Chapter 2

Vector Spaces

2.1 Definitions and Examples

Linear equations in n variables are *mathematical objects*: they can be added, and multiplied by numbers. We say that the linear equations of a fixed type form a *vector space.*[1]

Definition. *Any set E consisting in mathematical objects which may be added,*[2] *and multiplied by numbers,*[3] *having the usual properties (commutativity, associativity, distributivity...) is a* vector space.

It is essential to observe that in general, the elements of a vector space cannot be multiplied together: In a vector space, the presence of an inner multiplicative law is not required (not forbidden either!).

The elements of a vector space are also called *vectors*—or more precisely *generalized vectors*—while numbers are called *scalars.* It is suitable to use different alphabets for them. For example, if ρ denotes the row $(2, 3, -1)$, i.e. the homogeneous equation $2x_1 + 3x_2 - x_3 = 0$, the multiple $a\rho$ denotes $(2a, 3a, -a)$, namely the equation $2ax_1 + 3ax_2 - ax_3 = 0$. If $\rho' = (1, 2, 2)$ is another row of the same type (homogeneous equation in three variables), then

$$2\rho + \rho' = (4, 6, -2) + (1, 2, 2) = (5, 8, 0)$$

represents the equation $5x_1 + 8x_2 + 0x_3 = 0$ of the same type.

More traditionally—and for reasons that will appear later—scalars are often denoted by Greek letters[4] while the elements of a vector space are covered by an arrow.[5] We might just as well represent the row $\rho = (2, 3, -1)$ by $\vec{\mathbf{r}}$ (or $\overrightarrow{r}$, $\mathbf{r}, \ldots$) and its multiples by $\alpha\vec{\mathbf{r}}$ (resp. $a\mathbf{r}, \ldots$). The notation should only be chosen in such a way as to suggest the correct interpretation.

[1] We refer to the Appendix to this chapter for the general notation concerning set theory.

[2] Addition being an internal operation $E \times E \to E$. Here appears the Cartesian product of sets; cf. Appendix.

[3] This is an external operation $\mathbf{R} \times E \to E$. Here appears the canonical set $\mathbf{R}$ of real numbers.

[4] A Greek alphabet appears in the Appendix to this chapter.

[5] Still another type of *arrow*: Arrows are used for mappings, convergence, etc.

A vector space is a set—or space—in which finite sums of multiples of elements, called *linear combinations*, can be made. With rows $\rho_1, \ldots, \rho_m$, one may also consider the linear combinations

$$a_1\rho_1 + a_2\rho_2 + \cdots + a_m\rho_m$$

where $a_1, a_2, \ldots, a_m$ are numbers (not rows!). It is more economical to write such a linear combination in the form of a sum of the *generic term* $a_i\rho_i$ for the values of the index i between 1 and m:

$$a_1\rho_1 + a_2\rho_2 + \cdots + a_m\rho_m = \sum_{1\le i\le m} a_i\rho_i.$$

Examples. (1) The set of homogeneous linear equations in n variables is a vector space sometimes denoted by $\mathbf{R}_n$. Its elements are the rows $\rho = (a_1, \ldots, a_n)$. Linear combinations of rows are computed according to the rule

$$a\rho + \rho' = (aa_1 + a'_1, \ldots, aa_n + a'_n)$$

(with obvious notations). One can also consider another space, consisting of the linear equations in n variables, having elements $(a_1, \ldots, a_n\,;\, b)$ with a similarly defined addition and multiplication by scalars, like the corresponding operations on equations.

(2) The lists consisting of n numbers written vertically can also be amplified and added in a natural way. We have encountered this situation with solutions of linear systems, and in particular, in the general principle stating that the general solution of a linear system can be obtain by making the *sum* of a particular solution and the general solution of the associated homogeneous system

$$\begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = \begin{pmatrix} p_1 + h_1 \\ p_2 + h_2 \\ \vdots \\ p_n + h_n \end{pmatrix}.$$

This vector space, consisiting of lists of n numbers written vertically is conventionally denoted by $\mathbf{R}^n$.[6]

A geometrical representation of the space $\mathbf{R}^2$ is given by a choice of *Cartesian coordinates* in the Euclidean plane: to the list $\binom{x_1}{x_2} = \binom{x}{y}$ we associate the point P having coordinates x and y. In this way, lists correspond one-to-one with points. However, the amplification of a list componentwise, as well as the sum of lists componentwise, makes it more intuitive to replace the point P by the vector $\overrightarrow{OP}$. The addition law of components now corresponds to the usual *parallelogram law* for adding vectors

$$a\vec{v} + \vec{w} = a\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} av_1 + w_1 \\ av_2 + w_2 \end{pmatrix}.$$

A similar geometrical representation for the space $\mathbf{R}^3$ of lists of three numbers is given by the usual Euclidean, or physical space: To a list, we associate the

[6] It is the paradigm of finite-dimensional vector space.

point P or the vector $\overrightarrow{OP}$ having for coordinates the three numbers in the list.[7] It is more difficult to imagine representations for $\mathbf{R}^4$ (time-space), $\mathbf{R}^5$ or $\mathbf{R}^{100}$. We simply consider the elements of $\mathbf{R}^n$ as being lists of n numbers (written vertically): We also call them *n-tuples*. When $n = 1$, we obtain the vector space consisting of scalars.

