Math Review

F.A. Ramponi
Rev. 0.0.2, 2020-10-13

1 Linear algebra

1.1 Why abstract vector spaces?

For sure you're already familiar with the understanding of "vectors" as "columns of numbers"; but in the course we will need a more general (and rigorous) notion.

A *field* is a set $\mathbb{F}$ with two operations $+$ (addition) and $\cdot$ (multiplication) and all the properties that one expects from "well behaving numbers". Addition is associative and commutative, there exists a number 0 such that $a + 0 = 0 + a = a$ for all $a \in \mathbb{F}$, and for every $a \in \mathbb{F}$ there exists a $a' \in \mathbb{F}$ such that $a + a' = a' + a = 0$; as usual we denote a' as $-a$ and we write $a + (-b)$ as $a - b$ to simplify notation. Addition is associative and commutative, there exists a number 1 such that $a \cdot 1 = 1 \cdot a = a$ for all $a \in \mathbb{F}$, and for every $a \in \mathbb{F}$ there exists a $a' \in \mathbb{F}$ such that $a \cdot a' = a' \cdot a = 0$; as usual we denote a' as a^{-1} and (the multiplication being commutative) we write $a \cdot b^{-1}$ as $\frac{a}{b}$ to simplify notation. Multiplication is also distributive: $a \cdot (b+c) = a \cdot b + a \cdot c$. Finally, $0 \neq 1$ (this looks kind-of obvious, but it must be stated to exclude the trivial field with only one element. In computations, we omit the dot when dealing with number: $ab = a \cdot b$.

The usual fields that one encounters in engineering are $\mathbb{R}$, and $\mathbb{C}$; just be aware that field theory is an entire branch of algebra and that, far from being the only ones, these fields are just the most obvious examples (think at $\mathbb{Q}$; think at finite fields; and there are many others).

Definition.

A <u>*vector space over a field* $\mathbb{F}$</u> is a set V with two operations $+$ (addition) and $\cdot$ (multiplication by a scalar). The operation $+ : V \times V \to V$ is a function mapping two vectors into another vector, and $\cdot : \mathbb{F} \times V \to V$ is another function mapping a scalar (i.e. a number) and a vector into another vector. As usual, we write $\mathbf{v} + \mathbf{w}$ instead of $+(\mathbf{v}, \mathbf{w})$ and $a \cdot \mathbf{v}$ instead of $\cdot(a, \mathbf{v})$. The operations must satisfy these properties:

- the operation $+$ has all the properties usually associated with addition: it is associative, commutative, there exists a vector $0 \in V$ (zero-vector) such that $0 + \mathbf{v} = \mathbf{v} + 0 = v$ for all $\mathbf{v} \in V$, and for every $\mathbf{v} \in V$ there exists an inverse $-\mathbf{v}$ with respect to addition (we write $\mathbf{w} + (-\mathbf{v})$ as $\mathbf{w} - \mathbf{v}$ to simplify notation).
- multiplication by a scalar has some properties that we expect from multiplication: it is associative, that is $(ab) \cdot \mathbf{v} = a \cdot (b \cdot \mathbf{v})$ for all $a, b \in \mathbb{F}$ and $\mathbf{v} \in V$; it is distributive in two respects, namely $(a + b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$ and $a \cdot (\mathbf{v} + \mathbf{w}) = a \cdot \mathbf{v} + a \cdot \mathbf{w}$; and finally, $1 \cdot \mathbf{v} = \mathbf{v}$ for all $v \in V$. As it is customary, we will omit the dot when dealing with multiplication by a scalar: $a\mathbf{v} = a \cdot \mathbf{v}$.

From these axioms, and enriching the notion of "abstract" vector space with further algebraic structure, one can derive all linear algebra and a good lot of geometry: you can find the juicy details in any book about linear algebra. Now you can safely forget about the rigorous definition and trust your intuition: a vector space is a set *whose elements, called vectors, are objects that can be added and multiplied by a scalar.* Just be prepared to encounter objects that are legitimate "vectors" both rigorously and intuitively, and that are not just "columns of numbers".

