors of $f(B)$ in the basis B', that is,

$$[f]_{BB'}=\begin{pmatrix}a_{11}&a_{12}&\dots&a_{1n}\\a_{21}&a_{22}&\dots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\dots&a_{mn}\end{pmatrix}.$$

If $V=V'$ and $B=B'$, then we simply denote $[f]_B=[f]_{BB'}$.

Remark 3.4.2 We have to emphasize that we put the coordinates on the columns of the matrix of a linear map and not on the rows as we did for the matrix of a list of vectors.

Example 3.4.3 Consider the $\mathbb{R}$-linear map $f : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$f(x, y, z, t) = (x + y + z, y + z + t, z + t + x), \ \forall (x, y, z, t) \in \mathbb{R}^4 .$$

Let $E = (e_1, e_2, e_3, e_4)$ and $E' = (e'_1, e'_2, e'_3)$ be the canonical bases in $\mathbb{R}^4$ and $\mathbb{R}^3$ respectively. Since

$$\begin{cases} f(e_1) = f(1, 0, 0, 0) = (1, 0, 1) = e'_1 + e'_3 \\ f(e_2) = f(0, 1, 0, 0) = (1, 1, 0) = e'_1 + e'_2 \\ f(e_3) = f(0, 0, 1, 0) = (1, 1, 1) = e'_1 + e'_2 + e'_3 \\ f(e_4) = f(0, 0, 0, 1) = (0, 1, 1) = e'_2 + e'_3 \end{cases}$$

it follows that the matrix of f in the bases E and E' is

$$[f]_{EE'} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} .$$

Theorem 3.4.4 *Let $f : V \to V'$ be a K-linear map, $B = (v_1, \ldots, v_n)$ a basis of V, $B' = (v'_1, \ldots, v'_m)$ a basis of V' and $v \in V$. Then*

$$[f(v)]_{B'} = [f]_{BB'} \cdot [v]_B .$$

Proof. Let $[f]_{BB'} = (a_{ij}) \in M_{m,n}(K)$. Let $v = \sum_{j=1}^{n} k_j v_j$ and

$$f(v) = \sum_{i=1}^{m} k'_i v'_i$$

for some $k_i, k'_i \in K$. On the other hand, using the definition of the matrix of f in the bases B and B', we have

$$f(v) = f\left(\sum_{j=1}^{n} k_j v_j\right) = \sum_{j=1}^{n} k_j f(v_j) = \sum_{j=1}^{n} k_j \left(\sum_{i=1}^{m} a_{ij} v'_i\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} k_j\right) v'_i .$$

But the writing of $f(v)$ as a linear combination of the vectors of the basis B' is unique, hence we must have $k'_i = \sum_{j=1}^{n} a_{ij} k_j$ for every $i \in \{1, \ldots, m\}$. Therefore, $[f(v)]_{B'} = [f]_{BB'} \cdot [v]_B$. □

Now we give a connection between the ranks of a linear map and of its matrix in a pair of bases.

Theorem 3.4.5 *Let $f : V \to V'$ be a K-linear map. Then*

$$\operatorname{rank}(f) = \operatorname{rank}([f]_{BB'}) ,$$

where B and B' are any bases of V and V' respectively.

Proof. Let $B = (v_1, \ldots, v_n)$ and $[f]_{BB'} = A$. Using our results relating ranks and dimensions, we have

$$\begin{aligned} \operatorname{rank}(f) &= \dim(\operatorname{Im} f) = \dim f(V) = \dim f(\langle v_1, \ldots, v_n \rangle) \\ &= \dim \langle f(v_1), \ldots, f(v_n) \rangle = \operatorname{rank}(A^T) = \operatorname{rank}(A) = \operatorname{rank}([f]_{BB'}) . \end{aligned}$$

Now take some other bases $B_1 = (u_1, \ldots, u_n)$ of V and B'_1 of V' and denote $[f]_{B_1 B'_1} = A_1$. Then

$$\begin{aligned} \operatorname{rank}([f]_{B_1 B'_1}) &= \operatorname{rank}(A_1) = \operatorname{rank}(A_1^T) = \dim \langle f(u_1), \ldots, f(u_n) \rangle \\ &= \dim(\operatorname{Im} f) = \dim \langle f(v_1), \ldots, f(v_n) \rangle = \operatorname{rank}([f]_{BB'}) . \end{aligned}$$

This shows the result. □

Remark 3.4.6 Notice that the rank of a linear map does not depend on the pair of bases in which we write its matrix. Also notice that, considering matrices of a linear map in different pairs of bases, their ranks are the same. Some other connection between matrices of a linear map in different pairs of bases will be discussed in the next section.

Example 3.4.7 Consider the $\mathbb{R}$-linear map $f : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$f(x, y, z, t) = (x + y + z, y + z + t, z + t + x), \ \forall (x, y, z, t) \in \mathbb{R}^4 .$$

Let $E = (e_1, e_2, e_3, e_4)$ and $E' = (e'_1, e'_2, e'_3)$ be the canonical bases in $\mathbb{R}^4$ and $\mathbb{R}^3$ respectively. Using Example 3.4.3 it follows that

$$[f]_{EE'} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} .$$

Now by Theorem 3.4.5 it follows that $\operatorname{rank}(f) = \operatorname{rank}([f]_{EE'}) = 3$.

We end this section with a key result in Linear Algebra, connecting linear maps and matrices.

Theorem 3.4.8 *Let V, V' and V'' be vector spaces over K with $\dim V = n$, $\dim V' = m$ and $\dim V'' = p$ and let $B = (v_1, \ldots, v_n)$, $B' = (v'_1, \ldots, v'_m)$ and $B'' = (v''_1, \ldots, v''_p)$ be bases of V, V' and V'' respectively. Then $\forall f, g \in \operatorname{Hom}_K(V, V')$, $\forall h \in \operatorname{Hom}_K(V', V'')$ and $\forall k \in K$, we have*

$$\begin{aligned} [f + g]_{BB'} &= [f]_{BB'} + [g]_{BB'} , \\ [kf]_{BB'} &= k \cdot [f]_{BB'} , \\ [h \circ f]_{BB''} &= [h]_{B'B''} \cdot [f]_{BB'} . \end{aligned}$$

Proof. Let $[f]_{BB'} = (a_{ij}) \in M_{m,n}(K)$, $[g]_{BB'} = (b_{ij}) \in M_{m,n}(K)$ and $[h]_{B'B''} = (c_{ki}) \in M_{pm}(K)$. Then

$$f(v_j) = \sum_{i=1}^{m} a_{ij} v'_i , \quad g(v_j) = \sum_{i=1}^{m} b_{ij} v'_i , \quad h(v'_i) = \sum_{k=1}^{p} c_{ki} v''_k$$

$\forall j \in \{1, \ldots, n\}$ and $\forall i \in \{1, \ldots, m\}$.

Then $\forall k \in K$ and $\forall j \in \{1, \ldots, n\}$ we have

$$\begin{aligned} (f + g)(v_j) = f(v_j) + g(v_j) = \sum_{i=1}^{m} a_{ij} v'_i + \sum_{i=1}^{m} b_{ij} v'_i = \sum_{i=1}^{m} (a_{ij} + b_{ij}) v'_i , \\ (kf)(v_j) = kf(v_j) = k \cdot \left(\sum_{i=1}^{m} a_{ij} v'_i \right) = \sum_{i=1}^{m} (k a_{ij}) v'_i , \end{aligned}$$

hence $[f + g]_{BB'} = [f]_{BB'} + [g]_{BB'}$ and $[kf]_{BB'} = k \cdot [f]_{BB'}$.

Finally, $\forall j \in \{1, \ldots, n\}$ we have

$$\begin{aligned} (h \circ f)(v_j) = h(f(v_j)) = h \left(\sum_{i=1}^{m} a_{ij} v'_i \right) = \sum_{i=1}^{m} a_{ij} h(v'_i) \\ = \sum_{i=1}^{m} a_{ij} \left(\sum_{k=1}^{p} c_{ki} v''_k \right) = \sum_{k=1}^{p} \sum_{i=1}^{m} (c_{ki} a_{ij}) v''_k , \end{aligned}$$

hence $[h \circ f]_{BB''} = [h]_{B'B''} \cdot [f]_{BB'}$. □

Theorem 3.4.9 *Let V and V' be vector spaces over K with $\dim V = n$ and $\dim V' = m$, and let B and B' be bases of V and V' respectively. Then the map*

$$\varphi : \operatorname{Hom}_K(V, V') \to M_{m,n}(K), \quad \varphi(f) = [f]_{BB'} , \ \forall f \in \operatorname{Hom}_K(V, V')$$

is an isomorphism of vector spaces.

Proof. We have seen that $\mathrm{Hom}_K(V,V')$ is a vector space over K with respect to the following addition and scalar multiplication: $\forall f,g \in \mathrm{Hom}_K(V,V')$ and $\forall k \in K$, $f+g, k\cdot f \in \mathrm{Hom}_K(V,V')$, where $\forall x \in V$,

$$(f+g)(x) = f(x)+g(x)\,,$$
$$(kf)(x) = kf(x)\,.$$

Also, $M_{m,n}(K)$ is a vector space over K. By Theorem 3.4.8 it follows that φ is a K-linear map.