(3) The set $E = \mathcal{F}(\mathbf{R}, \mathbf{R})$ of all functions $\mathbf{R} \to \mathbf{R}$ is a vector space. Indeed, we can add real valued functions on $\mathbf{R}$ pointwise, as well as amplify them by (real) scalars. (It turns out that we can also multiply pointwise two functions, hence speak of an inner multiplication law in this vector space; but this possibility is irrelevant here). Hence a function $f : \mathbf{R} \to \mathbf{R}$ may also be called a (generalized) "vector" when we deal with this vector space. This vector space is huge, and it will often be more reasonable to use a smaller one $\mathcal{C}(\mathbf{R}, \mathbf{R})$ consisting of the *continuous* functions: The sum of two continuous functions is continuous—so we are told in an elementary calculus course—and so are the multiples of continuous funtions.

(4) Here is a generalization of the first examples. Consider the set of rectangular arrays of size $m \times n$ (corresponding to a homogeneous linear systems containing m equations in n variables). Define the addition of two arrays componentwise: For example with 2 rows and 3 columns

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{pmatrix}.$$

Define similarly the multiplication by scalars componentwise, e.g.

$$a \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} aa_{11} & aa_{12} & aa_{13} \\ aa_{21} & aa_{22} & aa_{23} \end{pmatrix}.$$

This set $\mathbf{M}_{m,n}(\mathbf{R}) := (\mathbf{R}_n)^m = (\mathbf{R}^m)_n =: \mathbf{R}^m_n$ is a vector space playing a central role in linear algebra.

2.2 Subspaces, Generators

Definition. *A subset $V \subset E$ of a vector space is a* subspace *when it contains the zero vector of E: $0_V = 0_E = \vec{0}$, and*

v and $w \in E$ implies $av + w \in E$ for any scalar a.

The zero vector of E alone constitutes the smallest subspace $\{0\}$ of E, sometimes called a *trivial* subspace. The whole space itself is also an example of *trivial* subspace. When $E \neq \{0\}$ we get two trivial—extreme—examples of subspaces. We are mainly interested in *nontrivial* subspaces of a vector space E, but the trivial examples often occur as particular cases in general statements, and it would be awkward to exclude them a priori.

Examples. (1) Let us give the general form of the nontrivial subspaces of $\mathbf{R}^3$. First, we observe that the lines going through the origin,[8] furnish infinitely

[7] We assume some familiarity with the use of vectors in the context of forces, velocities (not to be confused with *speed*), acceleration, etc.

[8] Also called *homogeneous* lines.

many examples of subspaces of $\mathbf{R}^3$. Secondly, the planes containing the origin furnish infinitely many other examples of subspaces of $\mathbf{R}^3$. It will be obvious later that these are all the nontrivial subspaces of $\mathbf{R}^3$.

(2) Consider a system (HS) of homogeneous linear equations in n variables. The solutions of this system form a subset of the space $\mathbf{R}^n$ of lists $(x_i)_{1\le i\le n}$ of n numbers. In Chapter 1, we observed that the sum of two solutions, as well as multiples of solutions, are again solutions of this homogeneous system. With the present terminology, we may say that the solutions of (HS) form a subspace of $\mathbf{R}^n$. But the solutions of a nonhomogeneous system do not constitute a subspace, simply since the trivial family—consisting of 0's only—is not a solution of a nonhomogeneous system.

(3) Let us consider the set $E = \mathcal{F}(\mathbf{R}, \mathbf{R})$ of all functions $\mathbf{R} \to \mathbf{R}$. Since we can add real valued functions on $\mathbf{R}$ pointwise, as well as amplify them by (real) scalars, this huge set is a vector space.[9] Since the sum of two continuous functions is again a continuous function, and so are the multiples of continuous functions, the subset $\mathcal{C}$ consisting of continuous functions is a subspace of E. The same properties are valid for polynomial functions instead of continuous ones. Hence the polynomial functions also constitute a subspace Π of E. They also form a subspace of $\mathcal{C}$:

$$\Pi \subset \mathcal{C} \subset E.$$

The *quadratic functions*, namely the functions having a representation in the form $f(x) = ax^2 + bx + c$ for some scalars a, b, and c, make up a subspace of Π and hence also of $\mathcal{C}$, and of E. The *degree* of a *nonzero polynomial* is by definition the highest power of the variable x that occurs (with a nonzero coefficient) in this polynomial. For example, the polynomials of degree less than or equal to 2 can be represented in the form $f(x) = ax^2 + bx + c$. By convention, we define the degree of the zero polynomial to be less than any possible degree. With this convention, the quadratic polynomials are precisely the polynomials of degree less than or equal to 2. The polynomial functions of degree exactly 2 *do not* produce a subspace: The sum of two polynomials of degree 2 may have degree less than 2, e.g.