Examples:

- $\mathbb{R}$ with its own operations $+, \cdot$ is a vector space over itself;
- real-valued functions $f : A \to \mathbb{R}$ can be added and multiplied by a real scalar thus obtaining new functions over the same domain (with the obvious, element-wise definitions $(f_1 + f_2)(u) := f_1(u) + f_2(u)$ and $(a \cdot f)(u) := af(u)$ for all $u \in A$), hence the set of all real-valued functions can be understood as a vector space over $\mathbb{R}$;
- *continuous* real-valued functions over an interval, say $f : [a, b] \to \mathbb{R}$, can be added and multiplied by a real scalar thus obtaining new continuous functions over the same interval, hence the set of all such functions is usually understood as a vector space over $\mathbb{R}$ (and denoted $C^0[a, b]$);
- *polynomials*[1] can be added and multiplied by a scalar thus obtaining another polynomial, hence the set of all the polynomials can be understood as a vector space;
- *infinite sequences* of real numbers are none other than real-valued functions over the set $\mathbb{N}$ (or $\mathbb{Z}$, if they are infinite in both directions: the typical *signals* of discrete-time signal analysis); hence sequences can be understood as vectors;
- sequences that converge to 0 can be added and multiplied by a scalar thus obtaining another sequence that also converges to 0; hence the set of all such sequences can be understood as a vector space;
- *matrices* in $\mathbb{R}^{m \times n}$ can be added and multiplied by a real scalar in the usual way: hence $\mathbb{R}^{m \times n}$ is itself a vector space over $\mathbb{R}$;
- *random variables* can be added and multiplied by a scalar: hence we will understand them as vectors.

In saying something like “let V be a vector space over the field $\mathbb{R}$ (or over $\mathbb{C}$)”, we will always left understood that suitable operations $+$ and $\cdot$ have been defined; if even the field is omitted, we will always mean $\mathbb{F} = \mathbb{R}$.

A final note: if you look again at the previous examples, you will realize that *every* real vector space in common use is indeed a set of real-valued functions. For example, random variables are indeed real-valued functions over a so-called “sample space”. A “column vector” in $\mathbb{R}^6$ is ultimately a function over the set of indices $S = \{1, 2, 3, 4, 5, 6\}$: if we write it as a column, and not in another shape, it is because we are implicitly assuming that S itself is endowed with some other structure encoding the column shape; a different structure would yield the very same vector as a row, or maybe a 2×3 matrix, and so on. This is not very important here, but it has an useful practical consequence: since the multiplication between real numbers is commutative, the multiplication of a real number by a real-valued function also commutes. Then we can accept, *as a convention*, that the scalar$\cdot$vector product $\cdot$ is commutative as well, i.e. we accept to write $\alpha\mathbf{v} = \alpha \cdot \mathbf{v} = \mathbf{v} \cdot \alpha = \mathbf{v}\alpha$ without arousing scandal. This is common practice and never causes any problem, although strictly speaking it is an abuse of notation.

[1] More properly, polynomial *functions.*

1.2 First concepts to review

1.2.1 Subspaces, spans, and linear independence

Definition.

A *linear combination* of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a *finite* sum of scalar·vector products like

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i.$$

Let me stress it again: a *finite* sum! Series like $\sum_{i=1}^{\infty} \alpha_i \mathbf{v}_i$, whether "convergent" or not, are *not* supposed to be linear combinations (or at least: not strictly speaking).

Definition.

The *span* of a set of vectors $S \subseteq V$ is the set of all (finite!) linear combinations of vectors in S:

$$\text{span } S = \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \ : \alpha_i \text{ scalars}, \ \mathbf{v}_i \in S \right\}.$$

Definition.