Finally, let us prove that φ is bijective. Consider $B=(v_1,\ldots,v_n)$ and $B'=(v_1',\ldots,v_m')$. Let $f,g \in \mathrm{Hom}_K(V,V')$ be such that $\varphi(f)=\varphi(g)$. Then $[f]_{BB'}=[g]_{BB'}=(a_{ij}) \in M_{m,n}(K)$, hence

$$f(v_j) = a_{1j}v_1' + a_{2j}v_2' + \cdots + a_{mj}v_m' = g(v_j)\,,$$

$\forall j \in \{1,\ldots,n\}$. We have seen that two K-linear maps are equal if and only if they have the same values at all vectors of a basis. Hence $f=g$, which shows that φ is injective. Now let $A=(a_{ij}) \in M_{m,n}(K)$, seen as a list of column-vectors $(a^1,\ldots,a^n)$, where $a^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$. Define a K-linear map $f: V \to V'$ on the basis of the domain by

$$f(v_j) = a_{1j}v_1' + \cdots + a_{mj}v_m'\,,$$

$\forall j \in \{1,\ldots,n\}$. Then $\varphi(f)=[f]_{BB'}=(a_{ij})=A$. Thus, φ is surjective. □

Remark 3.4.10 The extremely important isomorphism given in Theorem 3.4.9 allows us to work with matrices instead of linear maps, which is much simpler from a computational point of view. Under this isomorphism, the kernel and the image of a linear map $f: V \to V'$, where V and V' are vector spaces over K with $\dim(V)=n$ and $\dim(V')=m$, and bases B and B' respectively, correspond to the *null space* and to the *column space* of its associated matrix $A=[f]_{BB'} \in M_{m,n}(K)$ respectively. Thus, the *null space* of A consists of vectors $x \in K^n$ such that $Ax=0$, while the *column space* of A consists of all linear combinations of the columns of A. A vector $b \in K^m$ belongs to the column space of A if and only if the system $Ax=b$ has a solution. By the First Dimension Theorem it follows that the sum of the dimensions of the null space and the column space of A equals n.

Theorem 3.4.11 *Let V be a vector space over K with $\dim V = n$, and let B be a basis of V. Then the map*

$$\varphi: \mathrm{End}_K(V) \to M_n(K), \quad \varphi(f)=[f]_B\,,\ \forall f \in \mathrm{End}_K(V)$$

is an isomorphism of vector spaces and of rings.

Proof. Note that $(\mathrm{End}_K(V),+,\circ)$ and $(M_n(K),+,\cdot)$ are rings. The required isomorphisms follow by Theorem 3.4.9. □

Corollary 3.4.12 *Let $f \in \mathrm{End}_K(V)$. Then $f \in \mathrm{Aut}_K(V) \iff \det([f]_B) \neq 0$, where B is any basis of V.*

Proof. Let B a basis of V. By Theorem 3.4.11, $f \in \mathrm{Aut}_K(V) \iff f$ is invertible in the ring $(\mathrm{End}_K(V),+,\circ) \iff [f]_B$ is invertible in the ring $(M_n(K),+,\cdot) \iff \det([f]_B) \neq 0$. □

Extra: Hill cipher

Let $n \in \mathbb{N}^*$ and consider the canonical vector space $V=\mathbb{Z}_2^n$ over $\mathbb{Z}_2$ with canonical basis E. The vectors of V may be identified with n-bit binary strings. Suppose that Alice needs to send an n-bit plaintext $p \in \mathbb{Z}_2^n$ to Bob.

Hill cipher:

1. (*Key establishment*) Alice and Bob randomly choose an invertible matrix $K \in M_n(\mathbb{Z}_2)$ as a key, and compute its inverse.
2. (*Encryption*) Alice computes the ciphertext c according to the formula $[c]_E^T = [p]_E^T \cdot K$.

3. (*Decryption*) Bob computes the plaintext p according to the formula $[p]_E^T = [c]_E^T \cdot K^{-1}$.

Remark 3.4.13 The Hill cipher, which is nowadays insecure, was the first application of linear algebra to cryptography.

Example 3.4.14 Alice wants to send the message $p = (1, 0, 1) \in \mathbb{Z}_2^3$ to Bob.
Alice and Bob agree on the matrix

$$K = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in M_3(\mathbb{Z}_2)$$

as a key, and compute its inverse

$$K^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{Z}_2).$$

Alice encrypts the message by computing the ciphertext c as:

$$[c]_E^T = [p]_E^T \cdot K = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

Bob decrypts the message by computing the plaintext p as:

$$[p]_E^T = [c]_E^T \cdot K^{-1} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}.$$

Extra: Image transformations

Suppose that we have a 2D-image that we want to rotate counterclockwise with θ degrees around the origin. By such a rotation, the point of coordinates $(1, 0)$ becomes the point of coordinates $(\cos\theta, \sin\theta)$, while the point of coordinates $(0, 1)$ becomes the point of coordinates $(-\sin\theta, \cos\theta)$.
We look for an $\mathbb{R}$-linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ satisfying the following conditions:

$$f(1, 0) = (\cos\theta, \sin\theta),$$
$$f(0, 1) = (-\sin\theta, \cos\theta).$$

Recall that every linear map is determined by its values at the elements of a basis (the canonical basis in our case). Hence the matrix of the linear map f in the canonical basis E of the canonical real vector space $\mathbb{R}^2$ is:

$$[f]_E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

For any point $v = (x, y) \in \mathbb{R}^2$ of a 2D-image, its corresponding point in the rotated image is computed as $f(v) = (x', y') \in \mathbb{R}^2$, where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = [f(v)]_E = [f]_E \cdot [v]_E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

For instance, for a counterclockwise rotation of 90^0 around the origin one has the matrix:

$$[f]_E = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Extra: Graphs and networks (see [Crivei])

Course 9

3.5 Change of bases

We would like now to establish a connection between the matrices of a vector or of a linear map in different bases of a vector space.

Definition 3.5.1 Let V be a vector space over K, and let $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ be bases of V. Then we can uniquely write

$$\begin{cases} v'_1 = t_{11}v_1 + t_{21}v_2 + \cdots + t_{n1}v_n \\ v'_2 = t_{12}v_1 + t_{22}v_2 + \cdots + t_{n2}v_n \\ \cdots\cdots\cdots\cdots\cdots \\ v'_n = t_{1n}v_1 + t_{2n}v_2 + \cdots + t_{nn}v_n \end{cases}$$

for some $t_{ij} \in K$. Then the matrix $(t_{ij}) \in M_n(K)$, having as its columns the coordinates of the vectors of the basis B' in the basis B, is called the *change matrix* (or *transition matrix*) *from the basis B to the basis B'* and is denoted by $T_{BB'}$.

Remark 3.5.2 (1) B and B' are referred to as the "old" basis and the "new" basis respectively.

(2) The change matrix may be related to the matrix of a linear map as follows. For every $j \in \{1, \ldots, n\}$, the j^{th} column of $T_{BB'}$ consists of the coordinates of $v'_j = 1_V(v'_j)$ in the basis B, hence $T_{BB'} = [1_V]_{B'B}$.

Theorem 3.5.3 *Let V be a vector space over K, and let $B = (v_1, \ldots, v_n)$, $B' = (v'_1, \ldots, v'_n)$ and $B'' = (v''_1, \ldots, v''_n)$ be bases of V. Then*

$$T_{BB''} = T_{BB'} \cdot T_{B'B''}.$$

Proof. Using a previous theorem, we have

$$T_{BB'} \cdot T_{B'B''} = [1_V]_{B'B} \cdot [1_V]_{B''B'} = [1_V \circ 1_V]_{B''B} = [1_V]_{B''B} = T_{BB''}.$$

We also present a direct proof. Denote $T_{BB'} = (t_{ij}) \in M_n(K)$, $T_{B'B''} = (t'_{jk}) \in M_n(K)$ and $T_{BB''} = (t''_{ik}) \in M_n(K)$. Then we have

$$v'_j = \sum_{i=1}^{n} t_{ij}v_i\,, \quad \forall j \in \{1, \ldots, n\},$$

$$v''_k = \sum_{j=1}^{n} t'_{jk}v'_j\,, \quad \forall k \in \{1, \ldots, n\}\,,$$

$$v''_k = \sum_{i=1}^{n} t''_{ik}v_i\,, \quad \forall k \in \{1, \ldots, n\}.$$

For every $k \in \{1, \ldots, n\}$ it follows that

$$v''_k = \sum_{j=1}^{n} t'_{jk}\left(\sum_{i=1}^{n} t_{ij}v_i\right) = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} t_{ij}t'_{jk}\right)v_i.$$

By the uniqueness of writing of each v''_k as a linear combination of the vectors of the basis B, for every $i, k \in \{1, \ldots, n\}$ it follows that

$$t''_{ik} = \sum_{j=1}^{n} t_{ij}t'_{jk}.$$

This shows that $T_{BB''} = T_{BB'} \cdot T_{B'B''}$. □

Theorem 3.5.4 *Let V be a vector space over K, and let B and B' be bases of V. Then the change matrix $T_{BB'}$ is invertible and its inverse is the change matrix $T_{B'B}$.*

Proof. Using Theorem 3.5.3 for $B'' = B$, we have

$$T_{BB'}T_{B'B} = T_{BB} = I_n.$$

Using again Theorem 3.5.3 and changing the roles for B, B' and B'' by B', B and B' respectively, we have

$$T_{B'B}T_{BB'} = T_{B'B'} = I_n.$$

Hence $T_{BB'}$ is invertible and $T_{BB'}^{-1} = T_{B'B}$. □

Let us now see how one can use the change matrix from one basis to another in order to compute the coordinates of a vector in different bases or the matrix of a linear map in different bases.