$$(x^2 + x + 1) + (-x^2 + 1) = x + 2 \quad \text{has degree 1.}$$

More generally (with the preceding convention concerning the degree of the zero polynomial) the polynomials of degree less than or equal to any natural integer n form a subspace $\Pi_{\le n}$. For example, the subspace of quadratic polynomials is $\Pi_{\le 2}$. We have a sequence of inclusions of subspaces

$$\{0\} \subset \Pi_{\le 0} \subset \Pi_{\le 1} \subset \Pi_{\le 2} \subset \cdots \subset \Pi_{\le n} \subset \cdots \subset \Pi \subset \mathcal{C} \subset E$$

(when $n \ge 2$).[10]

Linear Span of a Family of Vectors

In a vector space E, the smallest subspace containing elements $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n \in E$ is denoted by

$$V = \mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) \subset E.$$

[9] It turns out that we can also multiply together two functions pointwise, but this possibility is irrelevant here.

[10] Observe that $\Pi_{\le 0}$ is the space of constants: The multiples of the constant function 1.

As any subspace, this subspace must contain all linear combinations

$$x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n = \sum x_i\vec{a}_i$$

of the given elements. Now, the set of all linear combinations is a subspace since multiples of linear combinations are again linear combinations[11]

$$a\,(x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n) = ax_1\vec{a}_1 + ax_2\vec{a}_2 + \cdots + ax_n\vec{a}_n,$$

and similarly for the sum of two linear combinations

$$(x_1\vec{a}_1 + \cdots + x_n\vec{a}_n) + (y_1\vec{a}_1 + \cdots + y_n\vec{a}_n) = (x_1+y_1)\vec{a}_1 + \cdots + (x_n+y_n)\vec{a}_n.$$

We can write

$$\begin{aligned}\mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) &= \text{set of linear combinations of the } \vec{a}_i \\ &= \Big\{ \sum_{1\le i\le n} x_i\vec{a}_i \ : \ \text{any scalars } x_i \Big\},\end{aligned}$$

and call it the *linear span* of the elements $\vec{a}_i$ $(1 \le i \le n)$. We also say that this space is *generated*, or *spanned*, by the elements $\vec{a}_i$.

Obviously

$$\mathcal{L}(\vec{a}_2, \vec{a}_1, \ldots, \vec{a}_n) = \mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n).$$

This subspace does not depend on the order in which the elements a_i are listed. It is as obvious that

$$\mathcal{L}(c\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) = \mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n),$$

when $c \neq 0$ is a nonzero scalar. The subspace $\mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n)$ does not change if we replaces one element a_i by a nonzero multiple ca_i $(c \neq 0)$.

Important Example. Let A denote a rectangular array of coefficients (corresponding to a linear system) containing m rows $\rho_1, \ldots, \rho_m$ of a certain type. In the vector space E of rows of the same type, consider the *row space* of A, namely the subspace

$$\mathcal{L}(\text{rows}\, A) := \mathcal{L}(\rho_1, \ldots, \rho_m)$$

generated by the rows of A. Let A' be the rectangular array obtained after one row operation has been performed on A. As has already been observed, permuting two rows, or amplifying one row with a nonzero scalar, does not alter the row space. Let us consider the third type of row operation. Typically, we may imagine that the first row of A' is obtained from the first row of A by addition of a multiple of its second row:

$$\rho_1' = \rho_1 + c\rho_2 \in \mathcal{L}(\rho_1, \ldots, \rho_m).$$

This proves

$$\mathcal{L}(\rho_1', \ldots, \rho_m) \subset \mathcal{L}(\rho_1, \ldots, \rho_m).$$

[11] For this and the next assertions, verify that the axioms of vector spaces only are used!

Since this row operation is invertible, explicitly $\rho_1 = \rho_1' - c\rho_2$, the opposite inclusion is also valid, and we deduce

$$\mathcal{L}(\rho_1', \ldots, \rho_m) = \mathcal{L}(\rho_1, \ldots, \rho_m).$$

The row space of A is equal to the row space of A'. Performing successive row operations until a reduced form is obtained

$$A \sim A' \sim \cdots \sim U$$

we conclude that the row spaces of A and of any row-equivalent form of A are the same:

$$\mathcal{L}(\text{rows}\, A) = \mathcal{L}(\text{rows}\, U).$$

Interpretation of a linear system (S) with m-tuples

If a linear system (S) is given, we may introduce the m-tuples

$$\vec{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbf{R}^m \quad (1 \le j \le n),$$

and rewrite the system in the equivalent form

$$(S) \qquad x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

More simply, in vector form

$$x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{b}.$$

This system is compatible when we can find some values for the coefficients x_i, namely when the vector $\vec{b}$ *is* a linear combination of the vectors $\vec{a}_j$:

$$\vec{b} \in \mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n).$$

An equivalent way of giving this condition is

$$\mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) = \mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n, \vec{b}).$$

