September 18, 2017

Contents

3. Linear Transformations

3.1. **Definition and examples.** Let U, V be two vector spaces over the same field F (which for us is $\mathbb{R}$ or $\mathbb{C}$).

Definition 1. *A linear transformation $T : U \to V$ is a function which is linear, in the sense that it satisfies*

$$T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in U \tag{1}$$

and

$$T(c\mathbf{x}) = cT(\mathbf{x}) \quad \text{for all } \mathbf{x} \in U,\ c \in F \tag{2}$$

The properties $(1), (2)$ which express linearity, are often written in one formula as

$$T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y}) \quad \text{for all } \mathbf{x}, \mathbf{y} \in U,\ c, d \in F$$

Note that if T is linear, then also

$$T\left(\sum_{k=1}^{r} c_k \mathbf{x}_k\right) = \sum_{k=1}^{r} c_k T(\mathbf{x}_k) \quad \text{for all } \mathbf{x}_1, \ldots, \mathbf{x}_r \in U,\ c_1, \ldots c_r \in F$$

Notation: for linear transformations it is customary to denote simply $T\mathbf{x}$ rather than $T(\mathbf{x})$.

Note that a linear transformation takes the zero vector to the zero vector: $T\mathbf{0} = \mathbf{0}$. (Indeed, take $c = 0$ in (2).)

Definition 2. *A linear transformation $\phi : V \to F$, from a vector space to its scalar field, is called a linear functional.*

By contrast, linear transformations $T : U \to V$ with target space V not the scalar field are also called *linear maps*, or *linear operators*.

Examples.

1. The transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{x}) = 2\mathbf{x}$ is linear.

More generally, *scaling transformation* is linear: for some fixed scalar $\lambda \in \mathbb{R}$, let $T\mathbf{x} = \lambda\mathbf{x}$. In particular, $T\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$ is linear, and so is the identity transformation, $I\mathbf{x} = \mathbf{x}$.

And even more generally, *dilations*: $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(x_1, x_2, \ldots, x_n) = (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_n x_n)$ are linear.

The rotation of the plane by θ (rad) counterclockwise: $R : \mathbb{C} \to \mathbb{C}$, $R(z) = e^{i\theta} z$ is linear (over $F = \mathbb{C}$).

2. The *projection operator* $P : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $P(x_1, x_2, x_3) = (x_1, x_2)$ is a linear operator.

The projection $P : \mathbb{R}^3 \to \mathbb{R}$, $P(x_1, x_2, x_3) = x_2$ is a linear functional.

3. The translation transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, given by $T(\mathbf{x}) = \mathbf{x} + \mathbf{e}_1$ is not linear.

The usually called "linear function" $f(x) = ax + b$ is not a linear transformation. It is more correctly called an *affine* transformation. An affine transformation is a linear transformation followed by a translation.

4. *A few examples in infinite dimensions.* Denote by $\mathcal{P}$ the vector space of polynomials with coefficients in F.

o) Evaluation at some t_0: the functional $E_{t_0} : \mathcal{P} \to F$ defined by $E_{t_0}(p) = p(t_0)$ is linear.

a) The differentiation operator: $D : \mathcal{P} \to \mathcal{P}$, defined by $(Dp)(t) = p'(t)$ is linear.

b) The integration operator: $J : \mathcal{P} \to \mathcal{P}$, defined by $(Jp)(t) = \int_0^t p(s)ds$ is linear, and

c) so is the functional $I : \mathcal{P} \to F$, by $Ip = \int_0^1 p(s)ds$. Note that I is the composition $I = E_1 \circ J$.

3.2. **The matrix of a linear transformation.**

I. *A linear transformation is completely determined by its action on a basis.*

Indeed, let $T : U \to V$ be a linear transformation between two finite dimensional vector spaces U and V. Let $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of U. Then any $\mathbf{x} \in U$ can be uniquely written as

$$\mathbf{x} = x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n \quad \text{for some } x_1, \ldots, x_n \in F \tag{3}$$

and by the linearity of T we have

$$T\mathbf{x} = T(x_1\mathbf{u}_1 + \ldots + x_n\mathbf{u}_n) = x_1 T\mathbf{u}_1 + \ldots + x_n T\mathbf{u}_n \tag{4}$$

hence, once we give the values $T\mathbf{u}_1, \dots, T\mathbf{u}_n$ then the action of T is determined on all the vectors in U.

II. To determine $T\mathbf{u}_1, \dots, T\mathbf{u}_n$ we need to specify the representation of these vectors in a given basis of V. Let then $\mathcal{B}_V = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a basis of V, and denote by M_{ij} the coefficients of these representations:

$$T\mathbf{u}_j = \sum_{i=1}^{m} M_{ij}\mathbf{v}_i, \quad \text{for some } M_{ij} \in F, \quad \text{for all } j = 1, \dots, n$$

The matrix $M = M_T = [M_{ij}]_{i=1,\dots m, j=1,\dots n}$ is called **the matrix representation of the linear transformation T in the bases $\mathcal{B}_U$ and $\mathcal{B}_V$.**

Note that the matrix representation of a linear transformation depends not only on the basis chosen for U and V, but also on the order on which the vectors in each basis are enumerated. It would have been more precise to write the two basis $\mathcal{B}_U$, $\mathcal{B}_V$ as a multiplet, rather than a set.

Note. Let M_T is the matrix representation of the linear transformation T in the bases $\mathcal{B}_U, \mathcal{B}_V$. Any $\mathbf{x} \in U$ can be represented as (3), and denote $T\mathbf{x} = \mathbf{y} = y_1\mathbf{v}_1 + \ldots + y_m\mathbf{v}_m$. Then, organizing the coordinates of $\mathbf{x}, \mathbf{y}$ as columns, the action of T is matrix multiplication on the coordinate vector:

$$M_T \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

which can be seen after a direct verification: $T\mathbf{x} = \sum_j x_j T\mathbf{u}_j = \sum_j x_j \sum_i M_{ij}\mathbf{v}_i = \sum_i (\sum_j M_{ij}x_j)\mathbf{v}_i = \sum_i y_i \mathbf{v}_i$.

Note also the block representation of the matrix of T: if we organize the coordinates $M_{1j}, \dots, M_{mj}$ of $T\mathbf{u}_j$ as columns,

$$M_T = [T\mathbf{u}_1 \,|\, T\mathbf{u}_2 \,|\, \dots \,|\, T\mathbf{u}_n]$$

From now on we have to write the coordinates of vectors as columns.

Conversely, every matrix determines a linear transformation. Let M be an $m \times n$ matrix, and consider the transformation which multiplies vectors $\mathbf{x} \in \mathbb{R}^n$ by the matrix M

$$T : \mathbb{R}^n \to \mathbb{R}^m, \quad T(\mathbf{x}) = M\mathbf{x}$$

T is a linear transformation (check!), whose matrix in the standard bases is exactly M. Indeed,

$$T\mathbf{x} = M\mathbf{x} = M\left(\sum_{j=1}^{n} x_j\mathbf{e}_j\right) = \sum_{j=1}^{n} x_j\, M\mathbf{e}_j$$

where $M\mathbf{e}_j$ *is the column* j *of* M (convince yourselves!), so

$$M\mathbf{e}_j = \sum_{i=1}^{m} M_{ij}\,\mathbf{e}_i \tag{5}$$

Note that (5) can be used to give a block representation of M where each block is a column:

$$M = [M\mathbf{e}_1 \,|\, M\mathbf{e}_2 \,|\, \ldots \,|\, M\mathbf{e}_n]$$

Therefore **any linear transformation is a matrix multiplication**.

Example 1. Consider the standard basis $\{\mathbf{e}_1,\, \mathbf{e}_2\}$ for $\mathbb{C}^2$, where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{6}$$

For any $\mathbf{z} \in \mathbb{C}^2$ we have

$$\mathbf{z} \equiv \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = z_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + z_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = z_1\mathbf{e}_1 + z_2\mathbf{e}_2$$

$\mathbb{C}$ is a vector space over $F = \mathbb{C}$, of dimension 1. Denote its standard basis (consisting of $\{\mathbf{f}_1\}$ for $\mathbb{C}$, where $\mathbf{f}_1 = [1]$.

