

2 Basic concepts

We will assume that the following concepts are known:

- Vector, column vector, row vector, transpose. Recall that \mathbf{x} is a column vector, \mathbf{x}^t is a row vector.
- Addition of vectors (always of the same size!) and multiplication of vectors by a number (scalar).
- Scalar product (or dot product) of vectors (always of the same size). Recall that $\mathbf{x} \cdot \mathbf{y}^t = \mathbf{y} \cdot \mathbf{x}^t$ but $\mathbf{x} \cdot \mathbf{y}^t \neq \mathbf{x}^t \cdot \mathbf{y}$ (!!!) (the latter is an $n \times n$ matrix if the size of \mathbf{x} is n). Two vectors with scalar product zero are called *orthogonal*.

Remark: In many books the t superscript is omitted, i.e. $\mathbf{x} \cdot \mathbf{y}$ is used for $\mathbf{x} \cdot \mathbf{y}^t$. With this slightly abusing convention the scalar product is commutative, i.e. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

- Matrix, transpose matrix, matrix addition, multiplication of matrices by a scalar. Note that the vector is a special case of the matrix ($n \times 1$ matrix)
- Matrix-vector and matrix-matrix multiplication (sizes must match). Not commutative.
- Basic algebraic manipulations with vectors and matrices (associativity, commutativity, distributivity for addition and scalar multiplication). Null vector and null matrix (or zero matrix), identity matrix.
- $n \times 1$ vectors can be represented in the n -dimensional Euclidean space \mathbf{R}^n by arrows starting from the origin and ending at the point (x_1, x_2, \dots, x_n) , where $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$. Of course we can “see it” only for $n = 2, 3$, but this is just our limitation. In almost all cases the geometric intuition based upon the 2 or 3-dimensional picture is valid in higher dimensions. It is usually worth making pictures.

- More generally, one can consider arrows pointing from any point $A = (a_1, a_2, \dots, a_n)$ to any other point $B = (b_1, b_2, \dots, b_n)$ of the n -dimensional space. This arrow is identified with the vector $(b_1 - a_1, b_2 - a_2, \dots, b_n - a_n)$. In other words, this arrow is the vector $(b_1 - a_1, b_2 - a_2, \dots, b_n - a_n)$ with its starting point shifted to A . In most cases people talk about the vector from A to B , or the vector \vec{AB} , which is a slight abuse of language for the arrow from A to B , but it never causes any confusion.
- Geometry of vector addition and scalar multiplication (parallelogram rule)
- Length of a vector \mathbf{x} is $\|\mathbf{x}\| = \sqrt{\mathbf{x}^t \cdot \mathbf{x}} = \sqrt{x_1^2 + \dots + x_n^2}$. The vector is called normalized if its length is 1. The angle θ between two vectors \mathbf{x}, \mathbf{y} is defined via its cosine as

$$\cos \theta = \frac{\mathbf{x}^t \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

In case of $n = 2, 3$ these definitions coincide with the usual geometric notions of the length and angle. In higher dimensional Euclidean spaces these are the basic definitions of the geometry in those spaces.

- Elementary geometry of matrix multiplications in the plane and space. Scaling, rotation, projection, reflection on the level of [D] Chapter 4. and Chapter 12 of Salas-Hille.

The following concepts are also supposedly known, nevertheless we list them to fix the terminology and notation:

Definition 2.1 *The linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ with the numbers (scalars) c_1, c_2, \dots, c_k is the vector*

$$c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_k \mathbf{x}_k$$

in \mathbf{R}^n . A vector $\mathbf{y} \in \mathbf{R}^n$ is said to be a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ if there exist numbers c_1, c_2, \dots, c_k such that

$$\mathbf{y} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k. \quad (2.1)$$

In this case we also say that \mathbf{y} depend linearly on the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. It can happen that the numbers c_1, c_2, \dots, c_k are not unique in the above representation.

Definition 2.2 The **span** of the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ is the set of all possible linear combinations of these vectors, i.e. all the vectors of the form

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k$$

as c_1, c_2, \dots, c_k independently runs through the real numbers.