The system can be solved *for all data* $\vec{b} \in \mathbf{R}^m$ precisely when the subspace generated by the vectors $\vec{a}_i$ is the whole space

$$\mathcal{L}(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n) = \mathbf{R}^m.$$

2.3 Linear Independence

When a linear system has two *distinct* solutions, their difference is a *nontrivial* solution of the associated homogeneous system (HS). Let us introduce the columns $\mathbf{a_j}$ so that

$$(HS) \qquad \mathbf{a_1}x_1 + \mathbf{a_2}x_2 + \cdots + \mathbf{a_n}x_n = \mathbf{0},$$

and there is a nontrivial linear combination

$$x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_n\mathbf{a_n} = \mathbf{0}.$$

At least one coefficient $x_i \neq 0$ (but we cannot say a priori for which index i it is the case!). On the contrary, when this system only has the trivial solution, we say that the columns $\mathbf{a_j}$ are *linearly independent*. This suggests a general definition, valid in any vector space.

Definition. *We say that a family of vectors* $(\vec{a}_i)_{i\in I}$ *is* free *or equivalently* linearly independent *when the only linear combination producing the zero vector is the trivial one, having all zero coefficients*

$$x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \sum_{\text{finite}} x_i\vec{a}_i = \vec{0} \quad \textit{implies} \quad \textit{all } x_i = 0.$$

In the opposite case, we say that the family is linearly dependent *or* linked*: There is a* dependence relation $x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{0}$*, namely a nontrivial linear combination producing the zero vector.*

If $x_i \neq 0$ in $x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{0}$, we can solve

$$\vec{a}_i = -\tfrac{1}{x_i}\sum_{j\neq i} x_j\vec{a}_j$$

and $\vec{a}_i$ is a linear combination of the other $\vec{a}_j$, explaining the terminology chosen. If a finite sum[12] $\sum_i x_i\vec{a}_i = 0$ has a nonzero coefficient, then one vector is a linear combination of the others: But we do not know a priori which one. The advantage of the definition is its symmetry.

Comment. If a subset of the vector space E is free, it does not contain the zero vector $\vec{0} \in E$. Indeed, $a\vec{0} = \vec{0}$ holds for any scalar a: taking $a = 1$, we get a nontrivial dependence relation.[13]

Example. The nonzero rows of a row-reduced array are independent.

$$\left(\begin{array}{cccccc} \lfloor\underline{p_1 \quad *} & & & & & \\ 0 & & \lfloor\underline{p_2 \quad *} & & & \\ \vdots & & 0 & \cdots & \lfloor\underline{p_r \quad *} & \\ \vdots & & \vdots & & 0 & \cdots \end{array}\right) \begin{array}{c} \rho_1 \\ \rho_2 \\ \rho_r \\ \vdots \end{array}$$

A linear combination $\sum a_i\rho_i$ can vanish only if $a_1 = 0$: This is seen by considering the first coefficient of the row. Having acquired this, one can look at the coefficient of index given by the second pivot, and successively prove that all coefficients are zero.

Although not needed in this book, let us establish a result which is very important in analysis.

[12] In the right-hand side, we ought to write $\vec{0}$ instead of "0": we shall now often make this abuse, relying on the reader for the proper interpretation!

[13] In the trivial vector space $E = \{0\}$, the only free subset is the empty set $\varnothing$.

Proposition. *Consider a family (λ_i) of distinct scalars. Then the family of functions $(x^j e^{\lambda_i x})$ is independent.*[14]

Proof. We have to prove that any finite linear combination $\sum c_{ij}x^j e^{\lambda_i x}$ that vanishes identically, has all $c_{ij} = 0$. Grouping terms with the same index i, we see that we have to prove that a finite sum $\sum p_i e^{\lambda_i x}$ having polynomial coefficients, that vanishes identically, has all $p_i = 0$. We show this by induction on the number m of terms in such a sum.

(a) Case $m = 1$. Let $p(x)e^{\lambda x}$ vanish identically, where p is a polynomial. Since $e^{\lambda x}e^{-\lambda x} = e^{\lambda x-\lambda x} = e^0 = 1$, the exponential never vanishes and the assumption implies that $p(x) = 0$ vanishes identically. This can only happen if $p = 0$ is the trivial polynomial (having all zero coefficients).

(b) Induction step. Assume that for some $m \geq 1$

$$\sum_{1\leq i\leq m} p_i(x)e^{\lambda_i x} = 0 \text{ for all } x \Longrightarrow p_i = 0 \quad (1 \leq i \leq m)$$

(where the p_i's are polynomials, and the λ_i are distinct scalars). Consider a dependence relation having one more term

$$\sum_{1\leq i\leq m} p_i(x)e^{\lambda_i x} + p_{m+1}(x)e^{\lambda_{m+1} x} = 0 \text{ for all } x$$