Define a linear transformation $T : \mathbb{C}^2 \to \mathbb{C}$ (note that here the scalars are $F = \mathbb{C}$, and $n = 2,\ m = 1$) be the linear transformation defined by its action on the basis of its domain: let $T(\mathbf{e}_1) = \mathbf{f}_1,\ T(\mathbf{e}_2) = -\mathbf{f}_1$. By linearity, its action on any element of $\mathbb{C}^2$ is completely determined:

$$T(\mathbf{z}) = T(z_1\mathbf{e}_1 + z_2\mathbf{e}_2) = z_1T(\mathbf{e}_1) + z_2T(\mathbf{e}_2) = z_1[1] + z_2[-1] = (z_1 - z_2)[1]$$

The same calculation in matrix notation: $[M_{ij}]_{i=1;j=1,2} = [1, -1]$ is the matrix of T, and $T(\mathbf{z}) = M\mathbf{z}$ (with $\mathbf{z}$ written as a vertical string).

Example 2. Consider the standard basis $p_j(t) = t^j,\ j = 0, 1, \ldots, n$ of $\mathcal{P}_n$, and let us find the matrix representation of the differentiation operator $D : \mathcal{P}_n \to \mathcal{P}_n$, $Dp = p'$.

Noting that $Dp_k = kp_{k-1}$, the matrix of D has the block form

$$M = M_D = [Dp_0 \,|\, Dp_1 \,|\, Dp_2 \,|\, \ldots \,|\, Dp_n] = [0 \,|\, 1 \,|\, 2p_1 \,|\, \ldots \,|\, np_{n-1}]]$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 2 & 0 & \ldots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & n \\ 0 & 0 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

3.3. **Operations with linear transformations and with their associated matrices.** The sum of two linear transformations $S, T : U \to V$ is defined like for any functions, as $(S+T)(\mathbf{u}) = S\mathbf{u} + T\mathbf{u}$ and so is multiplication by scalars, as $(cT)(\mathbf{u}) = cT\mathbf{u}$. It turns out that $S+T$ and cT are also linear transformations, and so is the composition:

Theorem 3. *Let U, V, W be vector spaces over F, and fix some bases B_U, B_V, B_W. Let $T, S : U \to V$ be linear transformations, with matrix representations M_T, M_S in the bases $\mathcal{B}_U$, $\mathcal{B}_V$.*

(i) Any linear combination $cS + dT : U \to V$ is a linear transformations, and its matrix representation is $cM_S + dM_T$ (in the bases $\mathcal{B}_U$, $\mathcal{B}_V$).

(ii) Let $R : V \to W$ be linear, with matrix M_R in the basis $\mathcal{B}_V$, $\mathcal{B}_W$. Then the composition $R \circ S : U \to W$, defined as (usually) by $(R \circ S)(\mathbf{x}) = R(S(\mathbf{x}))$ is a linear transformation, with matrix $M_R M_S$ (in the bases $\mathcal{B}_U, \mathcal{B}_W$).

The proof of the Theorem relies on immediate calculations; the calculations needed in part (ii) will be detailed below in §3.10.1 (and it shows why we multiply matrices using that strange rule...). □

For linear transformations we simply denote $R \circ S \equiv RS$.

It is easy to check that the operations with linear transformations satisfy the axioms of vector spaces, and *the set of all linear transformations from U to V, denoted $\mathcal{L}(U, V)$ is a linear space.* Moreover, for $U = V$, there is an extra operation, the composition of linear transformations, which behaves very nicely with respect to addition and scalar multiplication, in the sense that usual algebra rules apply: $(RS)T = R(ST)$, $R(cS + dT) = cRS + dRT$, $cRS + dTS = (cR + dT)S$ *except* that "multiplication" is not commutative, since, in general, RS does not equal SR (recall that, in general, for two functions f, g, $f \circ g \neq g \circ f$).

3.4. **Null space and range.** The following definitions and properties are valid in finite or infinite dimensions.

Let U, V be vector spaces over the same scalar field F.

Definition 4. *Let $T : U \to V$ be a linear transformation.*

The **null space** *(or* **kernel***) of T is*

$$\mathcal{N}(T) = \{\mathbf{x} \in U\,;\, T\mathbf{x} = \mathbf{0}\}$$

The **range** *of T is*

$$\mathcal{R}(T) = \{\mathbf{y} \in V \,|\, \mathbf{y} = T\mathbf{x} \text{ for some } \mathbf{x} \in U\}$$

Example. Let $T : \mathbb{R}^3 \to \mathbb{R}$ be the linear transformation defined by $T(x_1, x_2, x_3) = x_2 - 3x_1$ (this is a functional, to be more precise). Then $\mathcal{N}(T) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \,|\, x_2 - 3x_1 = 0\}$ (a plane) and $\mathcal{R}(T) = \mathbb{R}$.

Theorem 5. *Let $T : U \to V$ be a linear transformation. Then:*

(i) $\mathcal{N}(T)$ is a subspace of U.

(ii) $\mathcal{R}(T)$ is a subspace of V.

(iii) T is one to one if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$.

Why: the proof of (i)-(ii) is immediate.

To show (iii), first assume that T is one-to-one. If $\mathbf{x} \in \mathcal{N}(T)$ then $T\mathbf{x} = \mathbf{0} = T\mathbf{0}$ hence $x = \mathbf{0}$ and therefore the only element of $\mathcal{N}(T)$ is $\mathbf{0}$. Conversely, assume that $\mathcal{N}(T) = \{\mathbf{0}\}$. If $\mathbf{x}, \mathbf{y} \in U$ are so that $T\mathbf{x} = T\mathbf{y}$ then $T(\mathbf{x}-\mathbf{y}) = \mathbf{0}$ (because T is linear) which means that $\mathbf{x}-\mathbf{y} \in \mathcal{N}(T)$, hence $\mathbf{x}-\mathbf{y} = \mathbf{0}$ which implies that T is one-to-one. $\square$

One of the key features of linear transformations is that if a (linear) property is valid at $\mathbf{0}$ then it is valid at any point. Theorem 5 (iii) illustrates this principle: if T takes the value $\mathbf{0}$ only once, then, if T takes some value, then it takes it only once.

Some buzz-words: the **rank** of T is

$$rank(T) \equiv \dim \mathcal{R}(T)$$

and the **nullity** of T is

$$nullity(T) \equiv \dim \mathcal{N}(T)$$

3.5. **Column space and rank of a matrix.** Recall that matrices are convenient ways to represent linear transformations (once bases are chosen). It is then useful to transcribe the notions of null space, nullity, range, and rank for matrices.

Let M be an $n \times m$ matrix with entries in F, and T its associated linear transformation $T : F^n \to F^m$, $T_M\mathbf{x} = M\mathbf{x}$. Recall that M is the matrix of T_M associated to the standard bases.

Since vectors in $\mathcal{R}(T_M)$ have the form

$$M\mathbf{x} = M(\sum_j x_j\mathbf{e}_j) = \sum_j x_j M\mathbf{e}_j \qquad (*)$$

we see that

$$\mathcal{R}(T_M) = Sp(M\mathbf{e}_1, \ldots, M\mathbf{e}_n) \qquad (7)$$

which is the space spanned by the columns of M, called the **column space**, and which we will denote (by abuse of notation) by $\mathcal{R}(M)$.

A useful block notation for the matrix M is

$$M = [M\mathbf{e}_1 \,|M\mathbf{e}_2\,| \ldots \,|M\mathbf{e}_n]$$

and this block writing is compatible with operations with matrices if we write the coordinates of vectors vertically:

$$M\mathbf{x} = [M\mathbf{e}_1 \,|M\mathbf{e}_2\,| \ldots \,|M\mathbf{e}_n] \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix}$$

which by matrix multiplication equals (*), noting that $M\mathbf{e}_j x_j = x_j M\mathbf{e}_j$ since x_j is a scalar.

The $rank(T_M)$, is, by definition, the dimension of its range. It is natural to define the rank of a matrix as

$$rank(M) = rank(T_M) = \dim \text{of the column space of } M$$

Proposition 6. *If* $\dim Sp(\mathbf{x}_1, \ldots, \mathbf{x}_k) = r > 0$ *then there are* r *vectors independent among* $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

Here is an orderly procedure to choose them. If $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent then $r = k$ and we found our basis. Otherwise one of the $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is a linear combination of the others, say $\mathbf{x}_k$. Then $r < k$ and we disregard $\mathbf{x}_k$ since $Sp(\mathbf{x}_1, \ldots, \mathbf{x}_k) = Sp(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1})$. We continue: if $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$ are linearly independent then $r = k-1$ and we found our basis. Otherwise one of the $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$ is a linear combination of the others, say $\mathbf{x}_{k-1}$. Then $r < k-1$ and we disregard $\mathbf{x}_{k-1}$ etc. $\Box$.