Theorem 3.5.5 *Let V be a vector space over K, let $B = (v_1, \ldots, v_n)$ and $B' = (v_1', \ldots, v_n')$ be bases of V and let $v \in V$. Then*

$$[v]_B = T_{BB'} \cdot [v]_{B'} .$$

Proof. Using a previous theorem, we have

$$T_{BB'} \cdot [v]_{B'} = [1_V]_{B'B} \cdot [v]_{B'} = [1_V(v)]_B = [v]_B.$$

We also present a direct proof. Consider the writings of the vector $v \in V$ in the two bases B and B', say $v = \sum_{i=1}^n k_i v_i$ and $v = \sum_{j=1}^n k_j' v_j'$ for some $k_i, k_j' \in K$. Since $T_{BB'} = (t_{ij}) \in M_n(K)$, we have

$$v_j' = \sum_{i=1}^n t_{ij} v_i , \quad \forall j \in \{1, \ldots, n\}.$$

It follows that

$$v = \sum_{j=1}^n k_j' \left(\sum_{i=1}^n t_{ij} v_i \right) = \sum_{i=1}^n \left(\sum_{j=1}^n t_{ij} k_j' \right) v_i .$$

By the uniqueness of writing of v as a linear combination of the vectors of the basis B, it follows that $k_i = \sum_{j=1}^n t_{ij} k_j'$, whence $[v]_B = T_{BB'} \cdot [v]_{B'}$. □

Remark 3.5.6 Usually, we are interested in computing the coordinates of a vector v in the new basis B', knowing the coordinates of the same vector v in the old basis B and the change matrix from B to B'. Then by Theorem 3.5.5, we have

$$[v]_{B'} = T_{BB'}^{-1} \cdot [v]_B = T_{B'B} \cdot [v]_B .$$

Example 3.5.7 Consider the bases $E = (e_1, e_2, e_3)$ and $B = (v_1, v_2, v_3)$ of the canonical real vector space $\mathbb{R}^3$, where E is the canonical basis and $v_1 = (0, 1, 1)$, $v_2 = (1, 1, 2)$, $v_3 = (1, 1, 1)$. Let us determine the change matrices from E to B and viceversa. We have

$$\begin{cases} v_1 = \quad e_2 + e_3 \\ v_2 = e_1 + e_2 + 2e_3 \\ v_3 = e_1 + e_2 + e_3 \end{cases}$$

which implies

$$\begin{cases} e_1 = -v_1 \quad + v_3 \\ e_2 = v_1 - v_2 + v_3 \\ e_3 = \quad v_2 - v_3 \end{cases} .$$

Hence we get

$$T_{EB} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \quad T_{BE} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

We must have $T_{BE} = T_{EB}^{-1}$, so that we could have obtained T_{BE} by computing the inverse of T_{EB}.

Now consider the vector $u = (1, 2, 3)$. Clearly, its coordinates in the canonical basis E are 1, 2 and 3. By Theorem 3.5.5, it follows that

$$[u]_B = T_{BE} \cdot [u]_E = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Hence the coordinates of u in the basis B are 1, 1 and 0.

Next we give a theorem relating the matrices of a linear map in different bases.

Theorem 3.5.8 *Let $f \in \mathrm{Hom}_K(V, V')$, let B_1 and B_2 be bases of V and let B_1' and B_2' be bases of V'. Then*

$$[f]_{B_2B_2'} = T_{B_1'B_2'}^{-1} \cdot [f]_{B_1B_1'} \cdot T_{B_1B_2}.$$

Proof. We have

$$\begin{aligned} T_{B_1'B_2'}^{-1} \cdot [f]_{B_1B_1'} \cdot T_{B_1B_2} &= T_{B_2'B_1'} \cdot [f]_{B_1B_1'} \cdot T_{B_1B_2} \\ &= [1_V]_{B_1'B_2'} \cdot [f]_{B_1B_1'} \cdot [1_V]_{B_2B_1} = [1_V \circ f \circ 1_V]_{B_2B_2'} = [f]_{B_2B_2'}, \end{aligned}$$

which shows the result. □

Corollary 3.5.9 *Let $f \in \mathrm{End}_K(V)$, and let B and B' be bases of V. Then*

$$[f]_{B'} = T_{BB'}^{-1} \cdot [f]_B \cdot T_{BB'}.$$

Proof. This follows by Theorem 3.5.8 with $B_1 = B_1' = B$ and $B_2 = B_2' = B'$. □

Example 3.5.10 Consider the bases $E = (e_1, e_2, e_3)$ and $B = (v_1, v_2, v_3)$ of the canonical real vector space $\mathbb{R}^3$, where E is the canonical basis and $v_1 = (0, 1, 1)$, $v_2 = (1, 1, 2)$, $v_3 = (1, 1, 1)$. Also let $f \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^3)$ be defined by

$$f(x, y, z) = (x + y, y - z, z + x), \ \forall (x, y, z) \in \mathbb{R}^3.$$

Let us determine the matrix of f in the basis E and in the basis B. We have

$$\begin{cases} f(e_1) = (1, 0, 1) = e_1 + e_3 \\ f(e_2) = (1, 1, 0) = e_1 + e_2 \\ f(e_3) = (0, -1, 1) = -e_2 + e_3 \end{cases}$$

which implies that

$$[f]_E = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Using Corollary 3.5.9 and the change matrices T_{EB} and T_{BE}, that we have determined in Example 3.5.7, we have

$$\begin{aligned} [f]_B &= T_{EB}^{-1} \cdot [f]_E \cdot T_{EB} = T_{BE} \cdot [f]_E \cdot T_{EB} \\ &= \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -3 & -2 \\ 1 & 4 & 2 \\ 0 & -2 & 0 \end{pmatrix}. \end{aligned}$$

It is worth to be mentioned that we could have reached the same result using the definition of the matrix of a linear map and expressing the vectors $f(v_1)$, $f(v_2)$ and $f(v_3)$ as linear combinations of the vectors v_1, v_2 and v_3 of the basis B.

3.6 Eigenvectors and eigenvalues

The study of endomorphisms of vector spaces also makes use of vectors whose images are just scalar multiples of themselves, in other words vectors that are "stretched" by an endomorphism. They are the subject of the present section.

Definition 3.6.1 Let $f \in \mathrm{End}_K(V)$. A non-zero vector $v \in V$ is called an *eigenvector of* f if there exists $\lambda \in K$ such that $f(v) = \lambda \cdot v$. Here λ is called an *eigenvalue of* f.

Remark 3.6.2 Clearly, each eigenvector has a unique corresponding eigenvalue. But different eigenvectors may have the same corresponding eigenvalue.

For $f \in \mathrm{End}_K(V)$, denote $V(\lambda) = \{v \in V \mid f(v) = \lambda v\}$, that is, the set consisting of the zero vector and the eigenvectors of f with eigenvalue λ.

Theorem 3.6.3 *Let $f \in \mathrm{End}_K(V)$ and let λ be an eigenvalue of f. Then $V(\lambda)$ is a subspace of V.*

Proof. Clearly, $0 \in V(\lambda)$, hence $V(\lambda) \neq \emptyset$. Now let $k_1, k_2 \in K$ and $v_1, v_2 \in V(\lambda)$. Then we have $f(v_1) = \lambda v_1$ and $f(v_2) = \lambda v_2$. It follows that

$$\begin{aligned} f(k_1v_1 + k_2v_2) &= k_1f(v_1) + k_2f(v_2) = k_1(\lambda v_1) + k_2(\lambda v_2) \\ &= (k_1\lambda)v_1 + (k_2\lambda)v_2 = \lambda(k_1v_1 + k_2v_2)\,. \end{aligned}$$

Hence, $k_1v_1 + k_2v_2 \in V(\lambda)$ and consequently, $V(\lambda)$ is a subspace of V. □

Definition 3.6.4 Let $f \in \mathrm{End}_K(V)$ and let λ be an eigenvalue of f. Then $V(\lambda)$ is called the *eigenspace* (or the *characteristic subspace*) of λ with respect to f.

The next theorem offers the essence of the practical method to determine eigenvalues and eigenvectors.

Theorem 3.6.5 *Let V be a vector space over K, B a basis of V and $f \in \mathrm{End}_K(V)$ with the matrix $[f]_B = A = (a_{ij}) \in M_n(K)$. Then $\lambda \in K$ is an eigenvalue of f if and only if*

$$\det(A - \lambda \cdot I_n) = 0 \tag{1}$$

Proof. The element $\lambda \in K$ is an eigenvalue of f if and only if there exists a non-zero $v \in V$ such that $f(v) = \lambda v$. Consider $[v]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then it follows that

$$\begin{aligned} &f(v) = \lambda v \iff f(v) - \lambda v = 0 \iff (f - \lambda \cdot 1_V)(v) = 0 \iff [(f - \lambda \cdot 1_V)(v)]_B = [0]_B \\ &\iff [f - \lambda \cdot 1_V]_B \cdot [v]_B = [0]_B \iff ([f]_B - \lambda \cdot [1_V]_B) \cdot [v]_B = [0]_B \\ &\iff (A - \lambda \cdot I_n) \cdot [v]_B = [0]_B \\ &\iff \begin{pmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \\ &\iff \begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0. \end{cases} \end{aligned}$$

Then λ is an eigenvalue of f if and only if the final system (S) of linear equations has a non-zero solution if and only if its determinant $\det(A - \lambda \cdot I_n)$ is zero. □

Definition 3.6.6 The equality (1) is called the *characteristic equation* and the system (S) is called the *characteristic system*. The determinant $\det(A-\lambda I_n)$ may be seen as a polynomial $p_A(\lambda)$ in λ and it is called the *characteristic polynomial of* f with respect to A (or the *characteristic polynomial* of A).