(with polynomial coefficients p_j, and λ_{m+1} distinct from all preceding λ_i's). If we multiply this identity by $e^{-\lambda_{m+1}x}$, we get

$$\sum_{1\leq i\leq m} p_i(x)e^{\lambda'_i x} + p_{m+1}(x) = 0 \text{ for all } x,$$

where all $\lambda'_i = \lambda_i - \lambda_{m+1}$ are distinct and *nonzero* scalars. Differentiating this identity, we infer

$$\sum_{1\leq i\leq m} q_i(x)e^{\lambda'_i x} + p'_{m+1}(x) = 0 \text{ for all } x,$$

where $q_i = \lambda'_i p_i + p'_i$ *has the same degree as* p_i. Iterating this procedure $d + 1$ times where $d = \deg p_{m+1}$, we obtain a simpler identity

$$\sum_{1\leq i\leq m} r_i(x)e^{\lambda'_i x} = 0 \text{ for all } x,$$

still with polynomials r_i *having the same degree as* p_i. By induction assumption however, the only possibility is now $r_i = 0$ $(1 \leq i \leq m)$. The degree consideration shows that $p_i = 0$ for the same values of the index i. There only remains a dependence relation

$$p_{m+1}(x)e^{\lambda_{m+1}x} = 0 \text{ for all } x.$$

As we have seen in the first part of the proof, it implies $p_{m+1} = 0$ also. ∎

[14] Although we consider only *real* scalars here, the reader may observe that this proof works as well for *complex* scalars.

2.4 Bases, Dimension

Having discussed the notion of generators and of linear independence, we now gather the two concepts.

Theorem. *Let $A = \{\vec{a}_1, \ldots, \vec{a}_m\}$ be a finite subset in a finitely generated subspace $\mathcal{L}(\vec{b}_1, \ldots, \vec{b}_n)$. If $m > n$, then A is dependent.*

This basic result may be reformulated as:

Any family having more elements than a generating set is dependent.

By *logical contraposition*, we obtain the equivalent statement

Any free *subset of $\mathcal{L}(\vec{b}_1, \ldots, \vec{b}_n)$ has at most n elements.*

Proof. Let us start with m vectors

$$\vec{a}_1, \vec{a}_2, \vec{a}_3, \ldots, \vec{a}_m,$$

in the subspace generated by

$$\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n,$$

where $m > n$. Hence we may write the $\vec{a}_j$ as linear combinations of the $\vec{b}_i$'s

$$\vec{a}_j = a_{j1}\vec{b}_1 + a_{j2}\vec{b}_2 + \cdots + a_{jn}\vec{b}_n.$$

Now, let us form linear combinations of these

$$+\left\{\begin{array}{ccl} x_1\vec{a}_1 & = & x_1 a_{11}\vec{b}_1 + x_1 a_{12}\vec{b}_2 + \cdots + x_1 a_{1n}\vec{b}_n, \\ & \vdots & \\ x_m\vec{a}_m & = & x_m a_{m1}\vec{b}_1 + x_m a_{m2}\vec{b}_2 + \cdots + x_m a_{mn}\vec{b}_n, \end{array}\right.$$

$$\begin{array}{ccl} \Sigma_j\, x_j\vec{a}_j & = & (x_1 a_{11} + \ldots + x_m a_{m1})\vec{b}_1 \\ & & \quad + \quad \ldots \quad + \\ & & \qquad (x_1 a_{1n} + \ldots + x_m a_{mn})\vec{b}_n. \end{array}$$

To obtain zero with such a linear combination, we can simply choose the coefficients (x_i) solution of the homogeneous linear system

$$\left\{\begin{array}{rcl} x_1 a_{11} + \ldots + x_m a_{m1} & = & 0, \\ & \vdots & \\ x_1 a_{1n} + \ldots + x_m a_{mn} & = & 0. \end{array}\right.$$

Since this system has more variables (m) than equations (n), it has a nontrivial solution and we are done: The vectors $\vec{a}_i$ are linearly dependent. ■

Definition. *A* basis *of a vector space is a* free generating *family.*

We shall mainly be interested in *finitely generated* vector spaces, namely vector spaces E for which there is a finite family (a_i) of elements such that $E = \mathcal{L}(a_1, \ldots, a_n)$. If this family is not free, one element can be expressed in function of the other ones, and deleting it, we obtain a set of generators having one less element. Continuing in this way, we finally reach a basis of E. We have thus proved the first part of the following basic theorem.

Theorem. *Let E be a finitely generated vector space. Then E has a basis. Two bases of E have the same number of elements.*

PROOF. Take two bases A and B of E, and let $\operatorname{card} A$, $\operatorname{card} B$ denote their respective number of elements. Then

$$\begin{aligned} A \text{ generates and } B \text{ free} &\Longrightarrow \operatorname{card} B \le \operatorname{card} A, \\ B \text{ generates and } A \text{ free} &\Longrightarrow \operatorname{card} A \le \operatorname{card} B. \end{aligned}$$

This proves $\operatorname{card} A = \operatorname{card} B$. ■

Definition. *The common number of elements in all bases of a finitely generated vector space E is called the* dimension *of E: and is denoted by* $\dim E$.