Let $r = rank\, M$. Then among the columns $M\mathbf{e}_j$ of M there are r linearly independent ones, by Proposition 6; this is called a *basis column* and is a basis for $\mathcal{R}(T_M)$.

3.6. **Rank, minors, the dimension of the span of a finite set.** In Sec. 2.4 we found a test on how to decide if k vectors are linearly dependent or not. We are now asking even more, how to find the maximal number of independent vectors among them (this is the dimension of their span).

Let $M = [M_{ij}]_{i=1,\ldots n, j=1\ldots,k}$ be an $n \times k$ matrix. This could be the matrix of a linear transformation T, and then the rank of M is $\dim \mathcal{R}(T_M)$. Or, the matrix M may be constructed so that its columns represent the coordinates of k vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in a chosen basis $\mathbf{v}_1, \ldots \mathbf{v}_n$: $\mathbf{x}_j = \sum_i M_{ij}\mathbf{v}_i$; in this case the rank of M is $\dim Sp(\mathbf{x}_1, \ldots, \mathbf{x}_k)$. In this section we find a practical way to determine the rank of a matrix.

Recall the following *properties of determinants* from Chapter 2:

1) The value of a determinant equals the value of the determinant obtained by turning its lines into columns ($\det A = \det A^T$).

2) A determinant is zero if and only if some column is a linear combination of the others (its columns are linearly dependent).

3) A determinant is zero if and only if some row is a linear combination of the other rows (its rows are linearly dependent).

Definition. *Let* M *be an* $n \times k$ *matrix, and* $p \in \mathbb{Z}_+,\ p \leq \min\{k, n\}$. *Delete any* $n - p$ *rows and* $k - p$ *columns of* M*; we are left with a* $p \times p$ *sub-matrix. Its determinant is called a* **minor** *of order* p.

Of course, there are many minors of any order.

In the following we assume that M is not the zero matrix.

Remark 3. Denote $r = rank\, M$. Then any collection of $p \geq r + 1$ columns of M are linearly dependent, therefore any minor of order $\geq r + 1$ is zero.

Then by Proposition 6, *the rank of a matrix is the largest number* r *for which there is a nonzero minor of order* r *of* M.

Examples.
(i) Determine the dimension of $Sp(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ where $\mathbf{x}_1 = (2, 3, 1, 4)^T$, $\mathbf{x}_2 = (0, -2, 1, 3)^T$, $\mathbf{x}_3 = (2, 5, 0, 1)^T$.
Solution. We form the matrix having as columns the coordinates of these vectors:

$$M = [\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \mathbf{x}_3] = \begin{bmatrix} 2 & 0 & 2 \\ 3 & -2 & 5 \\ 1 & 1 & 0 \\ 4 & 3 & 1 \end{bmatrix} \tag{8}$$

The rank of M can be at most 3. The matrix has 4 minors of order 3 (obtained by deleting one of the four rows); a direct calculation shows that they are all zero, hence the rank of M is less than 3. Now we look at the minors of order 2, and we see that the upper left 2 by 2 submatrix has nonzero determinant, hence the rank of M is 2, and so is dim $Sp(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$. Furthermore, this also shows that $\mathbf{x}_1$, $\mathbf{x}_2$ are linearly independent hence they form a basis for $Sp(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$.

(ii) Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ given by $T\mathbf{x} = M\mathbf{x}$ where M is the matrix in (8). Find the dimension of the set of all vector $\mathbf{y}$ which can be written as $\mathbf{y} = T\mathbf{x}$ for some $\mathbf{x}$.
Solution. Of course, this is $\dim \mathcal{R}(T) = rank(T) = rank(M)$ which we found to be 2.

3.7. **Systems of linear equations, column and null spaces.** Note:

(0) A homogeneous system $M\mathbf{x} = \mathbf{0}$ has the set of solutions equal to $\mathcal{N}(M)$.

(1) The system $M\mathbf{x} = \mathbf{b}$ is *solvable* (it has solutions) if and only if $\mathbf{b} \in \mathcal{R}(M)$.

To check if $\mathbf{b}$ belongs to $\mathcal{R}(M)$ (the column space of M), one can check the rank of the so-called augmented matrix, obtained from M by adding $\mathbf{b}$ as one more column, $M_{aug} = [M \,|\, \mathbf{b}]$. There are two possibilities: either
1^o $\mathbf{b} \in \mathcal{R}(M)$ hence $\mathbf{b}$ is linearly dependent of the columns on M, therefore $rank\, M_{aug} = rank\, M$, or
2^o $\mathbf{b} \notin \mathcal{R}(M)$ hence $\mathbf{b}$ is linearly independent of the columns on M, and therefore $rank\, M_{aug} = 1 + rank\, M$.

(2) Suppose the system $M\mathbf{x} = \mathbf{b}$ is solvable. Then its solution is unique if and only if $\mathcal{N}(M) = \{\mathbf{0}\}$ (why?).

(3) *If $\mathbf{x}_0$ is a particular solution of $M\mathbf{x} = \mathbf{b}$, then the set of all solutions is*

$$\mathbf{x}_0 + \mathcal{N}(M) = \{\mathbf{x}_0 + \mathbf{z} \,|\, \mathbf{z} \in \mathcal{N}(M)\}$$

In other words:

(9) General solution=Particular solution + Homogeneous solutions

Indeed, on one hand it is clear that all the vectors in $\mathbf{x}_0 + \mathcal{N}(M)$ solve the same equation: $M(\mathbf{x}_0 + \mathbf{z}) = M\mathbf{x}_0 = \mathbf{b}$. Conversely, any two solutions $\mathbf{x}_1$, $\mathbf{x}_2$ of the same equation: $M\mathbf{x}_1 = \mathbf{b}$, $M\mathbf{x}_2 = \mathbf{b}$, differ by an element $\mathbf{x}_1 - \mathbf{x}_2 \in \mathcal{N}(M)$.

(4) *A solvable system* $M\mathbf{x} = \mathbf{b}$ *has a unique solution if and only if the homogeneous equation* $M\mathbf{x} = \mathbf{0}$ *has a unique solution (which is, of course,* $\mathbf{x} = \mathbf{0}$*).*

This is clear by (3) and property (iv) of Theorem 5.

3.7.1. *An example.* A theoretical way (not computationally efficient, usually) to solve a linear system $M\mathbf{x} = \mathbf{b}$ is illustrated here for M given by (8) and $\mathbf{b} = (2\beta, \alpha, 0, \beta)^T$. To establish if the system is solvable we find the rank of the augmented matrix

$$M_{aug} = \left[\begin{array}{ccc|c} 2 & 0 & 2 & 2\beta \\ 3 & -2 & 5 & \alpha \\ 1 & 1 & 0 & 0 \\ 4 & 3 & 1 & \beta \end{array}\right] \tag{10}$$

The largest minor of M_{aug} has order 4, and its determinant is zero (no calculation is needed, we saw that the first three columns are linearly dependent). The minors of order 3 are obtained by deleting one row and one column. Deleting the first column we obtain four minors:

$$\det \begin{bmatrix} -2 & 5 & \alpha \\ 1 & 0 & 0 \\ 3 & 1 & \beta \end{bmatrix} = \alpha - 5\,\beta, \quad \det \begin{bmatrix} 0 & 2 & 2\beta \\ 1 & 0 & 0 \\ 3 & 1 & \beta \end{bmatrix} = 0$$

$$\det \begin{bmatrix} 0 & 2 & 2\beta \\ -2 & 5 & \alpha \\ 3 & 1 & \beta \end{bmatrix} = 6\,\alpha - 30\,\beta, \quad \det \begin{bmatrix} 0 & 2 & 2\beta \\ -2 & 5 & \alpha \\ 1 & 0 & 0 \end{bmatrix} = 2\,\alpha - 10\,\beta$$

If $\alpha \neq 5\,\beta$ then $rank\, M_{aug} = 1 + rank\, M$ and the system is not solvable.