Now a question arises naturally: if we take another basis B' of V and use the matrix $[f]_{B'}$, do we get the same eigenvalues and eigenvectors of f? We will show that the answer is positive.

Theorem 3.6.7 *Let V be a vector space over K, B and B' bases of V and $f \in End_K(V)$ with the matrices $[f]_B = A \in M_n(K)$ and $[f]_{B'} = A' \in M_n(K)$. Then $p_A(\lambda) = p_{A'}(\lambda)$.*

Proof. We have $[f]_{B'} = T_{BB'}^{-1} \cdot [f]_B \cdot T_{BB'}$. Denote $T = T_{BB'}$. Hence we have $A' = T^{-1} \cdot A \cdot T$. Then

$$\begin{aligned} p_{A'}(\lambda) &= \det(A' - \lambda I_n) = \det(T^{-1}AT - \lambda I_n T^{-1}T) = \det(T^{-1}(A - \lambda I_n)T) \\ &= \det(T^{-1}) \cdot \det(A - \lambda I_n) \cdot \det(T) = \det(A - \lambda I_n) = p_A(\lambda), \end{aligned}$$

which proves the result. □

Remark 3.6.8 (1) Therefore, the eigenvalues and the eigenvectors *do not depend* on the basis chosen for writing the matrix of the endomorphism. Of course, the matrices might be different, but in the end we get the same characteristic polynomial. Consequently, we can say that the eigenvalues of an endomorphism (or simply, of a matrix) are just the roots in K of its unique characteristic polynomial.

(2) If V is a vector space over K with $\dim V = n$ and $f \in \mathrm{End}_K(V)$, then the degree of the characteristic polynomial of f is n, hence f may have at most n eigenvalues. If $K = \mathbb{C}$, then by the Fundamental Theorem of Algebra f has exactly n eigenvalues, not necessarily distinct.

(3) A non-zero vector $v \in K^n$ is an eigenvector of a matrix $A \in M_n(K)$ if and only if there exists $\lambda \in K$ such that $A[v]_E = \lambda[v]_E$, where E is the canonical basis of the canonical vector space K^n over K. In this case, λ is an eigenvalue of A.

Example 3.6.9 Let $f \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^3)$ be defined by

$$f(x, y, z) = (2x, y + 2z, -y + 4z), \ \forall (x, y, z) \in \mathbb{R}^3.$$

We write its matrix in the simplest basis, namely in the canonical basis E of $\mathbb{R}^3$. Then

$$[f]_E = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 4 \end{pmatrix}.$$

The characteristic polynomial is $p(\lambda) = -(\lambda - 2)^2(\lambda - 3)$, so the eigenvalues are $\lambda_1 = \lambda_2 = 2$ and $\lambda_3 = 3$.

Let us take first $\lambda_1 = \lambda_2 = 2$. An eigenvector (x_1, x_2, x_3) is a non-zero solution of the characteristic system

$$\begin{pmatrix} 2 - \lambda_1 & 0 & 0 \\ 0 & 1 - \lambda_1 & 2 \\ 0 & -1 & 4 - \lambda_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

that is,

$$\begin{cases} -x_2 + 2x_3 = 0 \\ -x_2 + 2x_3 = 0 \end{cases}.$$

Then $x_2 = 2x_3$ and $x_1, x_3 \in \mathbb{R}$, whence

$$V(2) = \{(x_1, 2x_3, x_3) \mid x_1, x_3 \in \mathbb{R}\} = \langle (1, 0, 0), (0, 2, 1) \rangle.$$

Any non-zero vector in $V(2)$ is an eigenvector of f with the associated eigenvalue $\lambda_1 = \lambda_2 = 2$.

Consider now $\lambda_3 = 3$. The corresponding characteristic system is

$$\begin{pmatrix} 2-\lambda_3 & 0 & 0 \\ 0 & 1-\lambda_3 & 2 \\ 0 & -1 & 4-\lambda_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

that is,

$$\begin{cases} -x_1 = 0 \\ -2x_2 + 2x_3 = 0 \\ -x_2 + x_3 = 0 \end{cases}.$$

We get the solution $x_1 = 0$, $x_2 = x_3$ and $x_3 \in \mathbb{R}$. Then

$$V(3) = \{(0, x_3, x_3) \mid x_3 \in \mathbb{R}\} = \langle (0,1,1) \rangle.$$

Any non-zero vector in $V(3)$ is an eigenvector of f with the associated eigenvalue $\lambda_3 = 3$.

For $A \in M_n(K)$, $\mathrm{Tr}(A)$ is the *trace* of A, that is, the sum of the elements of the main diagonal of A.

Theorem 3.6.10 *Let $A \in M_n(K)$ having eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:*
(i) $\lambda_1 + \cdots + \lambda_n = \mathrm{Tr}(A)$.
(ii) $\lambda_1 \cdots \lambda_n = \det(A)$.

The following famous theorem involves the characteristic polynomial.

Theorem 3.6.11 (Cayley-Hamilton Theorem) *Every matrix $A \in M_n(K)$ is a root of its characteristic polynomial.*

Corollary 3.6.12 *Let $A \in M_2(K)$. Then:*
(i) the characteristic polynomial of A is $p_A(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A)$.
(ii) $A^2 - \mathrm{Tr}(A) \cdot A + \det(A) \cdot I_2 = 0_2$.

Cayley-Hamilton Theorem may be used for computing the inverse or powers of a matrix.

Example 3.6.13 Let

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

Then $\det(A) = 2 \neq 0$, hence A is invertible. Its characteristic polynomial is

$$p_A(\lambda) = \det \begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 1 & 1-\lambda \end{pmatrix} = -\lambda^3 + 4\lambda^2 - 5\lambda + 2\,.$$

By Theorem 3.6.11, we have

$$A^3 - 4A^2 + 5A - 2I_3 = 0_3\,.$$

It follows that

$$A\left[\frac{1}{2}(A^2 - 4A + 5I_3)\right] = \left[\frac{1}{2}(A^2 - 4A + 5I_3)\right] A = I_3,$$

whence

$$A^{-1} = \frac{1}{2}(A^2 - 4A + 5I_3) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -2 & 2 \end{pmatrix}.$$

For $k \geq 3$, the powers A^k can be computed using the recurrence relation given by Theorem 3.6.11, namely

$$A^k = 4A^{k-1} - 5A^{k-2} + 2A^{k-3}\,.$$

The theory of eigenvectors and eigenvalues of an endomorphism is important for deciding whether an endomorphism is *diagonalizable* in the sense that there is a basis in which its matrix is diagonal (i.e., it has possibly non-zero entries only on its main diagonal), which is a much more useful computational form. As a sample result in this sense, we give the following theorem, whose proof will be omitted.

Theorem 3.6.14 *Let V be a vector space over K with $\dim V = n$ and $f \in \mathrm{End}_K(V)$. Then f is diagonalizable if and only if it has n linearly independent eigenvectors.*
In particular, if f has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then f is diagonalizable and

$$[f]_B = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

where B is the basis of the corresponding eigenvectors.

Extra: PageRank

PageRank is a number assigned by Google to each web page. Pages with higher rank come higher in search results. We describe a simplified version, following [N. Strickland, *Linear Mathematics for Applications.* `https://neilstrickland.github.io/linear_maths/notes/linear_maths.pdf`.].

- Consider pages $S_1, \ldots, S_n$, with some links between them. A link from S_j to S_i is a vote by S_j that S_i is important.
- Links from important pages should count for more (because the probability of visiting S_i will clearly increase); links from pages with many links should count for less (because that will decrease the probability that we click the one that leads to S_i).
- We want rankings $r_1, \ldots, r_n \geq 0$, normalized so that $\sum_{i=1}^n r_i = 1$.
- Say S_j links to N_j different pages, and assume $N_j > 0$. We use the rule: a link from S_j to S_i contributes $\frac{r_j}{N_j}$ to r_i.
- Thus, for every $i \in \{1, \ldots, n\}$, the following consistency condition should be satisfied:

$$r_i = \sum_{j \in J_i} \frac{r_j}{N_j},$$

  where $J_i = \{j \in \{1, \ldots, n\} \mid \text{page } S_j \text{ links to page } S_i\}$.
- Define the matrix $P = (p_{ij}) \in M_n(\mathbb{R})$ by

$$p_{ij} = \begin{cases} \frac{1}{N_j} & \text{if there is a link from } S_j \text{ to } S_i \\ 0 & \text{otherwise.} \end{cases}$$

- Hence, for every $i \in \{1, \ldots, n\}$, the consistency condition becomes:

$$r_i = \sum_{j \in J_i} p_{ij} r_j.$$

- But this is equivalent to the matrix equation $Pr = r$, and thus r is an eigenvector of the matrix P with eigenvalue 1.

Extra: Singular value decomposition (see [Crivei])

Linear Algebra

Course 10

Chapter 4. Introduction to Coding Theory

Part I

Contents

Coding theory

Starting points:

- Shannon 1948: Information Theory
- Hamming 1950: Error-Correcting Codes

Main classes of codes:

- source coding: data compression
- channel coding: error-correcting codes

Probabilities of errors

Suppose that we have a communication channel whose probability of a correct transmission is p. The probability of t errors in a message of length m is

$$C_m^t p^{m-t}(1-p)^t.$$

For instance, for $p = 0.99$ and $m = 50$, we have the following table:

t	Probability of t errors
0	0.605
1	0.3056
2	0.0756
3	0.0122
4	0.00145

These probabilities decrease if m is small enough, more precisely when $m < \frac{p}{1-p}$. Hence we should not expect too many errors during a transmission. But still they happen, and should be detected and corrected.