A *subspace* of a vector space V is a subset $W \subseteq V$ such that, for all $\mathbf{w}_1, \mathbf{w}_2$ and $\alpha_1, \alpha_2 \in \mathbb{F}$, it holds

$$\alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 \in W.$$

Another way to say the same thing is: W is a subset of V "closed with respect to addition and multiplication"; and yet another one is: if we restrict the operations $+$ and $\cdot$ to a set W, and W with these the restricted operations happens to be a vector space by itself, then W is a subspace of V.

Examples:

- the space of *differentiable* functions over $[a, b]$ (with continuous derivative), denoted $C^1[a, b]$, is a subspace of the space of *continuous* functions $C^0[a, b]$, because every differentiable function is continuous;
- the polynomial functions restricted to $[a, b]$ form a subspace of $C^1[a, b]$;
- the sequences that converge to 0 forms a subspace of the vector space of all sequences;
- a sequence of numbers (say, infinite in both directions) $(u_i)_{i=-\infty}^{\infty}$ is called *bounded* if there exists a constant $K_u \geq 0$ such that $|u_i| \leq K_u$ for all $i \in \mathbb{Z}$; since a sum of bounded sequences is also bounded, the set of all bounded sequence is a subspace of the vector space of all sequences;

- the subset $W \subset \mathbb{R}^{2\times 2}$ of matrices with this particular form[2],

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

is a subspace of $\mathbb{R}^{2\times 2}$.

Fact. The zero-vector $0 \in V$ must belong to any subspace of V.

Fact.

If S is an arbitrary non-empty subset of V, then span S is a subspace of V. This is evident if $S = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is finite: indeed any linear combination of linear combinations of the same vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is itself a linear combination of $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$. (Just group terms.)

Definition.

A set $S \subseteq V$ is said to *generate* V if $V = \text{span } S$. In words, if every vector in V can be obtained as a (finite!) linear combination of vectors in S. A vector space V is called *finitely generated* if there exists a *finite* set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$ such that $V = \text{span } S$. In words, if every vector in $\mathbf{v} \in V$ can be obtained as a linear combination of $\mathbf{v} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_n\mathbf{v}_n$. Unless otherwise stated, from now on we restrict our attention to finitely generated vector spaces.

Example. Assume that V is the set of polynomials with degree at most 2; then the polynomials $\mathbf{v}_1(x) = 1, \mathbf{v}_2(x) = x - 1, \mathbf{v}_3(x) = x^2, \mathbf{v}_4(x) = x + x^2$ generates V, because any polynomial in V has the form $\mathbf{v}(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2$, and hence it can be obtained as the linear combination $\mathbf{v}(x) = (\alpha_1 + \alpha_2)\mathbf{v}_1(x) + \alpha_2\mathbf{v}_2(x) + \alpha_3\mathbf{v}_3(x)$. By the way, it can be obtained also as the linear combination $\mathbf{v}(x) = (\alpha_1 + \alpha_2 - c)\mathbf{v}_1(x) + (\alpha_2 - \alpha_3)\mathbf{v}_2(x) + \alpha_3\mathbf{v}_4(x)$, i.e. the linear combination that yields $\mathbf{v}$ is by no means supposed to be unique.

Definition.

The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called *linearly independent in V* if their only linear combination that yields the zero-vector $0 \in V$ is the one with all coefficients equal to 0:

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_n\mathbf{v}_n = 0 \quad \Leftrightarrow \quad \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0.$$

In the opposite case, i.e. if there exists a linear combination such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_n\mathbf{v}_n = 0$$

where *at least one* $\alpha_i \neq 0$, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are called *linearly dependent*. An arbitrary set $S \subset V$ is said to be *linearly independent* if, however we choose finitely many vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S$, they are linearly independent. (Otherwise, it is called linearly dependent.)

[2] Every matrix with this form can be written as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = aI + bJ,$$

where it so happens that $J^2 = -I$. Do you recognize the subspace W?