If $\alpha = 5\,\beta$ all the other minors of oder 3 are zero as well[1], therefore $rank\, M_{aug} = 2 = rank\, M$ and the system is solvable.

Let $\alpha = 5\,\beta$, so that the system is solvable. To solve, identify one nonzero minor of maximal order. Keep the the equations corresponding to the lines present in the minor, and disregard the others (as they are linear combinations of the former ones). We then keep only the variables corresponding the the columns present in the minor, and take the others to the left side of the equation; these variables will be free parameters. In our example, the

[1]Calculation of these minors can be avoided by noting that each of the first three columns is a linear combination of the other two.

system is:

$$\begin{array}{llll} 2x_1 & & +2x_3 & = 2\beta \\ 3x_1 & -2x_2 & +5x_3 & = 5\,\beta \\ x_1 & +x_2 & & = 0 \\ 4x_1 & +3x_2 & +x_3 & = \beta \end{array}$$

Say we identify the top left 2×2 minor. We keep only the first two equations and disregard the others. (To satisfy our need to be sure that the last equations are just linear combinations of the first two, hence are redundant, we can see that the four equations, i.e. the four rows of the augmented matrix, satisfy the linear dependence: $A\,row_1 + B\,row_2 + C row_3 + D\,row_4 = 0$ for $C = 17\,B + 6\,A, D = -2\,A - 5\,B$ and A, B arbitrary.)

So we rewrite the system as

$$\begin{array}{llll} 2x_1 & & = 2\beta & -2x_3 \\ 3x_1 & -2x_2 & = 5\,\beta & -5x_3 \end{array}$$

where x_3 is arbitrary and it is treated as a parameter. The solution is

$$x_1 = \beta - t,\ x_2 = -\beta + t,\ x_3 = t,\ \text{with } t \text{ arbitrary}$$

or

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \beta \\ -\beta \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \tag{11}$$

which illustrates (9).

For numerical calculation, Gauss elimination is certainly more numerically efficient, see also §3.11.4.

3.8. **Dimensionality properties.**

Theorem 7. *Let $T : U \to V$ be a linear transformation between two finite dimensional vector spaces U and V. Let $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of U. Then:*

(i) $\mathcal{R}(T) = Sp\,(T\mathbf{u}_1, \ldots, T\mathbf{u}_n)$.

(ii) T is one-to-one if and only if $T\mathbf{u}_1, \ldots, T\mathbf{u}_n$ are linearly independent.

(iii) We have

$$\dim U = \dim \mathcal{R}(T) + \dim \mathcal{N}(T)$$

(in other words, $rank(T) + nullity(T) = n$).

Remarks.

1. Property (i) has already been established for standard bases, see (7).

1. Property (iii) can be interpreted intuitively as: when the operator T acts on n dimensions, some directions collapse to zero ($\dim\mathcal{N}(T)$ of them), and the rest span the range (their number is $\dim \mathcal{R}(T)$).

2. In the particular case of $T = T_M$ where $T(\mathbf{x}) = M\mathbf{x}$ as in §3.5:

∘ Property (i) has been established by (7).

∘ Property (ii) states that:

Corollary 8. *A solvable system $M\mathbf{x} = \mathbf{b}$ has a unique solution if and only if the columns of M are independent.*

If M is an $m \times n$ matrix, its columns are independent if and only if $n \leq m$ (the number of columns does not exceed the number of rows) and the rank$M = n$ (maximal rank).

Proof of Theorem 7.

Note that (i) is immediately visible from (3), (4).

For property (ii), note that $\sum_j c_j T\mathbf{u}_j = \mathbf{0}$ is equivalent to $\sum c_j\mathbf{u}_j \in \mathcal{N}(T)$. Now, T is one-to-one if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$ (by Theorem 5 (iii)). If $\mathcal{N}(T) = \{\mathbf{0}\}$ then $\sum c_j\mathbf{u}_j = \mathbf{0}$ which implies all $c_j = 0$ hence $T\mathbf{u}_1, \ldots, T\mathbf{u}_n$ are independent. Conversely, assume $T\mathbf{u}_1, \ldots, T\mathbf{u}_n$ are independent and let $\mathbf{x} \in \mathcal{N}(T)$. Then $\mathbf{x} = \sum_j c_j\mathbf{u}_j$ for some $c_j's$. Then $\mathbf{0} - T\mathbf{x} = \sum_j c_j T\mathbf{u}_j$ hence all $c_j = 0$ and therefore $\mathbf{x} = \mathbf{0}$, showing that $\mathcal{N}(T) = \{\mathbf{0}\}$.

Property (iii) is proved as follows.

Let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be a basis for $\mathcal{R}(T)$. Therefore, $\mathbf{v}_j = T(\mathbf{x}_j)$ for some $\mathbf{x}_j \in U$, for all $j = 1, \ldots, r$.

Now let $\mathbf{z}_1, \ldots \mathbf{z}_k$ be a basis for $\mathcal{N}(T)$. Property (iii) is proved if we show that $B = \{\mathbf{z}_1, \ldots \mathbf{z}_k, \mathbf{x}_1, \ldots \mathbf{x}_r\}$ is a basis for U.

To show linear independence, consider a linear combination

$$c_1\mathbf{z}_1 + \ldots + c_k\mathbf{z}_k + d_1\mathbf{x}_1 + \ldots + d_r\mathbf{x}_r = \mathbf{0} \tag{12}$$

to which we apply T on both sides, and it gives

$$T(c_1\mathbf{z}_1 + \ldots + c_k\mathbf{z}_k + d_1\mathbf{x}_1 + \ldots + d_r\mathbf{x}_r) = T\mathbf{0}$$

and by linearity

$$c_1T\mathbf{z}_1 + \ldots + c_kT\mathbf{z}_k + d_1T\mathbf{x}_1 + \ldots + d_rT\mathbf{x}_r = \mathbf{0}$$

and since $T\mathbf{z}_j = \mathbf{0}$, and $\mathbf{v}_j = T\mathbf{x}_j$

$$d_1\mathbf{v}_1 + \ldots + d_r\mathbf{v}_r = \mathbf{0}$$

which imply all $d_j = 0$ since $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent. Then (12) becomes

$$c_1\mathbf{z}_1 + \ldots + c_k\mathbf{z}_k = \mathbf{0}$$

which implies all $c_j = 0$ since $\mathbf{z}_1, \ldots \mathbf{z}_k$ are linearly independent. In conclusion B is a linearly independent set.

To complete the argument we need to show that $Sp(B) = U$. For any $\mathbf{u} \in U$ we have

$$T\mathbf{u} = d_1\mathbf{v}_1 + \ldots + d_r\mathbf{v}_r$$

for some scalars $d_1, \ldots d_r$. Note that

$$T(\mathbf{u} - d_1\mathbf{x}_1 - \ldots - d_r\mathbf{x}_r) = T\mathbf{u} - d_1\mathbf{v}_1 - \ldots - d_r\mathbf{v}_r = \mathbf{0}$$

therefore $\mathbf{u} - d_1\mathbf{x}_1 - \ldots - d_r\mathbf{x}_r \in \mathcal{N}(T)$, so $\mathbf{u} - d_1\mathbf{x}_1 - \ldots - d_r\mathbf{x}_r = c_1\mathbf{z}_1 + \ldots + c_k\mathbf{z}_k$ which shows that $\mathbf{u} \in Sp(B)$. □

3.8.1. *More consequences of the Dimensionality Theorem 7 (iii).* Let U, V be *finite dimensional* vector spaces, and $T : U \to V$ be a linear transformation.

1. Suppose $\dim U = \dim V$. Then T *is one to one* $\iff$ T *is onto*

For systems this means that for a given square matrix M we have: the homogeneous equation $M\mathbf{x} = \mathbf{0}$ has only the zero solution if and only if all $M\mathbf{x} = \mathbf{b}$ are solvable.

Reformulated as: *either all* $M\mathbf{x} = \mathbf{b}$ *are solvable, or* $M\mathbf{x} = \mathbf{0}$ *has nonzero solutions*, it is the celebrated **Fredholm's alternative**.

To prove **1.**, note that: T is one to one $\Leftrightarrow$ $\dim\mathcal{N}(T) = 0$ (and by Theorem 7 (iii)) $\Leftrightarrow$ $\dim \mathcal{R}(T) = \dim V$ which means $\mathcal{R}(T) = V$.