A first example

EAN-13 International Article Number

It is a sequence of 13 digits $a_1, a_2, \ldots, a_{13}$ that identifies a product. Digit a_{13} is a check digit that is computed as

$$a_{13} = 10 - (a_1 + 3a_2 + a_3 + 3a_4 + \cdots + a_{11} + 3a_{12}) \bmod 10.$$

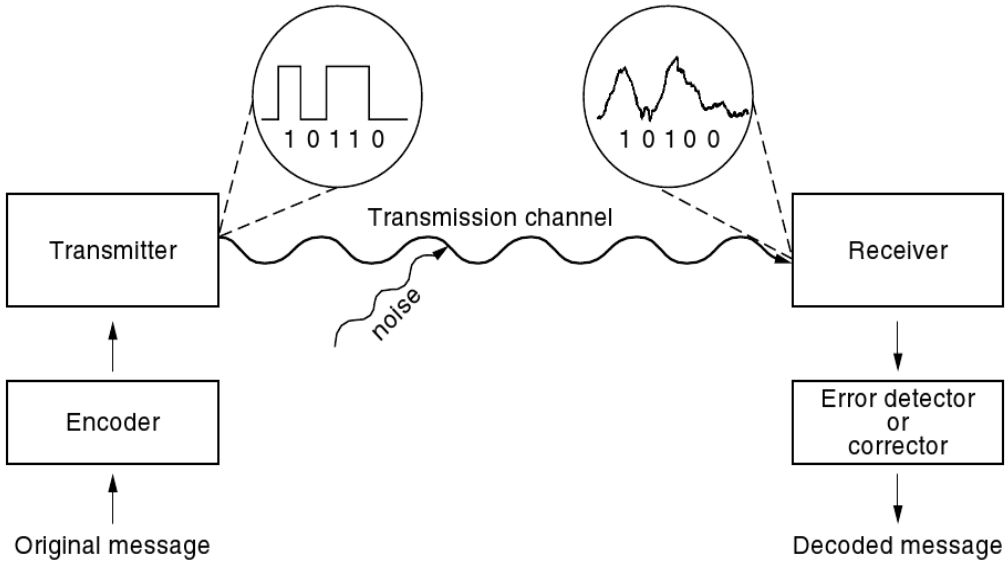
Digits are written in binary; black bars for 1, white bars for 0.

In particular:

- ISBN (International Standard Book Number)
- UPC (Universal Product Code) etc.

Error-correcting (detecting) codes

General scheme:

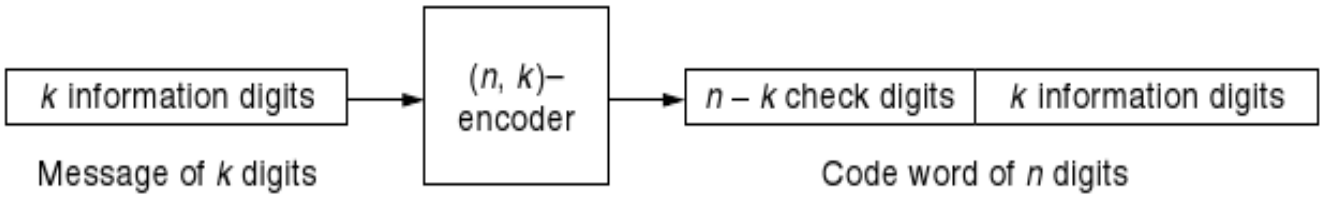


Different codes are suitable for different applications:

- satellite and space transmissions
- credit cards
- CD's, DVD's, Blu-ray discs etc.

The coding problem

- We discuss *binary codes*. In general: codes over finite fields.
- We consider *symmetric channels*: the probability of 1 being changed into 0 is the same as that of 0 being changed into 1.
- It is assumed that the number of errors is less than the number of correctly transmitted bits.
- We talk about (n, k)-*codes*:



There are 2^k possible messages, and so 2^k code words.
There are 2^n possible words received.

Aim

Find the right balance between k and $n - k$.

Two simple codes - The $(3, 2)$-parity check code

- The check digit is the sum modulo 2 of the message digits.
- Encoding:

Message	Code word
00	000
01	101
10	110
11	011

How many errors can this code detect/correct?

- Decoding:

Received words	101	111	100	000	110
Parity check	passes	fails	fails	passes	passes
Decoded words	01	-	-	00	10

Two simple codes - The $(3, 1)$-repeating code

- The two check digits repeat the message digit.
- Encoding:

Message	Code word
0	000
1	111

How many errors can this code detect/correct?

- Decoding:

Received words	111	010	011	000
Decoded words	1	0	1	0

Hamming distance

Definition

The *Hamming distance* between two words of the same length is the number of positions in which they difer.

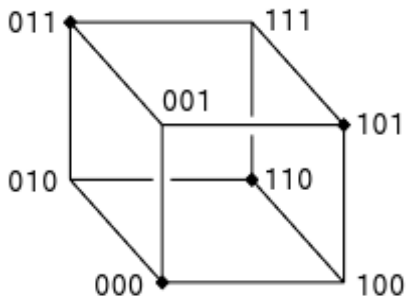
Notation $d(u, v)$.
Example: $d(101, 100) = 1$, $d(110, 001) = 3$, $d(101, 011) = 2$.

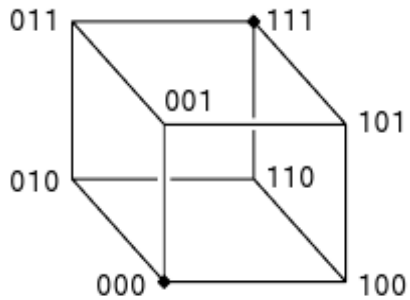
Theorem

The Hamming distance has the following properties hold for every $u, v, w \in \mathbb{Z}_2^n$*:*
(1) $d(u, v) = d(v, u)$.
(2) $d(u, v) + d(v, w) \geq d(u, w)$.
(3) $d(u, v) \geq 0$ *with equality if and only if* $u = v$.

Hamming distance - cont.

- In an (n, k)-code, the 2^n received words can be thought of as placed at the vertices of an n-dimensional cube with unit sides.
- The Hamming distance between two words is the shortest distance between their corresponding vertices along the edges of the n-cube.
- The 2^k code words form a subset of the 2^n vertices, and the code has better error-correcting and error-detecting capabilities the farther apart these code words are.
- Cube representations of the $(3, 2)$-parity check and $(3, 1)$-repeating codes:





Error detection/correction capabilities

Theorem

A code detects all sets of t or fewer errors $\iff$ the minimum Hamming distance between code words is at least $t+1$.

Theorem

A code corrects all sets of t or fewer errors $\iff$ the minimum Hamming distance between code words is at least $2t+1$.

Code	Minimum distance between words	No. of detectable errors	No. of correctable errors	Information rate
(n,k)-code	d	d-1	$\leq \frac{d-1}{2}$	$\frac{k}{n}$
$(3,2)$-parity check code	2	1	0	$\frac{2}{3}$
$(3,1)$-repeating code	3	2	1	$\frac{1}{3}$

Polynomial representation

- A binary n-digit word $a_0 a_1 \dots a_{n-1}$ may be identified with a polynomial $a_0 + a_1 X + \cdots + a_{n-1} X^{n-1} \in \mathbb{Z}_2[X]$.

Definition

Let $p \in \mathbb{Z}_2[X]$ be of degree $n - k$. The *polynomial code generated by* p is an (n, k)-code whose code words are those polynomials of degree less than n which are divisible by p. Then the polynomial p is called the *generator* of the code.

- A message of length k is represented by a polynomial $m \in \mathbb{Z}_2[X]$ of degree less than k.
- Since the message is stored in the right hand side of a word, the message digits are carried by the higher-order coefficients of a polynomial. So we consider $m \cdot X^{n-k}$.

Polynomial representation - cont.

- To encode the message polynomial m we first use the Division Algorithm to find unique $q, r \in \mathbb{Z}_2[X]$ such that

$$m \cdot X^{n-k} = q \cdot p + r\,, \quad degree(r) < degree(p) = n - k.$$

Then the code polynomial is

$$v = r + m \cdot X^{n-k}.$$

The check digits of the message are carried by r.

Theorem

With the above notation, the code polynomial v is divisible by p.

Proof. We have $v = r + m \cdot X^{n-k} = r + q \cdot p + r = q \cdot p$, because $r \in \mathbb{Z}_2[X]$, and so $r + r = 0$.

Polynomial representation - examples

Example 1. *Let* $p = 1 + X^2 + X^3 + X^4 \in \mathbb{Z}_2[X]$ *be the generator polynomial of a* $(7, 3)$*-code. Let us encode the message 101.*

Solution. Note that $n = 7$ and $k = 3$.

$$\begin{aligned}
\text{message } 101 &\rightsquigarrow m = 1 \cdot 1 + 0 \cdot X + 1 \cdot X^2 = 1 + X^2 \\
&\rightsquigarrow mX^{n-k} = (1 + X^2) \cdot X^4 = X^4 + X^6 \\
&\rightsquigarrow r = mX^{n-k} \bmod p = (X^4 + X^6) \bmod p = 1 + X \\
&\rightsquigarrow v = r + mX^{n-k} = 1 + X + X^4 + X^6 \\
&\rightsquigarrow \text{code word } \boxed{1100}\boxed{101}
\end{aligned}$$

Example 2. *If the generator polynomial of a* $(6, 3)$*-code is* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$*, test whether the following received words contain detectable errors: 100011, 100110.*

Solution. We check if the received words are code words, that is, their associated polynomials are divisible by p [...].