Example. Let as before V be the set of polynomials with degree at most 2; then the polynomials $\mathbf{v}_1(x) = 1, \mathbf{v}_2(x) = x - 1, \mathbf{v}_3(x) = x^2$ are linearly independent. Indeed if $\alpha_1\mathbf{v}_1(x) + \alpha_2\mathbf{v}_2(x) + \alpha_3\mathbf{v}_3(x) = (\alpha_1 - \alpha_2) + \alpha_2 x + \alpha_3 x^2$. The fundamental theorem of algebra (every polynomial with degree ≥ 1 has a root in $\mathbb{C}$) and Ruffini's theorem ($\bar{x}$ is a root of $\mathbf{p}(x)$ if and only if $(x - \bar{x})$ divides $\mathbf{p}$) imply that any polynomial in V, except for the polynomial 0, has exactly 2 roots (in $\mathbb{C}$; and *at most* 2 roots in $\mathbb{R}$); on the other hand every x is a root of the polynomial 0. Hence the only possibility is that $(\alpha_1 - \alpha_2) + \alpha_2 x + \alpha_3 x^2$ is itself the polynomial 0, and this implies $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

1.2.2 Bases

Definition.

A set of vectors $S \subset V$ is called a *basis* of V if it is linearly independent and it generates V.

Assume that V is finitely generated. The property that a finite set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ generates V and the property that a set $S \subset V$ is linearly independent in V are "dual" of each other, in the sense expressed by the following facts:

Fact.

If a finite set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \subset V$ generates V, then it is always possible to extract from S a linearly independent set $B = \{\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_n}\}$ that still generates V. In other words, *every finitely generated vector space has a basis*[a].

[a] The true general theorem says: *every vector space has a basis*, the restriction of the thesis to *finitely generated* vector spaces being the most interesting for our purposes. On the other hand, one could think that the extension to general vector spaces is, after all, innocuous. This is far from the truth: while the "fact" above is very intuitive and has a somewhat tedious, but otherwise straightforward proof, the general statement has *astonishing* consequences, and its proof is so abstract that it needs tools from the very foundations of axiomatic set theory.

Fact.

If a finite set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent in V, then it is always possible to add some vectors $\{\mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\} \subset V$ to S in such a way that the union $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}, \ldots, \mathbf{v}_n\}$ is still linearly independent *and generates* V. In other words, *every linear independent set in V can be extended to a basis.*

It is not a case that I have used the same letter n for the number of vectors in the above two "facts". Indeed:

Fact.

Fact. Every basis of the same finitely-generated space V has the same number of elements $n \in \mathbb{N}$. The number n is called the *dimension* of V and denoted $\dim V$. From now on, instead of the cryptic attribute "finitely generated" we will use the the standard, equivalent, terminology: V is *finite-dimensional*, or more precisely *n-dimensional.*

Example. Consider $\mathbb{R}^4$ as a vector space over $\mathbb{R}$. The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

generate $\mathbb{R}^4$ and are linearly independent. Hence they form a basis of $\mathbb{R}^4$ and, as expected, $\dim \mathbb{R}^4 = 4$. (The basis of $\mathbb{R}^n$ whose elements have a 1 at just one coordinate and 0 at every other ccordinate is called the *canonical basis.*)

Example. Consider $\mathbb{C}^2$ as a vector space over the field $\mathbb{C}$. The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

form a basis of $\mathbb{C}^2$ and, as expected, $\dim \mathbb{C}^2 = 2$ (over $\mathbb{C}$).

Example. Consider $\mathbb{C}^2$ as a vector space *over* $\mathbb{R}$. The vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} j \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ j \end{bmatrix}$$

generate $\mathbb{C}^2$ and are linearly independent. Hence they form a basis of $\mathbb{C}^2$ (over $\mathbb{R}$), and $\dim \mathbb{C}^2 = 4$ (over $\mathbb{R}$!).