2. If $\dim U > \dim V$ then T is not one to one.

Indeed, $\dim \mathcal{N}(T) = \dim U - \dim \mathcal{R}(T) \geq \dim U - \dim V > 0$.

3. If $\dim U < \dim V$ then T is not onto.

Indeed, $\dim \mathcal{R}(T) = \dim U - \dim \mathcal{N}(T) \leq \dim U < \dim V$.

3.9. **Invertible transformations, isomorphisms.**

3.9.1. *The inverse function.* Recall that if $f : A \to B$ is a function which is onto to one and onto, then the function f has an inverse, denoted f^{-1} defined as $f^{-1} : B \to A$ by $f(x) = y \;\Leftrightarrow\; x = f^{-1}(y)$.

Recall that the compositions $f \circ f^{-1}$ and $f^{-1} \circ f$ equal the identity functions:

$$(f^{-1} \circ f)(x) = f^{-1}\left(f(x)\right) = x, \quad (f \circ f^{-1})(y) = f\left(f^{-1}(y)\right) = y$$

Examples:
1) let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 5x$. Then $f^{-1} : \mathbb{R} \to \mathbb{R}$, $f^{-1}(x) = x/5$;
2) let $g : \mathbb{R} \to \mathbb{R}$, $f(x) = x + 3$. Then $f^{-1} : \mathbb{R} \to \mathbb{R}$, $f^{-1}(x) = x - 3$;
3) let $h : (-\infty, 0] \to [0, \infty)$ by $h(x) = x^2$ then $h^{-1} : [0, \infty) \to (-\infty, 0]$, $h^{-1}(x) = -\sqrt{x}$.

3.9.2. *The inverse of a linear transformation.* Let $T : U \to V$ be a linear transformation between two vector spaces U and V. If T is onto to one and onto, then the function T has an inverse T^{-1}, $T^{-1} : V \to U$.

Exercise. Show that the inverse of a linear transformation is also a linear transformation.

Definition 9. *A linear transformation* $T : U \to V$ *which is onto to one and onto is called an* **isomorphism** *of vector spaces, and* U *and* V *are called* **isomorphic vector spaces.**

Whenever two vector spaces are isomorphic[2] and $T : U \to V$ is an isomorphism, then any property of U that can be written using the vector space operations of U can be translated, using the isomorphism T, as a translation machine, into a similar property for V.

The following theorem shows that all the finite dimensional vector spaces are essentially $\mathbb{R}^n$ or $\mathbb{C}^n$:

Theorem 10.
Any vector space over $\mathbb{R}$ of dimension n is isomorphic to $\mathbb{R}^n$.
Any vector space over $\mathbb{C}$ of dimension n is isomorphic to $\mathbb{C}^n$.

Indeed, let U be a vector space of dimension n over F (where F is $\mathbb{R}$ or $\mathbb{C}$). Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be a basis for U. Define $T : U \to F^n$ by $T(\mathbf{u}_j) = \mathbf{e}_j$ for $j = 1, \ldots n$ and then extended by linearity to all the vectors in U. Clearly T is onto, therefore it is also one to one, hence it is an isomorphism. □

Example.

1. As a vector space, $\mathcal{P}_n$ is essentially $\mathbb{R}^{n+1}$.

3.10. **Change of basis, similar matrices.** Let $T : U \to V$ be a linear transformation between two finite dimensional vector spaces U and V. Let $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of U, and $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a basis of V. Recall (see §3.2) that T is completely determined by the vectors $T\mathbf{u}_1, \ldots, T\mathbf{u}_n$, and these vectors, in turn, are completely determined by their expansion in the basis of V:

$$T(\mathbf{u}_j) = \sum_{i=1}^{m} M_{ij}\mathbf{v}_i, \quad \text{for some } M_{ij} \in F, \quad \text{for all } j = 1, \ldots, n$$

and the matrix $M = [M_{ij}]_{i=1,\ldots m, j=1,\ldots n}$ is the matrix representation of the linear transformation T in the bases $\mathcal{B}_U$ and $\mathcal{B}_V$.

Example. Consider the identity transformation $I : U \to U,\ I\mathbf{x} = \mathbf{x}$. Then its matrix representation in the bases $\mathcal{B}_U, \mathcal{B}_U$ is the identity matrix I (the diagonal matrix with 1 on its diagonal).

3.10.1. *The matrix representation of a composition is the product of the matrix representations.* We stated without proof Theorem 3(ii):

Proposition 11. *Let U, V, W be vector spaces over F (the same for all three spaces). Let $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a basis of U, $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ be a basis of V and $\mathcal{B}_W = \{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ be a basis of W.*

Let $T : U \to V$ and $S : V \to W$ be linear transformations. Denote by M_T the matrix representation of T in the basis $\mathcal{B}_U, \mathcal{B}_V$, and by M_S the matrix representation of S in the basis $\mathcal{B}_V, \mathcal{B}_W$. Composing the two maps

[2] The prefix *iso-* comes from the Greek word *isos*, which means "equal", "same". Combined with the Greek word *morphos*, meaning "shape", or "structure", then *isomorphic* means "having the same shape or structure".

$$\begin{array}{ccccc} U & \stackrel{T}{\longrightarrow} & V & \stackrel{S}{\longrightarrow} & W \\ \mathcal{B}_U & & \mathcal{B}_V & & \mathcal{B}_W \end{array}$$

we obtain the linear transformation $ST\mathbf{v} = S(T\mathbf{v})$*:*

$$\begin{array}{ccc} U & \stackrel{ST}{\longrightarrow} & W \\ \mathcal{B}_U & & \mathcal{B}_W \end{array}$$

Then the matrix representation of $ST : U \to W$ *in the basis* $\mathcal{B}_U, \mathcal{B}_W$ *is the product of the two matrices* $M_S M_T$*:*

$$M_{ST} = M_S M_T$$

Proof. We have

$$T\mathbf{u}_j = \sum_{i=1}^{m} M_{T,ij}\mathbf{v}_i \text{ for all } j = 1, \ldots, n \tag{13}$$

and

$$S\mathbf{v}_i = \sum_{k=1}^{p} M_{S,ki}\mathbf{w}_k, \quad \text{for all } i = 1, \ldots, m$$

Therefore

$$(ST)\mathbf{u}_j = S\left(T\mathbf{u}_j\right) = S\left(\sum_{i=1}^{m} M_{T,ij}\mathbf{v}_i\right) = \sum_{i=1}^{m} M_{T,ij}\, S\mathbf{v}_i$$

$$= \sum_{i=1}^{m} M_{T,ij}\left(\sum_{k=1}^{m} M_{S,ki}\mathbf{w}_k\right) = \sum_{k=1}^{m}\left(\sum_{i=1}^{m} M_{T,ij}\, M_{S,ki}\right)\mathbf{w}_k$$

and noting that the coefficient of $\mathbf{w}_k$ is the (k, j) entry of the matrix $M_S M_T$ the proof is complete. □

Recall the formula for the product of two matrices A and B: the element of AB in the position kj is obtained by multiplying elements on line k of A by elements on column j of B, hence $(AB)_{kj} = \sum_i A_{ki}B_{ij}$.

Exercise. Show that if $T : U \to V$ is invertible (recall that necessarily U and V must have the same dimension), and if M_T is its matrix representation in the basis $\mathcal{B}_U, \mathcal{B}_V$ then the matrix representation of T^{-1} in the basis $\mathcal{B}_V, \mathcal{B}_U$ is the inverse of the matrix M_T: $M_{T^{-1}} = M_T^{-1}$.

3.10.2. *The transition matrix.* Suppose the vector space U has a basis $\mathcal{B}_U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Let $\tilde{\mathcal{B}}_U = \{\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_n\}$ be another basis of U. Each $\tilde{\mathbf{u}}_j$ is a linear combination of the vectors in $\mathcal{B}_U$: there exist scalars S_{ij} so that

$$\tilde{\mathbf{u}}_j = \sum_{i=1}^{n} S_{ij}\mathbf{u}_i, \quad \text{for all } j = 1, \ldots, n \tag{14}$$

The matrix $S = [S_{ij}]_{i,j=1,\ldots,n}$ is called the **transition matrix** from the basis $\mathcal{B}_U$ to the basis $\tilde{\mathcal{B}}_U$.