Definition 2.3 The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ is **linearly independent** if

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k = \mathbf{0}$$

implies $c_1 = c_2 = \dots = c_k = 0$. Otherwise this set is called **linearly dependent**.

REMARK: Linear independence or dependence is the property of a SET OF VECTORS! With a slight abuse of language we often say that the vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ are linearly dependent or independent, but keep in mind that this is a property of the ensemble of these vectors and not a property of individual vectors. It can happen that $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ is linearly independent, but adding one more vector \mathbf{x}_{k+1} to this set, the new set becomes linearly dependent.

The following statements are equivalent definitions of linear independence:

Equivalent definition 2.4 The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ is **linearly dependent** if at least one of them is expressible as a linear combination of the rest.

Equivalent definition 2.5 *The set of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ is linearly independent if any vector in the span of these vectors can be uniquely represented as a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$, i.e. the numbers c_1, c_2, \dots, c_k in (2.1) are unique.*

REMARK: The key word here is the “unique”. Every vector in the span can be represented as a linear combination of the spanning vectors, but only linearly independent spanning vectors give rise to unique representation. In particular the vector $\mathbf{y} = \mathbf{0}$ has a unique representation using $c_1 = c_2 = \dots = c_k = 0$, this is exactly the first definition.

REMARK: It cannot happen that some vectors in the span of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ have a unique representation, some others have several. Either ALL vectors in the span are uniquely represented (case of lin. independent vectors) **or** ALL vectors have many (in fact infinitely many) representations (case of lin. dependent vectors).

We already mentioned that the linear dependence/independence is a the property of the *set of vectors*. If you remove a vector from a linearly independent set, it remains lin. independent. Similarly, if you add a new element to a lin. dependent set, it remains lin. dependent. However, if you remove a vector from a lin. dependent set, both situations can occur, i.e. the smaller set could remain lin. dependent, but it also could become lin. independent (Give examples for both (*)). Similarly, if you add a new vector to a lin. dependent set, both situations can occur (again, give examples (*)).

The following fact is often useful

Lemma 2.6 *Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ be nonzero, pairwise orthogonal vectors in \mathbf{R}^n . Then they are linearly independent.*

Proof: Suppose that there is a linear combination which gives the zero vector

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \dots + c_k\mathbf{x}_k = \mathbf{0} \tag{2.2}$$

Take the scalar product of this equation with the vector \mathbf{x}_1^t

$$\mathbf{x}_1^t \cdot (c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_k \mathbf{x}_k) = \mathbf{x}_1^t \cdot \mathbf{0} = 0$$

Notice that all products $\mathbf{x}_1^t \cdot \mathbf{x}_j$ are zero by orthogonality, except for $j = 1$. Hence

$$c_1 \|\mathbf{x}_1\|^2 = 0$$

and since \mathbf{x}_1 is nonzero, we must have $c_1 = 0$. Similarly, if you multiply the equation (2.2) by \mathbf{x}_2^t , you get $c_2 = 0$ etc. Hence from (2.2) it follows that all coefficients are zero, i.e. the original vectors are linearly independent. \square

Definition 2.7 *A subset S of vectors in \mathbf{R}^n is called a **linear subspace** if it is closed under addition and scalar multiplication. This means that if $\mathbf{v}, \mathbf{w} \in S$ then $\mathbf{v} + \mathbf{w} \in S$ and $\alpha \mathbf{v} \in S$ as well for any number α .*

REMARK: In most cases we omit the word “linear” and we just refer to it as “subspace” or even only “space”. Sometimes we add that a “subspace of \mathbf{R}^n ”.

REMARK: The linear span of any set of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbf{R}^n$ is a subspace (CHECK(*)).

Definition 2.8 *A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ in a given subspace S of \mathbf{R}^n is called a **basis** of S if:*

$$S = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$$

AND $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ are linearly independent.

Notice that it is not clear that basis exists. You might think it is an unnecessary subtlety, but as we explain below, this allows us thinking of subspaces as objects like “lines”, “planes” etc. It is actually the first nontrivial theorem of linear algebra, and we will prove it below.