Example. Let V be the space of polynomials over $\mathbb{R}$ with degree at most 3. The vectors/polynomials

$$\mathbf{v}_1(x) = 1, \quad \mathbf{v}_2(x) = x, \quad \mathbf{v}_3(x) = x^2, \quad \mathbf{v}_4(x) = x^3$$

form a basis of V, hence $\dim V = 4$. Indeed V is a 4-dimensional subspace of the space of all polynomials.

Fact.

Suppose that $\dim V = n$. Then any $n+1$ vectors in V are linearly *dependent.*

Example. Let V be the space of infinite sequences $(v_i)_{i=1}^{\infty}$ over $\mathbb{R}$. The vectors

$$\begin{aligned}
\mathbf{v}_1 &= (1, 0, 0, 0, 0, 0, 0, \ldots), \\
\mathbf{v}_2 &= (0, 1, 0, 0, 0, 0, 0, \ldots), \\
\mathbf{v}_3 &= (0, 0, 1, 0, 0, 0, 0, \ldots), \\
\mathbf{v}_4 &= (0, 0, 0, 1, 0, 0, 0, \ldots), \\
&\vdots \\
\mathbf{v}_i &= (0, \ldots, 0, 1, 0, \ldots, 0, \ldots), \\
&\vdots
\end{aligned}$$

are linearly independent: indeed if we set to 0 any (finite!) linear combination of an arbitrary number of them, say

$$\begin{aligned}\alpha_1\mathbf{v}_{i_1}+\alpha_2\mathbf{v}_{i_2}+\ldots+\alpha_m\mathbf{v}_{i_m} &= (0,0,\alpha_1,0,\alpha_2,0,0,\ldots,0,\alpha_m,0,0,\ldots)\\ &= (0,0,0,0,0,0,0,0,0,0,0,0,\ldots),\end{aligned}$$

then it must hold $\alpha_i = 0$ for all $i = 1,\ldots,m$. Now assuming that V is finite-dimensional would mean implying that it admits a basis of n vectors, and hence that any $n+1$ vectors are linearly dependent. But we have just shown that we can find *arbitrarily many* ($m \geq n+1$) linearly independent vectors in V, hence such assumption would be contradictory. In fact, V is *not* finite-dimensional. It is an "*infinite-dimensional*" space[3].

Fact.

Suppose that V is n-dimensional. The fundamental property of a basis $\{\mathbf{b}_1,\ldots,\mathbf{b}_n\}$ of V is that *every vector* $\mathbf{v}\in V$ *can be expressed as a linear combination of* $\mathbf{b}_1,\ldots,\mathbf{b}_n$ *in a unique way* (i.e. the coefficients of the linear combination are determined unambiguously).

Indeed, the fact that *every* vector $\mathbf{v}$ can be written as a linear combination is because the basis generates V; on the other hand, suppose that we write $\mathbf{v}$ in two ways:

$$\begin{aligned}\mathbf{v} &= \alpha_1\mathbf{b}_1+\alpha_2\mathbf{b}_2+\ldots+\alpha_n\mathbf{b}_n,\\ \mathbf{v} &= \beta_1\mathbf{b}_1+\beta_2\mathbf{b}_2+\ldots+\beta_n\mathbf{b}_n.\end{aligned}$$

Then

$$\begin{aligned}0 = \mathbf{v}-\mathbf{v} &= (\alpha_1\mathbf{b}_1+\alpha_2\mathbf{b}_2+\ldots+\alpha_n\mathbf{b}_n)-(\beta_1\mathbf{b}_1+\beta_2\mathbf{b}_2+\ldots+\beta_n\mathbf{b}_n)\\ &= (\alpha_1-\beta_1)\mathbf{b}_1+(\alpha_2-\beta_2)\mathbf{b}_2+\ldots+(\alpha_n-\beta_n)\mathbf{b}_n :\end{aligned}$$

but the basis is by definition also linearly independent, and the only linear combination that yields 0 is the one with null coefficients, i.e. $(\alpha_i-\beta_i)=0$, i.e. $\alpha_i=\beta_i$ for all $i=1,\ldots,n$. Hence, besides the different symbols used to denote the coefficients, the linear combination is indeed unique.