Note: S is the matrix of the identity transformation $I : U \to U$, $I\mathbf{x} = \mathbf{x}$ corresponding to basis $\tilde{\mathcal{B}}_U$ on the domain, and $\mathcal{B}_U$ on the codomain:

$$\begin{array}{ccc} U & \xrightarrow{I} & U \\ {\scriptstyle \tilde{\mathcal{B}}_U} & & {\scriptstyle \mathcal{B}_U} \end{array} \tag{15}$$

A practical way of calculating the transition matrix when the vectors of the two bases are specified by giving their coordinates in some common basis (usually the standard basis of $\mathbb{R}^n$ or $\mathbb{C}^n$). Then we can organize each basis as a matrix, whose columns has the coordinates of the vectors, let us denote them as

$$\mathbf{u} = [\mathbf{u}_1 \,|\, \ldots |\mathbf{u}_n], \text{ respectively } \tilde{\mathbf{u}} = [\tilde{\mathbf{u}}_1 \,|\, \ldots |\tilde{\mathbf{u}}_n]$$

We then have

$$\tilde{\mathbf{u}} = \mathbf{u}S \tag{16}$$

which is easy to check by an immediate (and careful) calculation. Indeed, first note that k'th component of $\mathbf{u}_i$ is the element $\mathbf{u}_{ki}$ of the matrix $\mathbf{u}$, and similarly, k'th component of $\tilde{\mathbf{u}}_j$ is the element $\tilde{\mathbf{u}}_{kj}$ of the matrix $\tilde{\mathbf{u}}$. Looking at the k'th component in (14) we see that $\tilde{\mathbf{u}}_{kj} = \sum_{i=1}^n S_{ij}\mathbf{u}_{ki}$ which is the k, j component of the product $\mathbf{u}S$.

From (16) we obtain the practical formula $S = \mathbf{u}^{-1}\,\tilde{\mathbf{u}}$.

The transition matrix from the basis $\tilde{\mathcal{B}}_U$ to the basis $\mathcal{B}_U$ is, of course, S^{-1} (why?).

3.10.3. *How coordinates of vectors change upon a change of basis.* Suppose a vector $\mathbf{x} \in U$ has coordinates $(x_1, \ldots, x_n)^T$ in the basis $\mathcal{B}_U$, i.e. $\mathbf{x} = \sum_j x_j\mathbf{u}_j$. What are the coordinates $(\tilde{x}_1, \ldots, \tilde{x}_n)^T$ of $\mathbf{x}$ the basis $\tilde{\mathcal{B}}_U$?

We calculate: $\mathbf{x} = \sum_j x_j\mathbf{u}_j = \sum_j x_j \sum_i S^{-1}_{ij}\mathbf{u}'_i = \sum_i \left(\sum_j S^{-1}_{ij}x_j\right)\tilde{\mathbf{u}}_i$ so $\tilde{x}_i = \sum_j S^{-1}_{ij}x_j$, or, in matrix notation,

$$\begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_m \end{bmatrix} = S^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

showing that the coordinates of vectors change by multiplication with the inverse of the transition matrix: *vectors are contravariant.*

3.10.4. *How matrices associated to linear transformation change upon changing bases.* Assume we have a linear transformation between two vector spaces $T : U \to V$ and its matrix associated to the bases $\mathcal{B}_U$ of U and $\mathcal{B}_V = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ of V is M:

$$\underset{\mathcal{B}_U}{U} \xrightarrow{T} \underset{\mathcal{B}_V}{V} \tag{17}$$

I. We want to find the matrix of T when the basis of U is changed to $\tilde{\mathcal{B}}_U$. We can do that by composing (17) with the identity transformation (15):

$$\underset{\tilde{\mathcal{B}}_U}{U} \xrightarrow{I} \underset{\mathcal{B}_U}{U} \xrightarrow{T} \underset{\mathcal{B}_V}{V}$$

giving

$$\underset{\tilde{\mathcal{B}}_U}{U} \xrightarrow{TI} \underset{\mathcal{B}_V}{V} \tag{18}$$

where, of course, $TI = T$. By Proposition 11, the matrix of (18) is MS.

II. Suppose now that we want to change the basis of V, to a new basis $\tilde{B}_V$. Denote by R the transition matrix from $\mathcal{B}_V$ to $\tilde{B}_V$, hence R is the matrix of the identity transformation

$$\underset{\tilde{\mathcal{B}}_V}{V} \xrightarrow{I} \underset{\mathcal{B}_V}{V}$$

and R^{-1} is the matrix of the identity transformation in the opposite direction

$$\underset{\mathcal{B}_V}{V} \xrightarrow{I} \underset{\tilde{\mathcal{B}}_V}{V} \tag{19}$$

and by composition

$$\underset{\mathcal{B}_U}{U} \xrightarrow{T} \underset{\mathcal{B}_V}{V} \xrightarrow{I} \underset{\tilde{\mathcal{B}}_V}{V}$$

we obtain

$$\underset{\mathcal{B}_U}{U} \xrightarrow{IT} \underset{\tilde{\mathcal{B}}_V}{V}$$

whose matrix is $R^{-1}M$, which is the matrix of T in the bases $\mathcal{B}_U$, $\tilde{\mathcal{B}}_V$ (since $IT = T$).

Combining I. and II. we deduce

Proposition 12. *Let $T : U \to V$ be a linear transformation whose matrix representation is M in the bases $\mathcal{B}_U$, $\mathcal{B}_V$ of U, respectively V.*

Let $\tilde{\mathcal{B}}_U$ be a basis of U and let S be the transition matrix from $\mathcal{B}_U$ to $\tilde{\mathcal{B}}_U$, and let $\tilde{\mathcal{B}}_V$ be a basis of V with transition matrix R.

Then the matrix representation of T in the bases $\tilde{\mathcal{B}}_U$, $\tilde{\mathcal{B}}_V$ is $R^{-1}MS$.

3.10.5. *Similar matrices.* In the particular case when T is a linear transformation from U to itself: $T : U \to U$, T is called an *endomorphism*[3] of U. Let M be the matrix representation of T in the basis B_U of U (the same basis for the domain and the codomain). Suppose we want to change the basis to $\tilde{B}_U$. Then the matrix representation of T in the basis $\tilde{B}_U$ is $S^{-1}MS$, by Proposition 12.

Definition 13. *Two* square *matrices N, M are called* **similar** *if there exists an invertible matrix S so that $N = S^{-1}MS$.*

As seen above, **similar matrices represent the same linear transformation (endomorphism), but in different bases**.

3.11. **Gauss elimination.** This section brings to light the fact that the allowed transformations in Gauss elimination method correspond to changes of bases.

Suppose we need to solve a linear system $M\mathbf{x} = \mathbf{b}$; say M is an $m \times n$ matrix, and $\mathbf{b}$ is an m-dimensional vector. If G is an invertible $m \times m$ matrix, then our system is equivalent to the system $GM\mathbf{x} = G\mathbf{b}$ (in the sense that the two systems have the same solutions).

We can multiply by G separately M, and $\mathbf{b}$, or we can work with the augmented matrix $M_{aug} = [M \,|\, \mathbf{b}]$, it is the same thing, since

$$GM_{aug} = G[M \,|\, \mathbf{b}] = [GM \,|\, G\mathbf{b}]$$

Viewing M as the matrix of a linear transformation $F^n \to F^m$ (corresponding to the standard bases) then GM is the matrix of the same linear transformation with respect to a new basis on $\mathbb{R}^m$ having the matrix of change of basis equal to $S = G^{-1}$.

Here are the allowed operations in Gauss elimination and the corresponding change of basis.

3.11.1. *Swapping two rows of a matrix.* This operation corresponds to exchanging the place of two vectors in the basis. For example, say we have a linear transformation whose matrix in the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is

$$M = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

Swapping its first two rows means writing the linear transformation in the basis of the codomain where the first two vectors are swapped, namely

[3]Linear transformations preserve the structure of vector spaces in the sense that they take sums to sums, and scalar multiples to scalar multiples, hence they are "vector spaces morphisms". The prefix *endo* comes from the Greek word *endon*, meaning "within". Endomorphisms are, therefore, morphisms from a space into itself.