Theorem 2.9 *Any subspace S of \mathbf{R}^n has a basis*

REMARK: A basis is the “most economical” set to span a given subspace S in a sense that it contains the fewest possible elements. In particular, if you remove any element from a basis, then it does not span S any more (otherwise the original set were not lin. independent CHECK (*)). If, on contrary, you add a new element to a basis, then it stops being linearly independent (CHECK (*)). In fact you can use these properties to give equivalent definitions of the basis

Equivalent definition 2.10 *Any maximal set of linearly independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ in a given subspace S of \mathbf{R}^n is a basis of S . Maximality means that if you add any more vector to this set from S , it stops being linearly independent.*

Equivalent definition 2.11 *Any minimal set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ that span a given subspace S of \mathbf{R}^n is a basis of S . Minimality means that if you remove any vector from this set, then it will not span S any more.*

It is not completely trivial that these definitions are really equivalent to the original one. The key is the following lemma, which states that every linearly independent set can be extended to a basis by adding more vectors, and conversely, any set of spanning vectors can be reduced to a basis by deleting some vectors. In particular it gives good way to construct bases and in particular it proves Theorem 2.9.

Lemma 2.12 (i) *Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a set of linearly independent vectors in a nonzero subspace S of \mathbf{R}^n . Then it can be extended by adding some more vectors to form a basis of S , i.e. there exists further vectors $\mathbf{v}_{m+1}, \dots, \mathbf{v}_k$ such that $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is a basis in S .*

(ii) *Let $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a (finite) set of vectors that span a nonzero subspace S of \mathbf{R}^n . Then it can be reduced to a basis by deleting certain elements from it.*

We omit the detailed proof, but here is the idea (DO THE PROOF yourself (*)):

Part (i) If the set does not span yet, then add any vector (call \mathbf{v}_{k+1}) to it from the complement of the span. First show that this new bigger set is also linearly independent (assume not and prove by contradiction!). Then ask again if this bigger set spans or not. If yes, then you have a basis. If not, then again adjoin a new vector, call \mathbf{v}_{k+2} , show that the new set is linearly independent etc. The procedure must stop at most when you have n vectors, otherwise you would have $n + 1$ linearly independent vector in \mathbf{R}^n . This is apparently impossible. But the rigorous proof is not completely trivial, and in fact we will need the Gaussian elimination to see this (see Theorem 3.3).

Part (ii) If the given vectors are linearly independent, then we have a basis. If not, then

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_m\mathbf{v}_m = \mathbf{0}$$

for some numbers x_j , and not all them are zero. Suppose $x_j \neq 0$. Then \mathbf{v}_j can be expressed as a linear combination of the rest (CHECK (*)), hence one can remove \mathbf{v}_j from the set and the remaining vectors still span S (CHECK!) Now keep on reducing until what remains is independent. Since you started with finitely many vectors, it will happen at worst by the time we have reduced the set to a single nonzero vector. \square

DELICATE REMARK: Note that both constructions prove independently Theorem 2.9. However we needed some extra information. If we want to use (i) to prove Theorem 2.9, then in the proof of (i) we needed Theorem 3.3 (to be proven later by Gaussian elimination). It seems simpler to use (ii) to prove Theorem 2.9, i.e. start with a spanning set and reduce it to a basis. Apparently nothing extra was used, EXCEPT, that you need a finite spanning set! Infinite spanning set exists, just take all vectors in S , but if you read the proof of (ii) above carefully, then finiteness of the set was heavily used (WHY?(*)). It seems completely trivial that every subspace S of \mathbf{R}^n has a finite spanning set if you think about lines, planes etc. But

do not forget that a subspace was defined only by a few properties (Definition 2.7), and a priori it is not clear that a subspace “looks like” a line or a plane. In fact it is exactly Theorem 3.3 which proves that every subspace is spanned by a finitely many linearly independent vectors, hence it “looks like” a line or a plane or higher dimensional generalization of them.

The conclusion is that we can prove Theorem 2.9 if we EITHER know Theorem 3.3 OR we deal with subspaces with a finite spanning set.