Example. The set V of *sinusoids* with a fixed frequency $\bar{\omega} = 2\pi\bar{F}$ considered as $\mathbb{R}\to\mathbb{R}$ functions, forms a 2-dimensional space (a subspace of $C^1(\mathbb{R})$). Indeed, any such sinusoid $\mathbf{s}(t)$ with amplitude A and phase ϕ can be written as

$$\begin{aligned}\mathbf{s}(t) &= A\sin(\bar{\omega}t+\phi)\\ &= (A\cos\phi)\sin(\bar{\omega}t)+(A\sin\phi)\cos(\bar{\omega}t)\\ &= \alpha\sin(\bar{\omega}t)+\beta\cos(\bar{\omega}t),\end{aligned}$$

that is, the two sinusoids $\mathbf{s}_1(t)=\sin(\bar{\omega}t)$ and $\mathbf{s}_2(t)=\cos(\bar{\omega}t)$ generate V. On the other hand, we have

$$\sqrt{\alpha^2+\beta^2} = \sqrt{A^2\cos^2\phi+A^2\sin^2\phi} = A,$$
$$\frac{\beta}{\alpha} = \frac{A\sin\phi}{A\cos\phi} = \tan\phi.$$

[3] Please refrain from guessing that the sequences $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_i,\ldots$ defined above, with just one term equal to 1 and all other terms equal to 0, form a basis of V. *They do not*, because "infinite linear combinations" of the kind $\sum_{i=1}^{\infty}\alpha_i\mathbf{v}_i$ are not allowed and strictly speaking do not make any sense. A basis of V does exist in the abstract set-up, but it is not this one. The sequences $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_i,\ldots$ form indeed a basis of the space W of infinite sequences *with finitely many non-zero terms*, but W is not the entire V.

This shows on one hand that $\mathbf{s}_1$ and $\mathbf{s}_2$ are linearly independent (i.e. they form a basis of V), because imposing

$$0 = \alpha \sin(\bar{\omega}t) + \beta \cos(\bar{\omega}t)$$

is equivalent to impose $A = \sqrt{\alpha^2 + \beta^2} = 0$, and hence $\alpha = \beta = 0$; on the other hand it shows that (A, ϕ) determine (α, β) uniquely, *and vice versa*, because $\phi = \arctan_2(\alpha, \beta)$.

1.3 Re-think matrix products!

I give for granted that you know the meaning of the product of a "row vector" by a "column vector" (of equal length) as second-nature: if $v, w \in \mathbb{R}^n$,

$$v^\top w = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n = \sum_{i=1}^n v_i w_i.$$

Now I invite you to apply the same rule, "sum of component-wise products" to "rows" and "columns" containing objects of *any* nature, insofar as rows and columns have the same length and insofar as sums and products make sense. Let the following be the scheme of a "practical" definition:

$$\begin{bmatrix} \square_1 & \square_2 & \cdots & \square_n \end{bmatrix} \begin{bmatrix} \diamondsuit_1 \\ \diamondsuit_2 \\ \vdots \\ \diamondsuit_n \end{bmatrix} = \square_1 \diamondsuit_1 + \square_2 \diamondsuit_2 + \ldots + \square_n \diamondsuit_n = \sum_{i=1}^n \square_i \diamondsuit_i,$$

whatever $\square_i$ and $\diamondsuit_i$ may mean. In other words, let this rule hold whenever a product like $\square_i \diamondsuit_i$ and sums like $\square_1 \diamondsuit_1 + \square_2 \diamondsuit_2$ make sense. I promise it will just work; and you will have fun whenever you encounter new applications.