$\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3$. (The basis of the domain remains the same.) The matrix of change of basis is, of course,

$$E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and noting that $E_{12}^{-1} = E_{12}$ (swapping two vectors twice returns them to the original position), the matrix in the new basis is

$$E_{12}M = \begin{bmatrix} b_1 & b_2 & b_3 \\ a_1 & a_2 & a_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

3.11.2. *Multiplying a row by a non-zero number.* Suppose we want to mulitiply the second row by the scalar t. This corresponds to changing the basis (of the target space) to $\mathbf{e}_1, \frac{1}{t}\mathbf{e}_2, \mathbf{e}_3$, with the matrix of change of basis

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{t} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{with } G_2(t) = S^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

we have

$$G_2(t)M = \begin{bmatrix} a_1 & a_2 & a_3 \\ tb_1 & tb_2 & tb_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

3.11.3. *Adding to one row a scalar multiple of another row.* This happens when the new basis consists on adding to one vector a scalar multiple of another. For example, if the new basis is $\mathbf{e}_1 - \alpha\,\mathbf{e}_2, \mathbf{e}_2, \mathbf{e}_3$ (check that this is a basis indeed!) the matrix of change of basis is

$$F_{1,2}(-\alpha) = \begin{bmatrix} 1 & -\alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

then $F_{1,2}(-\alpha)^{-1} = F_{1,2}(\alpha)$ (which is obvious by thinking of the reverse change of basis) and

$$F_{1,2}(\alpha)M = \begin{bmatrix} a_1 + \alpha\, b_1 & a_2 + \alpha\, b_2 & a_3 + \alpha\, b_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$$

3.11.4. *Gauss elimination is achieved by a composition of the three types of matrices.* We will illustrate this on the example (8) and its usage to solving the system of §(3.7.1).

Consider the linear system with the augmented matrix (10). The aim is to multiply M to the left by an invertible matrix G, so that GM is upper triangular.

We scan the first column, from top to bottom, until we find the first nonzero element. In our example this is 2. Aiming to make all the element below it equal to zero, we multiply M_aug by $F_{21}(-\frac{3}{2})$ obtaining

$$F_{21}(-\frac{3}{2})M_{aug} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3/2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} M_{aug} = \begin{bmatrix} 2 & 0 & 2 & 2\beta \\ 0 & -2 & 2 & -3\beta + \alpha \\ 1 & 1 & 0 & 0 \\ 4 & 3 & 1 & \beta \end{bmatrix}$$

To make the lower column terms zero we further multiply by

$$F_{31}(-\frac{1}{2}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad F_{41}(-2) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

and we obtain

$$F_{41}(-2)\, F_{31}(-\frac{1}{2})\, F_{21}(-\frac{3}{2})\, M_{aug} = \begin{bmatrix} 2 & 0 & 2 & 2\beta \\ 0 & -2 & 2 & -3\beta + \alpha \\ 0 & 1 & -1 & -\beta \\ 0 & 3 & -3 & -3\beta \end{bmatrix} := M_1$$

For the second column we next multiply $F_{42}(\frac{3}{2})\, F_{32}(\frac{1}{2})\, M_1$:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \frac{3}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} M_1 = \begin{bmatrix} 2 & 0 & 2 & 2\beta \\ 0 & -2 & 2 & -3\beta + \alpha \\ 0 & 0 & 0 & -\frac{5}{2}\beta + \frac{1}{2}\alpha \\ 0 & 0 & 0 & -\frac{15}{2}\beta + \frac{3}{2}\alpha \end{bmatrix}$$

which is a matrix upper triangular. In this basis it is easy to see when the system is soluble: the last column belongs to the span of the previous ones if and only if the last two coordinates are zero, hence $\alpha = 5\beta$.

For this α the system has the augmented matrix

$$\begin{bmatrix} 2 & 0 & 2 & 2\beta \\ 0 & -2 & 2 & 2\beta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and can be easily be solved, yielding (11).

3.12. **Projections, symmetries, rotations.** We will now see a few fundamental examples of linear transformation: projections, symmetries, rotations.

3.12.1. *Projections.* An example of (orthogonal) projection in the x_1x_2-plane, is the projection on the x_1 line. It is defined as $P_0 : \mathbb{R}^2 \to \mathbb{R}^2$ by $P_0(x_1, x_2) = (x_1, 0)$. This is a linear transformation (check!) and to find its matrix in the standard basis note that $P_0(\mathbf{e}_1) = \mathbf{e}_1$ and $P_0(\mathbf{e}_2) = \mathbf{0}$ therefore

$$A_{P_0} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

and $P_0(\mathbf{x}) = A_{P_0}\mathbf{x}$ (matrix A_{P_0} times vector $\mathbf{x}$).

As an example of (orthogonal) projection in space, let $Q : \mathbb{R}^3 \to \mathbb{R}^3$ be the projection on the x_1x_2-plane: $Q(x_1, x_2, x_2) = (x_1, x_2, 0)$.

Exercise. Check that $P^2 = P$ and $Q^2 = Q$ (recall that by P^2 we mean $P \circ P$). Find the matrix of Q in the standard basis.

Square matrices satisfying $M^2 = M$ could also be oblique projections (for example, the shadow cast by an object on the surface of the Earth is an oblique projection. Orthogonal projections satisfy one more property...but first we need to define what we mean by orthogonality.

3.12.2. *Rotations.* Consider the transformation of the plane which rotates points by and angle θ counterclockwise around the origin. The simplest way to deduce the formula for this transformation is to consider the plane as the complex plane $\mathbb{C}$, where rotation by angle θ is achieved by multiplication with $\exp(i\theta)$: let $L : \mathbb{C} \to \mathbb{C}$, defined by $L(z) = e^{i\theta}z$.

Now we only need to convert to real coordinates: if $z = x_1 + ix_2$ then

$$e^{i\theta}z = (\cos\theta + i\sin\theta)(x_1 + ix_2) = (\cos\theta x_1 - \sin\theta x_2) + i(\sin\theta x_1 + \cos\theta x_2)$$

therefore define the rotation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by $R_\theta(x_1, x_2) = ((\cos\theta x_1 - \sin\theta x_2), (\sin\theta x_1 + \cos\theta x_2))$ whose matrix in the standard basis is (check!)

$$A_{R_\theta} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and the rotation is the map

$$R\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Remarks:

(i) $R_{\theta_1}R_{\theta_2} = R_{\theta_1+\theta_2}$ for any $\theta_{1,2}$

(ii) R_θ is invertible, and $(R_\theta)^{-1} = R_{-\theta}$.

(iii) $\det A_{R_\theta} = 1$.

3.12.3. *Projections on other lines in* $\mathbb{R}^2$. Consider the orthogonal projection $P_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ on the line $x_2 = \tan\theta\, x_1$. We can obtain its formula using P_0 and changes of coordinates.

We rotate the $\mathbf{x}$-plane by an angle $-\theta$, thus taking the projection line on the x_1-axis, then project using P_0, and then rotate back. We obtain the projection P_θ as the composition: $P_\theta = R_\theta P_0 R_{-\theta}$.

Aa an alternative approach, we can think in terms of changes of coordinates. Let $\tilde{\mathbf{e}}_1 = R_\theta \mathbf{e}_1$ and $\tilde{\mathbf{e}}_2 = R_\theta \mathbf{e}_2$ (then $\tilde{\mathbf{e}}_1$ is a unit vector on the line $x_2 = \tan\theta\, x_1$). In coordinates $\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2$ the projection P_θ has the simple formula $P_\theta(c_1\tilde{\mathbf{e}}_1 + c_2\tilde{\mathbf{e}}_2) = c_1\tilde{\mathbf{e}}_1$ having the matrix A_{P_0} (as it is clear from the geometry of the problem). We now need to write P_θ in standard coordinates $\mathbf{e}_1$, $\mathbf{e}_2$:

$$\begin{array}{ccccccc} \mathbb{R}^2 & \stackrel{I}{\longrightarrow} & \mathbb{R}^2 & \stackrel{P_\theta}{\longrightarrow} & \mathbb{R}^2 & \stackrel{I}{\longrightarrow} & \mathbb{R}^2 \\ \mathbf{e}_{1,2} & & \tilde{\mathbf{e}}_{1,2} & & \tilde{\mathbf{e}}_{1,2} & & \mathbf{e}_{1,2} \end{array}$$

hence the matrix of P_θ in the standard basis of $\mathbb{R}^2$ is $A_{P_\theta} = A_{R_\theta} A_{P_0} A_{R_{-\theta}}$.