Example 1 The vector space $E = \mathbf{R}^n$ has dimension n. To prove this, we have to give a basis of this space. I claim that the following n-tuples

$$\vec{\mathbf{e}}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \vec{\mathbf{e}}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \vec{\mathbf{e}}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

constitute a basis of the space. Let us make arbitrary linear combinations of these vectors. By definition

$$x_1\vec{\mathbf{e}}_1 + x_2\vec{\mathbf{e}}_2 + \cdots + x_n\vec{\mathbf{e}}_n = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Hence taking suitable coefficients x_i, we can obtain any n-tuple: These elements $\vec{\mathbf{e}}_i$ make up a set of generators of $\mathbf{R}^n$. Moreover, a linear combination of these can furnish the zero n-tuple only if all coefficients x_i vanish: They are independent. This basis has n elements: The dimension of $\mathbf{R}^n$ is n.[15] When $n \ge 1$, there are many other bases of this space. However, the preceding one is more natural, and is therefore called *the canonical basis* of $\mathbf{R}^n$. In an abstract vector space, there is usually no way of selecting a prefered basis.

Example 2 In a similar vein, consider the vector space $\mathbf{R}^n_m$ consisting of arrays of size $m \times n$. It has a canonical basis

$$(E_{ij})_{1 \le i \le m,\, 1 \le j \le n},$$

consisting of the matrices

$$E_{11} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ \vdots & & & \ddots & \vdots \\ 0 & & & \ldots & 0 \end{pmatrix}, \quad E_{12} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & & \\ \vdots & & & \ddots & \vdots \\ 0 & & & \ldots & 0 \end{pmatrix}, \ldots$$

[15] The field of scalars is a one-dimensional vector space: Any nonzero element is a basis for this space.

where E_{ij} has only one nonzero coefficient—this being a 1—placed at the intersection of the ith row and jth column. Hence this vector space has dimension mn.

Example 3 Consider the subspace V of $\mathbf{R}^n$ formed with the solutions of a homogeneous system $Ax = 0$.[16] Attributing successively the values $1, 0, \ldots, 0$, and then $0, 1, 0, \ldots, 0$ etc. to the free variables, we find a basis of V: Its dimension is $n - r$ and

$$\operatorname{rank} A + \dim V = r + (n - r) = n.$$

Fundamental Application. Let A be any rectangular array of coefficients. Using row operations, we can find a reduced row-equivalent form of A, say $A \sim U$. We have seen $\mathcal{L}(\text{rows}\, A) = \mathcal{L}(\text{rows}\, U)$. Now, the nonzero rows of U form a system of generators of this space. Since they are independent, they constitute a *basis* of the row-space:

$$r = \dim \mathcal{L}(\text{rows}\, A).$$

Two procedures $A \sim U$, resp. $A \sim U'$, leading to row-reduced forms of A furnish two bases of $\mathcal{L}(\text{rows}\, A)$, hence have the same number of elements: $r = r'$. This proves that the rank r is independent of the particular method of reduction,[17] and the row-rank of A can be defined by

$$\operatorname{rank} A := \dim \mathcal{L}(\text{rows}\, A).$$

Let us quote an easy refinement of the preceding theorem. A basis of E appears as a minimal generating set of this space. Symmetrically, a basis also appears as a *maximal free* family. We get a basis by successive adjunction of elements which cannot be expressed by means of the former ones. In a finitely generated vector space, this procedure will eventually furnish a set of generators. For example, starting from a basis of a subspace, we may complete it into a basis of the whole space. This is the content of the next theorem which will be refered to as the *incomplete basis theorem.*

Theorem. *Let F be a free subset of a finitely generated vector space $E = \mathcal{L}(v_1, v_2, \ldots, v_m)$. Then one can obtain a basis of E by addition of suitable elements v_j to F.*

Proof. If the free set F does not generate E, then at least one of the generators v_j does not belong to $\mathcal{L}(F)$. Hence the union of F with this element is still free and one can continue until a free generating set is obtained: This process stops after at most m steps. ■

For reference, we also quote explicitly the following particular case.

Corollary. *For any $0 \neq x \in E$, there is a basis of E containing x.*

[16] Also called nilspace of the array A.

[17] One can also show that the ranks of the pivots are well defined, independently from the method of reduction: In fact, there is uniqueness of the row-echelon form of any array A.

2.5 Appendix

2.5.1 Set Theory, Notation

A *set* is a collection of mathematical objects. It is given by a list between brackets, e.g. the set consisting of the two numbers 1 and 2 is $\{1, 2\}$. For infinite sets we use dots, e.g. the set of *natural integers* is

$$\mathbf{N} = \{0, 1, 2, 3, \ldots\}.$$

or we list the property which is characteristic of the set. For example, the set of even numbers is

$$E = \{0, 2, 4, 6, \ldots\} = \{2n : n \text{ is a natural integer}\}.$$

This set is contained in—or is a subset of—the set of natural numbers $\mathbf{N}$. The inclusion of two sets is represented by the sign $\subset$,[18] e.g. $E \subset \mathbf{N}$. Another subset is the set of *prime integers*, or simply the set of *primes*

$$P = \{2, 3, 5, 7, 11, \ldots\} \subset \mathbf{N}.$$

To indicate that an element belongs to a set, we use the $\in$ symbol: Instead of "23 is a prime" we may equivalently write "$23 \in P$", which is read "23 is an element of—or belongs to—the set P of primes". The negation of $\in$ is denoted by $\notin$, e.g. $1 \notin P$: the integer 1 is not a prime.[19]