Polynomial representation - examples

Example 3. *Write down all the code words for the* $(6, 3)$*-code generated by the polynomial* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$.

Solution. Note that $n = 6$, $k = 3$, and we have $2^k = 8$ code words. We obtain the following table:

message	code word
000	000000
001	111001
010	011010
011	100011
100	110100
101	001101
110	101110
111	010111

E.g.: $110 \rightsquigarrow m = 1 + X \rightsquigarrow mX^{n-k} = X^3 + X^4$

$\rightsquigarrow r = mX^{n-k} \bmod p = (X^3 + X^4) \bmod p = 1 + X^2$

$\rightsquigarrow v = r + mX^{n-k} = 1 + X^2 + X^3 + X^4 \rightsquigarrow$ | 101 | 110 |

Linear Algebra

Course 11

Chapter 4. Introduction to Coding Theory

Part II

Contents

Matrix representation

- A binary n-digit word $a_0 a_1 \dots a_{n-1}$ may be identified with a matrix $\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} \in M_{n,1}(\mathbb{Z}_2)$.
- For an (n, k)-code, we see the 2^k possible messages as the elements of the vector space $\mathbb{Z}_2^k$ over $\mathbb{Z}_2$, and the 2^n possible received words as the elements of the vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$.

Definition

- An *encoder* of an (n, k)-code is an injective function $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ (or equivalently, $\gamma : M_{k,1}(\mathbb{Z}_2) \to M_{n,1}(\mathbb{Z}_2)$).
- An (n, k)-code is called *linear* if its encoder is a linear map.

Theorem

Any (n, k)-code generated by a polynomial of degree $n - k$ is linear.

E.g. *Reed-Solomon code*, used for CD's, DVD's, Blu-ray discs etc.

Generator matrix

Definition

Consider a linear (n, k)-code with encoder $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$. Let E, E' be the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^k$ and $\mathbb{Z}_2^n$ respectively. Then the matrix

$$G = [\gamma]_{EE'}$$

is called the *generator matrix* of the code.

A message $m \in \mathbb{Z}_2^k$ encodes as $\gamma(m)$.

But for $m \in \mathbb{Z}_2^k$, we have $[\gamma(m)]_{E'} = [\gamma]_{EE'} \cdot [m]_E$.

Hence a message $m \in M_{k,1}(\mathbb{Z}_2)$ encodes as $G \cdot [m]_E$.

Generator matrix - cont.

Use the above notation.

Theorem

(i) The code words of the (n, k)-code are the vectors in the subspace $\operatorname{Im}\gamma$ of $\mathbb{Z}_2^n$. Hence a binary (n, k)-code means a k-dimensional subspace of the vector space $\mathbb{Z}_2^n$.
(ii) The columns of G form a basis of this subspace, and so a vector is a code vector if and only if it is a unique linear combination of the columns of G.

Remark. A code word contains the message digits on the last k positions. Hence the generator matrix G of an (n, k)-code is always of the form

$$G = \begin{pmatrix} P \\ I_k \end{pmatrix} \in M_{n,k}(\mathbb{Z}_2),$$

where $P \in M_{n-k,k}(\mathbb{Z}_2)$ and $I_k \in M_k(\mathbb{Z}_2)$ is the identity matrix.

Parity check matrix

Definition

With the above notation, the matrix

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$$

is called the *parity check matrix* of the code.

Theorem

Consider a linear (n, k)-code with parity check matrix $H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$. Then a received vector $u \in \mathbb{Z}_2^n$ (or $u \in M_{n,1}(\mathbb{Z}_2)$) is a code vector if and only if $H \cdot [u]_{E'} = [0]_{E'}$.

Matrix representation - example

Example. *Determine the generator matrix and the parity check matrix of the* $(6,3)$*-code generated by the polynomial* $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$*, and characterize the code vectors.*

Solution. Note that $n = 6$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, i.e. $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2^3$ and $\mathbb{Z}_2^6$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \rightsquigarrow m = 1 \rightsquigarrow m \cdot X^{n-k} = X^3 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^3 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^3 \\ &\rightsquigarrow \boxed{110}\boxed{100} \rightsquigarrow (1,1,0,1,0,0) = \gamma(e_1). \end{aligned}$$

Similarly, $e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,1,0) = \gamma(e_2)$ and $e_3 = (0,0,1) \rightsquigarrow (1,1,1,0,0,1) = \gamma(e_3)$.

Matrix representation - example

- Hence $G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}$.
- The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_3 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

- $(u_1, u_2, u_3, u_4, u_5, u_6) \in \mathbb{Z}_2^6$ is a code word $\Leftrightarrow H \cdot [u]_{E'} = [0]_{E'}$

$$\Leftrightarrow \begin{cases} u_1 + u_4 + u_6 = 0 \\ u_2 + u_4 + u_5 + u_6 = 0 \\ u_3 + u_5 + u_6 = 0 \end{cases} \Leftrightarrow \begin{cases} u_1 = u_4 + u_6 \\ u_2 = u_4 + u_5 + u_6 \\ u_3 = u_5 + u_6 \end{cases}.$$

Error-correcting and decoding

A naive method:

- Given a received word, compute all Hamming distances to the code words.
  (Recall that the Hamming distance between two words of the same length is the number of positions in which they difer.)
- The code word closest to the received word will be assumed to be the most likely transmitted word.

Not practical!

Intermezzo: quotient vector spaces

Theorem

Let U be a K-vector space and let V be a subspace of U. For $u \in U$ we denote $u + V = \{u + v \mid v \in V\}$. Then

$$U/V = \{u + V \mid u \in U\}$$

is a vector space over K with respect to the addition and the scalar multiplication given by

$$\begin{aligned} (u_1 + V) + (u_2 + V) &= (u_1 + u_2) + V, \quad \forall u_1, u_2 \in U, \\ k \cdot (u + V) &= (k \cdot u) + V, \quad \forall k \in K, \forall u \in U. \end{aligned}$$

Then U/V is called a quotient vector space*, and $u + V$ $(u \in U)$ is called a* coset*.*

Coset leaders

Consider an (n, k)-code with encoding function $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ and denote $V = \operatorname{Im} \gamma$ (the subspace of code vectors).

- Start with a code vector $v \in V = \operatorname{Im} \gamma \leq \mathbb{Z}_2^n$, and assume that an error $e \in \mathbb{Z}_2^n$ occurs during transmission.
- Then the received vector is $u = v + e \in \mathbb{Z}_2^n$. The receiver determines the most likely transmitted vector by finding the most likely error pattern (called the *coset leader*)

$$e = u - v = u + v \in u + V.$$

- The coset leader will usually be the coset containing the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's in the error pattern are bunched together as much as possible.

Syndromes

Theorem

Consider a linear (n, k)-code with generator and parity check matrices $G \in M_{n,k}(\mathbb{Z}_2)$ and $H \in M_{n-k,n}(\mathbb{Z}_2)$ respectively. Let

$$\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n \text{ and } \eta : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$$

be the $\mathbb{Z}_2$-linear maps corresponding to G and H respectively. Then $V = \operatorname{Im}\gamma = \operatorname{Ker}\eta$.

If E' and E'' are the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^n$ and $\mathbb{Z}_2^{n-k}$ respectively, then $[\eta(u)]_{E''} = [\eta]_{E'E''} \cdot [u]_{E'} = H \cdot [u]_{E'}$.

Definition

With the above notation, the vector $\eta(u) \in \operatorname{Im}\eta \leq \mathbb{Z}_2^{n-k}$ (or $H \cdot u \in M_{n-k,1}(\mathbb{Z}_2)$) is called the *syndrome* of u.

The number of syndromes for an (n, k)-code is 2^{n-k}.

A general method for decoding

1. Calculate the syndrome of the received word.
2. Find the coset leader of the coset corresponding to the syndrome.
3. Subtract the coset leader from the received word to obtain the most likely transmitted word.
4. Drop the check digits to obtain the most likely message.

Decoding - examples

Example 1. Construct a table of coset leaders and syndromes for the $(6, 3)$-code with parity check matrix

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

and then decode the received words 011100 and 100011.

Solution.

- We have an (n, k)-code, where $n = 6$ and $k = 3$.
- The number of syndromes is $2^{n-k} = 2^3 = 8$.
- We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.

Decoding - examples

syndrome	coset leader
000	
001	
010	
011	
100	
101	
110	
111	

syndrome	coset leader
000	000000
001	001000
010	010000
011	000010
100	100000
101	000110
110	000100
111	000001

Decoding - examples

- The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.
- We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.
- We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

Decoding - examples

- The syndrome of $u = 000000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 100000$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 010000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 001000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000100$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 000010$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000001$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Decoding - examples

- For the last syndrome, namely 101, we try with 110000, 011000, 001100, 000110 or 000011. The correct one is $u = 000110$, because $H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.
- To decode $u_1 = 011100$, compute its syndrome $H \cdot [u_1] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.
  Its corresponding coset leader is $e_1 = 000110$.
  The most likely code vector is $v_1 = u_1 + e_1 = 011010$.
  Hence the most likely message is 010.
- To decode $u_2 = 100011$, compute its syndrome $H \cdot [u_2] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
  Its corresponding coset leader is $e_2 = 000000$.
  The most likely code vector is $v_2 = u_2 + e_2 = 100011$.
  Hence the most likely message is 011.