The following crucial property follows immediately from these definitions:

Theorem 2.13 *Fix a basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ in a given subspace S of \mathbf{R}^n . Then any vector $\mathbf{v} \in S$ can be written as a linear combination of the basis vectors as*

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m$$

with some numbers c_1, c_2, \dots, c_m , and these numbers are uniquely determined. In this case we say that the vector \mathbf{v} is written in the basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ with coefficients (or coordinates) c_1, c_2, \dots, c_m .

There are many many different bases for a given subspace. However, the number of elements in all basis sets is the same. This is the second nontrivial theorem of linear algebra (for the proof see Proposition 2.4.20 of [HH])

Theorem 2.14 *Any two bases of a subspace S of \mathbf{R}^n has the same number of elements and this number is called the dimension of S .*

EXAMPLE 1: The zero vector itself, $S = \{\mathbf{0}\}$, and the full space $S = \mathbf{R}^n$ are subspaces of \mathbf{R}^n . The dimension of $S = \{\mathbf{0}\}$ is zero, the dimension of $S = \mathbf{R}^n$ is n . Sometimes these are

called trivial subspaces. The standard basis of \mathbf{R}^n is denoted by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ n \end{pmatrix}. \quad (2.3)$$

EXAMPLE 2: For $n = 2$, apart from the trivial subspaces of \mathbf{R}^2 , there are one dimensional subspaces. These are straight lines passing through the origin, i.e. they are of the form $S = \{t\mathbf{v} : t \in \mathbf{R}\}$ for some nonzero vector $\mathbf{v} \in \mathbf{R}^2$. A line that does not go through the origin is NOT a subspace (is not closed under addition).

EXAMPLE 3: For $n = 3$, apart from the trivial subspaces of \mathbf{R}^3 , there are one and two dimensional subspaces. Every straight line passing through the origin is a one dimensional subspace, they are of the form $S = \{t\mathbf{v} : t \in \mathbf{R}\}$ for some nonzero vector $\mathbf{v} \in \mathbf{R}^3$. Every plane passing through the origin is a two dimensional subspace. They can be represented as the span of their two lin. independent vectors, i.e. they are of the form $S = \{t\mathbf{v} + s\mathbf{w} : t, s \in \mathbf{R}\}$ for some linearly independent set $\{\mathbf{v}, \mathbf{w}\}$ of vectors in \mathbf{R}^3 . Naturally, these representations are not unique.

Lines or planes not passing through the origin are NOT subspaces. However, these can be represented as $\mathbf{u} + S$, where S is a line or plane and $\mathbf{u} \notin S$. These are called *affine sets*. (Sometimes people call them “affine subspaces”, but this is too much abuse of language, since they are not subspaces).

The dimension measures the “size” of a subspace; clearly a point (zero dimensional space) is “smaller” than a line (one dimensional space) which is “smaller” than a plane etc. We have

Theorem 2.15 *Let S_1 and S_2 be subspaces of \mathbf{R}^n , and suppose that $S_1 \subset S_2$. Then $\dim S_1 \leq \dim S_2$. Moreover if S_1 is strictly contained in S_2 , i.e. $S_1 \neq S_2$, then $\dim S_1 < \dim S_2$*

Proof: Start with a basis in S_1 . These vectors form a linearly independent set in S_2 , but they do not span the whole S_2 unless $S_1 = S_2$. Hence one can add more vectors to this set from S_2 to keep linear independence. In other words the number of maximal linearly independent set in S_2 is bigger than the number of basis elements of S_1 . \square

Finally we introduce the concept of orthogonal and orthonormal basis:

Definition 2.16 *A basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ of a subspace S of \mathbf{R}^n is called **orthogonal** if the basis vectors are pairwise orthogonal, i.e. $\mathbf{v}_i^t \cdot \mathbf{v}_j = 0$ for all $i \neq j$. If, in addition, the vectors are normalized, i.e. $\|\mathbf{v}_i\| = 1$, then the basis is called **orthonormal**.*

A different and bit more formal treatment of this material can be found on

<http://www.math.gatech.edu/~carlen/1502/html/pdf/dim.pdf>