When two sets E, F are given, we may define their *intersection*, denoted by $E \cap F = F \cap E$, consisting of their common elements. For example, the intersection of the set of even numbers and the set of primes is $\{2\}$, a set consisting of a single element, also called a *singleton set*. Here are two equivalent notations:

$$2 \in P \quad \text{and} \quad \{2\} \subset P.$$

The *union* of two sets E and F, denoted by $E \cup F = F \cup E$,[20] is the set consisting of the elements which belong to at least one of the sets in question. For example, the union of the set of natural numbers $\mathbf{N}$ and the set of negative integers $\{-1, -2, -3, \ldots\}$ is the set $\mathbf{Z}$ of *rational integers*.[21]

The notation $A \subset B$ is also symmetrically denoted by $B \supset A$. To prove an equality of two sets A and B, we may proceed by *double inclusion*, namely prove $A \subset B$ and $B \subset A$.

If E is a set, the subsets of E constitute a new set

$$\mathcal{P}(E) = \{A : A \text{ is a subset of } E\}.$$

The *empty set* $\emptyset$ is a subset of any set E, hence $\emptyset \in \mathcal{P}(E)$ and this shows that $\mathcal{P}(E)$ is never empty! By convention, E itself is also a subset of E: $E \in \mathcal{P}(E)$. If two subsets A and B verify an inclusion $A \subset B$, we denote by $B - A$ the

[18] The symbol $\subset$ is a reminder of the first letter in "contained".

[19] Here is a reason for this: We like to have *unique* prime decompositions (up to order) of integers, e.g. $6 = 2 \cdot 3$. If we had admitted 1 as a prime, we could write $6 = 2 \cdot 3 = 1 \cdot 2 \cdot 3 = 1 \cdot 1 \cdot 2 \cdot 3 = \ldots$, whence nonuniqueness.

[20] The symbol $\cup$ is a reminder of the first letter in "union".

[21] The chosen letter is the initial of the German "Zahl" (=Number).

relative complement consisting of the elements of B not in A. The *complement* of a subset $A \subset E$ is the subset

$$A^c = E - A = \{x \in E : x \notin A\}.$$

By definition, the complement of A^c is A itself

$$(A^c)^c = A.$$

The complement of a union is the intersection of the complements

$$(A \cup B)^c = A^c \cap B^c.$$

The union of any family of sets is the set consisting of the elements which belong to at least one. For three sets, A, B, and C, this union is the set $A \cup B \cup C$. It can be obtained by first taking the union of A and B, and then the union of $A \cup B$ and C. Hence

$$(A \cup B) \cup C = A \cup B \cup C = A \cup (B \cup C).$$

This is the associativity of the "$\cup$" operation. Similar considerations hold for the intersection:

$$(A \cap B) \cap C = A \cap B \cap C = A \cap (B \cap C).$$

There is also a distributivity relation[22]

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

These operations make the basis of the *Boolean algebra* of subsets of a set E.

When E and F are two sets, their *Cartesian product* $E \times F$ denotes the set of (ordered) *pairs* (x, y) where x is taken in E and y in F. In particular, when $E = F$, the Cartesian product $E \times E$ consists of pairs of elements of E. Here, we note $E_2 := E \times E$, while E^2 rather denotes the set of *vertical* pairs of elements of E. Similar considerations apply to $E_n := E \times \cdots \times E$ consisting of rows with n elements of E, while E^n denotes the set of n-tuples written in column.

A *mapping* or *map* from a set E into a set F is a correspondence which to *each* element $x \in E$ associates *one* element $y \in F$. We often denote a map by $f : E \to F$,[23] and we indicate the correspondence at the level of the elements by $x \mapsto y = f(x)$. Thus a map $f : E \to F$ is defined on all of E, and for each $x \in E$, there is only one element $f(x) \in F$. The subset of F consisting of all $f(x)$ when $x \in E$ is the *image* of f

$$\operatorname{im} f = f(E) := \{y \in F : \text{ there exists } x \in E \text{ with } y = f(x)\}.$$

In the case $f(E) = F$, we say that f is *surjective*, or *onto*. When

$$x \neq y \Longrightarrow f(x) \neq f(y),$$

we say that f is *injective*, or 1-1 (read "one-to-one"). When both conditions hold, we say that f is *bijective*, or 1-1 onto. When there is a bijection between two sets E and F, they are *equipotent*, and this is a definition of the fact that they have the *same cardinality* (same number of elements).

[22] Exercise: Make a proof by double inclusion. Hint: Make a picture!

[23] This arrow has to be distinguished from the "convergence" arrow, or the "vector" arrow placed on top of letters.