Decoding - examples

Example 2. Construct a table of coset leaders and syndromes for the $(7, 3)$-code generated by the polynomial

$$p = 1 + X + X^4 \in \mathbb{Z}_2[X].$$

Solution. Note that $n = 7$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, i.e. $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^7$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2^3$ and $\mathbb{Z}_2^7$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \rightsquigarrow m = 1 \rightsquigarrow m \cdot X^{n-k} = X^4 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^4 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^4 \\ &\rightsquigarrow \boxed{1100}\boxed{100} \rightsquigarrow (1,1,0,0,1,0,0) = \gamma(e_1). \end{aligned}$$

Similarly, $e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,0,1,0) = \gamma(e_2)$ and $e_3 = (0,0,1) \rightsquigarrow (0,0,1,1,0,0,1) = \gamma(e_3)$.

Decoding - examples

- Hence $G = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}$.
- The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_4 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Decoding - examples

- The number of syndromes is $2^{n-k} = 2^4 = 16$.
- We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.
- The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.
- We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.
- We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

Decoding - examples

After computations, we obtain the following table:

syndrome	coset leader
0000	0000000
0001	0001000
0010	0010000
0011	0000001
0100	0100000
0101	0000011
0110	0000010
0111	0100001

syndrome	coset leader
1000	1000000
1001	1001000
1010	0000110
1011	1000001
1100	0000100
1101	0001100
1110	0010100
1111	0000101

Course 12

3.7 Linear systems of equations

In this section we present some aspects of the theory of linear systems of equations which can be approached by using tools of Linear Algebra.

As usual, throughout this section K will be a field. When needed, we use superior indices to denote vectors in K^n and inferior indices to denote their components. For instance, $x^0 = (x_1^0, \ldots, x_n^0) \in K^n$.

Definition 3.7.1 Consider a *linear system* of m equations with n unknowns $x_1, \ldots, x_n$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \qquad (S)$$

where the coefficients are $a_{ij}, b_i \in K$ $(i = 1, \ldots, m,\ j = 1, \ldots, n)$.
If $b_1 = \cdots = b_m = 0$, then the system (S) is called *homogeneous* and is denoted by (S_0).
The matrix $A = (a_{ij}) \in M_{m,n}(K)$ is called the *matrix of the system* (S).
The matrix

$$\bar{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{pmatrix}$$

is called the *augmented* (or *extended*) *matrix of the system* (S).

Next we present two equivalent forms of a linear system of equations.
First, denote

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Then the systems (S) and (S_0) can be written as:

$$A \cdot x = b\,, \qquad (S)$$

$$A \cdot x = 0\,. \qquad (S_0)$$

Furthermore, we know that there exists a bijective correspondence between K-linear maps and matrices. Thus, since $A \in M_{m,n}(K)$, there exists $f_A \in \mathrm{Hom}_K(K^n, K^m)$ such that $[f_A]_{EE'} = A$, where E and E' are the canonical bases in K^n and K^m respectively.

Denoting $x = (x_1, \ldots, x_n) \in K^n$ and $b = (b_1, \ldots, b_m) \in K^m$, it follows that

$$[f_A(x)]_{E'} = [f_A]_{EE'} \cdot [x]_E = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = [b]_{E'}\,.$$

Hence $f_A(x) = b$. Thus, the systems (S) and (S_0) can be written as:

$$f_A(x) = b\,, \qquad (S)$$

$$f_A(x) = 0\,. \qquad (S_0)$$

Remark 3.7.2 (1) Thus, for a linear system of equations we have three equivalent forms, namely: the classical one with coefficients and unknowns, the one using matrices and the one using the corresponding linear map.

(2) We have denoted by x and b first column-matrices and then row-matrices to get nicer results, without using any transposed matrices.

Definition 3.7.3 An element $x^0 \in M_{n1}(K)$ ($x^0 \in K^n$) is called a:
(1) *(particular) solution* of (S) if $A \cdot x^0 = b$ (or equivalently $f_A(x^0) = b$).
(2) *(particular) solution* of (S_0) if $A \cdot x^0 = 0$ (or equivalently $f_A(x^0) = 0$).

Denote the sets of solutions of (S) and (S_0) by

$$S = \{x^0 \in M_{n1}(K) \mid A \cdot x^0 = b\} \quad \text{or} \quad S = \{x^0 \in K^n \mid f_A(x^0) = b\},$$

$$S_0 = \{x^0 \in M_{n1}(K) \mid A \cdot x^0 = 0\} \quad \text{or} \quad S_0 = \{x^0 \in K^n \mid f_A(x^0) = 0\}.$$

Theorem 3.7.4 *The set S_0 of solutions of the homogeneous linear system of equations (S_0) is a subspace of the canonical vector space K^n over K and*

$$\dim S_0 = n - \operatorname{rank}(A).$$

Proof. Since

$$S_0 = \{x^0 \in K^n \mid f_A(x^0) = 0\} = \operatorname{Ker} f_A$$

and the kernel of a linear map is always a subspace of the domain vector space, it follows that $S_0 \leq K^n$. Now by the First Dimension Theorem, it follows that

$$\dim S_0 = \dim(\operatorname{Ker} f_A) = \dim K^n - \dim(\operatorname{Im} f_A) = n - \operatorname{rank}(f_A) = n - \operatorname{rank}(A),$$

which finishes the proof. □

Theorem 3.7.5 *If $x^1 \in S$ is a particular solution of the system (S), then*

$$S = x^1 + S_0 = \{x^1 + x^0 \mid x^0 \in S_0\}.$$

Proof. Since $x^1 \in S$, we have $Ax^1 = b$. We prove the requested equality by double inclusion.
First, let $x^2 \in S$. Then

$$Ax^2 = b \Longrightarrow Ax^2 = Ax^1 \Longrightarrow A(x^2 - x^1) = 0 \Longrightarrow x^2 - x^1 \in S_0 \Longrightarrow x^2 \in x^1 + S_0.$$

Conversely, let $x^2 \in x^1 + S_0$. There exists $x^0 \in S_0$ such that $x^2 = x^1 + x^0$. Then:

$$Ax^2 = A(x^1 + x^0) = Ax^1 + Ax^0 = b + 0 = b,$$

and consequently $x^2 \in S$.
Therefore, $S = x^1 + S_0$. □

Remark 3.7.6 By Theorem 3.7.5, the general solution of the system (S) can be obtained by knowing the general solution of the homogeneous system (S_0) and a particular solution of (S).

In the sequel, we are going to see when a linear system of equations has a solution.

Definition 3.7.7 The system (S) is called *compatible* (or *consistent*) if it has at least one solution. A compatible system (S) is called *determinate* if it has a unique solution.

Remark 3.7.8 (1) The system (S) is compatible if and only if $\exists x^0 \in K^n$ such that $f_A(x^0) = b$ if and only if $b \in \operatorname{Im} f_A$.

(2) The system (S_0) is compatible if and only if $\exists x^0 \in K^n$ such that $f_A(x^0) = 0$ if and only if $0 \in \operatorname{Im} f_A$. But the last condition always holds, since $\operatorname{Im} f_A$ is a subspace of K^m. Hence any homogeneous linear system of equations is compatible, having at least the zero (trivial) solution.

Theorem 3.7.9 *The system (S_0) has a non-zero solution if and only if* $\operatorname{rank}(A) < n$.

Proof. By Theorem 3.7.4, we have

$$S_0 = \operatorname{Ker} f_A \neq \{0\} \iff \dim S_0 \neq 0 \iff n - \operatorname{rank}(A) \neq 0 \iff \operatorname{rank}(A) < n\,,$$

which proves the result. □

Corollary 3.7.10 *Let $A \in M_n(K)$. Then*

$$S_0 = \{0\} \iff \operatorname{rank}(A) = n \iff \det(A) \neq 0\,.$$

Definition 3.7.11 If $A \in M_n(K)$ and $\det(A) \neq 0$, then the system (S) is called a *Cramer system.*

Theorem 3.7.12 *A Cramer system $Ax = b$ has a unique solution. More precisely, its unique solution $(x_1, \ldots, x_n)$ is computed by*

$$x_i = \det(A)^{-1} \cdot d_i,$$

where d_i is the determinant obtained from $\det(A)$ by replacing its i^{th} column by the column b for every $i \in \{1, \ldots, n\}$.

Proof. The matrix of a Cramer system is an invertible matrix $A \in M_n(K)$. Then we deduce that $x = A^{-1}b$ is the unique solution. Moreover, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A^{-1} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \det(A)^{-1} \cdot A^* \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \det(A)^{-1} \cdot \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}.$$

Hence $x_i = \det(A)^{-1} \cdot d_i$ for every $i \in \{1, \ldots, n\}$. □

Corollary 3.7.13 *A homogeneous Cramer system has only the zero solution.*

Let us now give two classical compatibility theorems.

Theorem 3.7.14 (Kronecker-Capelli Theorem) *The system (S) is compatible if and only if* $\operatorname{rank}(\bar{A}) = \operatorname{rank}(A)$.