3.12.4. *Reflexions.* The reflexion of the plane about its x_1-axis: $T_0 : \mathbb{R}^2 \to \mathbb{R}^2$ is obviously given by $T_0(x_1, x_2) = (x_1, -x_2)$, and its matrix representation in the standard basis is

$$A_{T_0} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Now suppose we want to write the formula for the reflexion $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ about a line of slope θ (with equation $x_2 = \tan\theta\ x_1$. We can obtain his by composition of the simpler transformations we wrote before as follows: rotate the domain of T_θ by an angle $-\theta$; this will take the line into the x_1-axis, we then apply the reflexion T, then we go back to the original coordinates by a rotation of angle θ. This gives $T_\theta = R_\theta T_0 R_{-\theta} = R_\theta T_0 R_\theta^{-1}$.

Exercises.

1) Write the matrix associated to T_θ in the canonical basis.

2) Use the result above to write the matrix associated to a reflexion about the x_2 axis in the canonical basis and check the results using geometrical arguments.

3) Use composition of transformations to find the formula for the orthogonal projection of the plane onto a line $x_1 = m\, x_2$ and check your result geometrically.

4) Use calculations, then check geometrically that reflexions and rotations in the plane almost commute: $R_\theta T_0 = T_0 R_{-\theta}$.

5) Show that the product of two reflexions in the plane is a rotation.

3.13. **Summary of some properties of n dimensional vector spaces.** The following hold for an n-dimensional vector space V. (We assume $n > 0$.)

Theorem. Any linearly independent set of vectors can be completed to a basis.

As a consequence of this theorem (proved earlier):

• Any set of $n+1$ or more vectors in V are linearly dependent.
• Any set of n linearly independent vectors in V is a basis for V.

Theorem. Any spanning set of V contains a basis of V.

Proof. Consider a spanning set: $Sp(\mathbf{v}_1, \ldots, \mathbf{v}_p) = V$.

Step 1. If $\mathbf{0}$ is among the vectors $\mathbf{v}_j$, then we remove it, and the remaining set is still a spanning set.

Step 2. If $\mathbf{v}_1, \ldots, \mathbf{v}_p \in V$ are linearly independent, then they form a basis of V, and the claim is proved.

Otherwise, there exists a linear relation $c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p = \mathbf{0}$ with not all scalars c_j equal to zero. Say $c_p \neq 0$. Then $\mathbf{v}_p$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}$ and therefore $Sp(\mathbf{v}_1, \ldots, \mathbf{v}_p) = Sp(\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}) = V$, and we now have a spanning set with $p-1$ element.

Continue with Step 1. The procedure ends with a basis of V. (It cannot end with the zero vector, since this means that $\dim V = 0$.) □

As a consequence of this theorem:
• Any set of n vectors in V which span V is a basis for V.
• Any spanning set for V must contain at least n vectors.

3.14. **Left inverse, right inverse of functions and of matrices.**

3.14.1. *Inverses for functions.* Let $f : D \to F$ be a function defined on a set D (the domain of F) with values in a set F (the codomain of f).

If f is one to one and onto, then there exists the inverse function $f^{-1} : F \to D$, which satisfies:
(i) $f^{-1}(f(x)) = x$ for all $x \in D$, in other words, $f^{-1} \circ f = Id$ where Id is the identity function of D ($Id : D \to D,\ Id(x) = x$), *and*
(i) $f(f^{-1}(x)) = x$ for all $x \in F$, in other words, $f \circ f^{-1} = Id$ where Id is the identity function of F ($Id : F \to F,\ Id(x) = x$).
(We should normally write Id_D, Id_F for the identity of D, respectively, F.)

In other words, f^{-1} is the inverse of f with respect to composition. Composition of functions is not commutative ($f \circ g \neq g \circ f$) so in order to conclude that f^{-1} is the inverse of F then *both relations* $f^{-1} \circ f = Id$, $f \circ f^{-1} = Id$ must be satisfied.

If $f^{-1} \circ f = Id$ then we say that f^{-1} is a **left inverse** of f.

If $f \circ f^{-1} = Id$ then we say that f^{-1} is a **right inverse** of f.

It is not hard to see that:
1) f in one to one if and only if f has a left inverse;
2) f is onto if and only if f has a right inverse.

If f is not invertible, then these one-sided inverses are not unique.

In what follows we will prove these results and construct one-sided inverses for linear transformations and matrices.

3.14.2. *Left inverse.* Let $T : U \to V$ be a linear transformation.

T has a left inverse $\Longrightarrow$ T is one to one.

Indeed, assume there is $L : V \to U$ so that $L(T(\mathbf{u})) = \mathbf{u}$ for all $\mathbf{u} \in U$. If $\mathbf{u} \in \mathcal{N}(T)$ then $T\mathbf{u} = \mathbf{0}$. But $\mathbf{u} = L(T(\mathbf{u})) = L(\mathbf{0}) = \mathbf{0}$ so $\mathbf{u} = \mathbf{0}$ and T is one to one.

Conversely, *T is one to one $\Longrightarrow$ T has a left inverse.*

Let T be one to one. If it is also onto, then we can take L to be the inverse of T, and the problem is solved.

Otherwise, split V as $V = R(T) \bigoplus W$ where W is a complement of $R(T)$. Any $\mathbf{v} \in V$ can be uniquely written as $\mathbf{v} = \mathbf{y} + \mathbf{w}$ with $\mathbf{y} \in R(T)$ and $\mathbf{w} \in W$ (by Remark 22 of § 1.7). There is a unique $\mathbf{u} \in U$ so that $\mathbf{y} = T\mathbf{u}$ (since T is 1-to-1). Define $L\mathbf{v} = \mathbf{u}$ and we can check that $LT\mathbf{u} = \mathbf{u}$.

Note that we can as well define $L\mathbf{v} = \mathbf{u} + L_0\mathbf{w}$ where $L_0 : W \to U$ is any linear transformation.

Now formulate this result in the language of matrices. Let M be an $m \times n$ matrix, and consider $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T\mathbf{x} = M\mathbf{x}$. The assumption that T is one to one means that the columns of M are linearly independent, hence $n \leq m$ and $\operatorname{rank} M = n$ (by Theorem 7 (iii)), therefore:

Theorem 14. *Let M be $m \times n$ matrix with $n \leq m$.*

M has rank n (maximal rank) if and only if M has a left inverse: an $n \times m$ matrix L so that $LM = I$.

If $n = m$ then $L = M^{-1}$, while if $n \neq m$ then L is not unique.

3.14.3. *Right inverse.* Let $T : U \to V$ be a linear transformation.

T has a right inverse $\Longrightarrow$ T is onto.

Indeed, there is $R : V \to U$ so that $T(R(\mathbf{v})) = \mathbf{v}$ for all $\mathbf{v} \in V$. Therefore any $\mathbf{v} \in V$ is $\mathbf{v} = T\mathbf{x}$ for $\mathbf{x} = R\mathbf{v}$, so T is onto.

Conversely, *T is onto $\Longrightarrow$ T has a right inverse.*

Let T be onto. If T is also one to one, then we can take R to be the inverse of T.

Otherwise, we want to define $R\mathbf{v} = \mathbf{u}$ where $\mathbf{u}$ is so that $T\mathbf{u} = \mathbf{v}$, the problem is that there are many such $\mathbf{u}$, as any vector in $\mathbf{u} + \mathcal{N}(T)$ works as well (and we cannot choose arbitrarily, since we want a linear transformation). To attain a linear "choice", split $U = \mathcal{N}(T) \oplus W$ where W is a complement. For any $\mathbf{v} \in V$ let $\mathbf{u}$ be such that $T\mathbf{u} = \mathbf{v}$. Write $\mathbf{u} = \mathbf{u}_N + \mathbf{u}_W$ (note that $\mathbf{u}_W$ depends linearly on $\mathbf{u}$, it is a projection, usually oblique) and define $R\mathbf{v} = \mathbf{u}_W$.

In the language of matrices: let M be an $m \times n$ matrix, and consider $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\mathbf{x}) = M\mathbf{x}$. The assumption that T is onto means that $n \geq m$ and that the column space $Sp(M\mathbf{e}_1, \ldots, M\mathbf{e}_n) = \mathbb{R}^m$, which means that $\operatorname{rank} M = m$.

Theorem 15. *Let M be an $m \times n$ matrix with $n \geq m$.*

M has maximal rank m if and only if M has a right inverse: an $m \times n$ matrix R so that $MR = I$.