Fundamental Sets of Numbers

We have introduced the canonical sets **N** and **Z**. Here are larger ones

Q : set of rational numbers $m/n \quad (m, n \in \mathbf{Z}, n \neq 0)$,
R : set of real numbers (scalars),
C : set of complex numbers (to be introduced later).

2.5.2 Axioms for Commutative Fields K

Scalars will here be represented by Greek letters: $\alpha, \beta, \gamma, \ldots \in K$

$$\left.\begin{array}{l} 1.\ \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma \\ 2.\ \alpha + \beta = \beta + \alpha \\ 3.\ \exists\, 0 \in K, \forall\, \alpha \in K :\ 0 + \alpha = \alpha \\ 4.\ \forall\, \alpha \in K, \exists\, -\alpha \in K :\ \alpha + (-\alpha) = 0 \end{array}\right\} \begin{array}{l} \textit{K is an} \\ \textit{additive} \\ \textit{Abelian group} \end{array}$$

$$\left.\begin{array}{l} 5.\ \alpha(\beta\gamma) = (\alpha\beta)\gamma \\ 6.\ \alpha\beta = \beta\alpha \\ 7.\ \exists\, 1\,(\neq 0) \in K, \forall\, \alpha \in K :\ 1\,\alpha = \alpha \\ 8.\ \forall\, \alpha \in K,\ \alpha \neq 0,\ \exists\, \alpha^{-1} \in K : \alpha\, \alpha^{-1} = 1 \end{array}\right\} \begin{array}{l} K - \{0\}\ \textit{is a} \\ \textit{multiplicative} \\ \textit{Abelian group} \end{array}$$

9. $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$ (*Distributivity*)

2.5.3 Axioms for Vector Spaces E

The elements of E will here carry an arrow: $\vec{\mathbf{x}}, \vec{\mathbf{y}}, \vec{\mathbf{z}}, \ldots \in E$

The sum is an *inner operation* for which E is an Abelian group:

1. $\vec{\mathbf{x}} + (\vec{\mathbf{y}} + \vec{\mathbf{z}}) = (\vec{\mathbf{x}} + \vec{\mathbf{y}}) + \vec{\mathbf{z}}$ (Associativity)
2. $\vec{\mathbf{x}} + \vec{\mathbf{y}} = \vec{\mathbf{y}} + \vec{\mathbf{x}}$ (Commutativity)
3. There is a unique $\vec{0} \in E$ such that $\vec{\mathbf{x}} + \vec{0} = \vec{\mathbf{x}}$ for all $\vec{\mathbf{x}} \in E$
4. Any $\vec{\mathbf{x}}$ has a unique opposite $-\vec{\mathbf{x}}$ such that $\vec{\mathbf{x}} + (-\vec{\mathbf{x}}) = \vec{0}$

The product by a scalar is an *external operation* satisfying

5. $1\,\vec{\mathbf{x}} = \vec{\mathbf{x}}$ (1 denotes the unit scalar)
6. $\alpha(\beta\vec{\mathbf{x}}) = (\alpha\beta)\vec{\mathbf{x}}$
7. $\alpha(\vec{\mathbf{x}} + \vec{\mathbf{y}}) = \alpha\vec{\mathbf{x}} + \alpha\vec{\mathbf{y}}$
8. $(\alpha + \beta)\vec{\mathbf{x}} = \alpha\vec{\mathbf{x}} + \beta\vec{\mathbf{x}}$

2.6 Exercises

** Let a, b, c be three elements of a vector space. (a) If a and b are independent, b and c are independent, can you prove that a and c are independent?
(b) If a and b are independent, b and c are independent, a and c are independent, can you prove that a, b, and c are independent?

1. By definition, the monomials

$$1,\ x,\ x^2, \ldots,\ x^n, \ldots$$

form a set of generators of the subspace of polynomials Π Show that the polynomials

$$1,\ x-1,\ (x-1)^2, \ldots,\ (x-1)^n, \ldots$$

also form a set of generators of the space Π (use the Taylor formula).

2. Show that the columns containing pivots of a row-reduced array are also independent. (Hint: Start by showing that in any linear combination of pivot columns producing the zero column, the coefficient of the *last* pivot column is zero.)

3. The set of linear equations $a_1x_1 + \cdots a_nx_n = b$ is a vector space V. What is its dimension? Is the subset consisting in equations having a solution a vector subspace of V?

4. Show that a space of dimension ≥ 2 (over the real field $\mathbf{R}$) is not a union of finitely many 1-dimensional subspaces.

5. (a) Show that for any field K and any set E

$$\mathcal{F}(E;K) = \{f : E \to K\}$$

is a vector space over K.

(b) Check that the set $\mathbf{F}_2 = \{0,1\}$ with addition and multiplication defined by

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

is a field (field of integers mod 2: Instead of 0, one may think of the "*even*" class, and instead of 1, one may think of the "*odd*" class).

(c) The vector space $\mathcal{F}(E;\mathbf{F}_2)$ is in 1-1 correspondence with the power set $\mathcal{P}(E)$ (set of subsets of E). To which operation on subsets do addition and multiplication of functions correspond?