Proof. Let $(e_1, \ldots, e_n)$ be the canonical basis of the canonical vector space K^n over K and denote by $a^1, \ldots, a^n$ the columns of the matrix A. Then we have

$$\begin{aligned}(S) \text{ is compatible } &\iff \exists x^0 \in K^n : \ f_A(x^0) = b \iff b \in \operatorname{Im} f_A \\ &\iff b \in f_A(\langle e_1, \ldots, e_n\rangle) \iff b \in \langle f_A(e_1), \ldots, f_A(e_n)\rangle \\ &\iff b \in \langle a^1, \ldots, a^n\rangle \iff \langle a^1, \ldots, a^n, b\rangle = \langle a^1, \ldots, a^n\rangle \\ &\iff \dim\langle a^1, \ldots, a^n, b\rangle = \dim\langle a^1, \ldots, a^n\rangle \iff \operatorname{rank}(\bar{A}) = \operatorname{rank}(A)\,,\end{aligned}$$

which proves the result. □

Definition 3.7.15 A minor d_p of the matrix A is called a *principal determinant* if $d_p \neq 0$ and d_p has the order $\operatorname{rank}(A)$.
We call *characteristic determinants associated to a principal determinant* d_p of A the minors of the augmented matrix $\bar{A}$ obtained by completing the matrix of d_p with a column containing the corresponding constants b_i and a row containing the corresponding elements of a row of $\bar{A}$.

Now we give the second compatibility theorem.

Theorem 3.7.16 (Rouché Theorem) *The system (S) is compatible if and only if all the characteristic determinants associated to a principal determinant are zero.*

Proof. $\boxed{\Longrightarrow}$ Suppose that the system (S) is compatible. Then by Theorem 3.7.14, $\text{rank}(\bar{A}) = \text{rank}(A)$. Denote this rank by r. Then there exists a principal determinant d_p of order r. Since $r = \text{rank}(A)$, any determinant of order $r+1$ is zero and consequently any characteristic determinant associated to d_p is zero.

$\boxed{\Longleftarrow}$ Suppose that all the characteristic determinants associated to a principal determinant are zero. Denote $r = \text{rank}(A)$. Then $r \leq \text{rank}(\bar{A})$ and there exists a non-zero minor, actually a principal determinant, d_r of A. But d_r is also a minor of $\bar{A}$ of order r.

Now let d_{r+1} be a minor of $\bar{A}$ of order $r+1$. We have two possibilities, namely either d_{r+1} is a minor of $\bar{A}$ or d_{r+1} is just a minor of A. In the first case, d_{r+1} is a characteristic determinant associated to the principal determinant d_r, hence $d_{r+1} = 0$ by hypothesis. In the second case, we have $d_{r+1} = 0$, since $\text{rank}(A) = r$.

Thus, $\text{rank}(\bar{A}) = r = \text{rank}(A)$. Now by Theorem 3.7.14, (S) is compatible. □

3.8 Gauss method

In this section we briefly present a very useful practical method to solve linear systems of equations, called the *Gauss method* (or *Gaussian elimination*).

In the sequel, suppose that $m \leq n$, that is, we talk about systems with less equations than unknowns. In fact, this is the interesting case.

The **Gauss method** consists of the following steps:

(1) *Write the augmented matrix $\bar{A}$ of the system (S).*

(2) *Apply elementary operations on rows for $\bar{A}$ to get to an echelon form A'.*

(3) *Use the Kronecker-Capelli Theorem to decide if the system is compatible or not.*

(4) *If compatible, write and solve the system corresponding to the echelon form, starting with the last equation.*

Remark 3.8.1 (1) Actually, the Gauss method simulates working with equations. When we apply an elementary operation on the rows of $\bar{A}$, say multiply a row by a scalar and add it to another row, in fact we multiply an equation by a scalar and add it to another equation. That is why it is important to apply elementary operations only on rows, in order not to interchange the order of the unknowns.

(2) The initial system and the system corresponding to the echelon form are equivalent, that is, they have the same solutions. The great advantage is that the last system can be easily solved, starting with the last equation.

(3) The Gauss method includes checking compatibility, done by the Kronecker-Capelli Theorem.

(4) If the system is compatible, we have a principal determinant of order $r = \text{rank}(\bar{A}) = \text{rank}(A)$ and it is possible to continue the procedure on the matrix A' to get to a diagonal form having r elements on the principal diagonal and all the other elements zero. Then, when writing the equivalent system, in fact we directly get the solution. This completion of the Gauss method is called the *Gauss-Jordan method*.

Example 3.8.2 (a) Consider the system

$$\begin{cases} x + y - z = 2 \\ 3x + 2y - 2z = 6 \\ -x + y + z = 0 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A} = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix}.$$

Since $\operatorname{rank}(\bar{A}) = 3 = \operatorname{rank}(A)$, the system is determinate compatible. The equivalent system is

$$\begin{cases} x + y - z = 2 \\ -y + z = 0 \\ 2z = 2. \end{cases}$$

We immediately get the solution $x = 2$, $y = 1$, $z = 1$.

We could have got to the same solution by continuing with the Gauss-Jordan method. Indeed,

$$\bar{A} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

whence we immediately read the solution $x = 2$, $y = 1$, $z = 1$.

(b) Consider the system

$$\begin{cases} x + y + z = 0 \\ x + 4y + 10z = 3 \\ 2x + 3y + 5z = 1 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 4 & 10 & 3 \\ 2 & 3 & 5 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & 9 & 3 \\ 0 & 1 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 1 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\operatorname{rank}(\bar{A}) = 2 = \operatorname{rank}(A)$, the system is non-determinate compatible. The equivalent system is

$$\begin{cases} x + y + z = 0 \\ y + 3z = 1. \end{cases}$$

Then x and y are principal unknowns and z is a secondary unknown. We immediately get the solution

$$\begin{cases} x = 2z - 1 \\ y = 1 - 3z \\ z \in \mathbb{R}. \end{cases}$$

We could have got to the same solution by continuing with the Gauss-Jordan method. Indeed,

$$\bar{A} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The equivalent system is

$$\begin{cases} x - 2z = -1 \\ y + 3z = 1 \end{cases}$$

whence we get the solution

$$\begin{cases} x = 2z - 1 \\ y = 1 - 3z \\ z \in \mathbb{R}. \end{cases}$$

(c) Consider the system

$$\begin{cases} x+y+z=3 \\ x-y+z=1 \\ -2x+y-2z=-3 \\ x+z=4 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A}=\begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ -2 & 1 & -2 & -3 \\ 1 & 0 & 1 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 3 & 0 & 3 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\operatorname{rank}(\bar{A})=3$ and $\operatorname{rank}(A)=2$, the system is not compatible.

Remark 3.8.3 Following [Robbiano], let us analyze how many operations are required to solve a linear system of equations $Ax=b$ with $A\in M_n(K)$ invertible by the Gauss method. We assume that the operations of interchanging rows are negligible.

Let us first compute the cost of the reduction to a triangular echelon form having all elements on the principal diagonal equal to 1. We may assume that $a_{11}\neq 0$. We reduce a_{11} to 1 by dividing the first row of A by a_{11}. Then we produce zeros on the first column under the $(1,1)$-entry. For each of the $n-1$ rows we need n multiplications and n additions. Adding the corresponding operations for b, we have 1 more division, $n-1$ more multiplications and $n-1$ more additions. After finishing working with the first row, we move on to the second row, and so on til we get the required triangular form with elements 1 on the principal diagonal. Counting up the operations we have

- $n+(n-1)+\cdots+1$ divisions on A and n divisions on b;
- $n(n-1)+(n-1)(n-2)+\cdots+2\cdot 1$ multiplications on A and $(n-1)+\cdots+1$ multiplications on b;
- $n(n-1)+(n-1)(n-2)+\cdots+2\cdot 1$ additions on A and $(n-1)+\cdots+1$ additions on b.

So far we have

- $\frac{n(n+1)}{2}+n$ divisions;
- $\frac{n^3-n}{3}+\frac{n(n-1)}{2}$ multiplications;
- $\frac{n^3-n}{3}+\frac{n(n-1)}{2}$ additions.

Now let us compute the cost of substitutions in the reduced triangular system. From the last equation we already have the unknown x_n. For the substitution on the previous but last equation to find x_{n-1} we need 1 multiplication and 1 addition. Continuing the procedure, for the first equation to find x_1 we need $n-1$ multiplications and $n-1$ additions. Counting up the operations, we have

- $(n-1)+\cdots+1=\frac{n(n-1)}{2}$ multiplications;
- $(n-1)+\cdots+1=\frac{n(n-1)}{2}$ additions.

Adding up the numbers of operations from the above two stages, it turns out that one needs:

(1) $\frac{n(n+1)}{2}+n$ divisions;

(2) $\frac{n^3-n}{3}+n(n-1)$ multiplications;

(3) $\frac{n^3-n}{3}+n(n-1)$ additions.

Hence the order of magnitude is $\frac{2}{3}n^3$ operations.

Extra: LU decomposition and Gauss method (see [Crivei])

Extra: Simple authentication scheme

Let us consider the following simple authentication scheme from cryptography, following [Klein]. We denote by E the canonical basis of the canonical vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$.

- The password is a vector $v = (x_1, \ldots, x_n) \in \mathbb{Z}_2^n$.
- As a challenge, Computer sends a random vector $u = (u_1, \ldots, u_n) \in \mathbb{Z}_2^n$.
- As the response, Human sends back the dot-product vector

$$u \cdot v = u_1 x_1 + \cdots + u_n x_n \in \mathbb{Z}_2.$$

- The challenge-response interaction is repeated until Computer is convinced that Human knows password v.

Eve eavesdrops and learns m pairs (a_1, b_1), $\ldots$, (a_m, b_m) such that each b_i is the correct response to challenge a_i. For every $i \in \{1, \ldots, m\}$, denote $a_i = (a_{i1}, \ldots, a_{in})$.
Then the password $v = (x_1, \ldots, x_n)$ is a solution of the linear system of equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Once the rank of the matrix of the system reaches n, the solution is unique, and Eve can use the Gauss method to find it, obtaining the password.