Example. Multiplying a row of vectors by a column of scalars (or a row of scalars by a column of vectors) you will obtain a linear combination of the vectors with the scalars as coefficients[4]:

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \mathbf{v}_1 \alpha_1 + \mathbf{v}_2 \alpha_2 + \ldots + \mathbf{v}_n \alpha_n = \sum_{i=1}^n \alpha_i \mathbf{v}_i.$$

Now think at the product of a $m \times n$ matrix by a column vector in $\mathbb{R}^n$. The standard way of seeing the product is as a stack of row×column products, in this way:

$$Av = \begin{bmatrix} — & r_1 & — \\ — & r_2 & — \\ & \vdots & \\ — & r_m & — \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix} = \begin{bmatrix} r_1 v \\ r_2 v \\ \vdots \\ r_m v \end{bmatrix}.$$

Think differently! You may interpret the matrix also as a *row of columns*, and the column vector as a column of *coefficients*. Applying the above "rule", the result happens to be a linear

[4] Here and below I use the convention $\alpha \cdot \mathbf{v} = \mathbf{v} \cdot \alpha$ without scandalizing anybody.

combination of the columns:

$$Av = \begin{bmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} | \\ c_1 \\ | \end{bmatrix} \alpha_1 + \begin{bmatrix} | \\ c_2 \\ | \end{bmatrix} \alpha_2 + \ldots + \begin{bmatrix} | \\ c_n \\ | \end{bmatrix} \alpha_n = \sum_{i=1}^{n} \alpha_i \begin{bmatrix} | \\ c_i \\ | \end{bmatrix}.$$

Example. Let V be any n-dimensional real vector space, and let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of V. Then every $\mathbf{v} \in V$ can be written *in a unique way* as

$$\mathbf{v} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

for some column of coefficients v_i. Once the basis is fixed, you can forget about it and work only with column vectors in $\mathbb{R}^n$. This is the reason why $\mathbb{R}^n$ is all over the place even when the "true" vectors have a completely different nature than that of a column. To tell the whole story, this reasoning applies also to $\mathbb{R}^n$ itself: *to use a column of coefficients, i.e. of coordinates, by definition a basis must have been fixed beforehand.* If in a linear-algebraic problem involving $\mathbb{R}^n$ the basis is never made explicit, you can take for understood that it is the canonical one:

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{b}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Stacked in a row, the basis yields the identity matrix:

$$\mathbf{v} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

And this is why in "good ol' $\mathbb{R}^n$" the basis is not even worth mentioning. But it is there.

There's more. Think at a product of two matrices $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times m}$. Sure, the result is the matrix $AB = M \in \mathbb{R}^{n \times m}$ whose ij-th entry is $m_{ij} = \sum_{h=1}^{k} a_{ih} b_{hj}$. But this is complicated stuff with too many boring indices. Think differently! Think A = a *row of k columns*, and B = a *column of k rows*: then

$$AB = \begin{bmatrix} | & | & & | \\ c_1 & c_2 & \cdots & c_k \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ & \vdots & \\ - & r_k & - \end{bmatrix} = \sum_{i=1}^{k} \begin{bmatrix} | \\ c_i \\ | \end{bmatrix} \begin{bmatrix} - & r_i & - \end{bmatrix}.$$

The result fits nicely (and coincides with the standard definition) because each product $c_i r_i$ makes perfect sense and yields an $n \times m$ matrix.

The above way of interpreting row×column products can be generalized to *matrix* products; for example we can multiply block-partitioned matrices in the same way as we multiply matrices of numbers:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix}\begin{bmatrix} E & F \end{bmatrix} + \begin{bmatrix} B \\ D \end{bmatrix}\begin{bmatrix} G & H \end{bmatrix}$$

As long as the blocks have compatible dimensions in such a way that products and sums make sense, the result is the same that we would obtain without partitioning: in my opinion, this flexibility in interpreting products is the true reason why matrices are so useful.
Exercise: prove, in 5 seconds, that the product of block-triangular matrices ($C = G = 0$) is itself block-triangular.

1.4 Linear maps and their representations

1.4.1 Linear maps

Definition.

Definition. A *linear map* (or *linear function*, or sometimes *linear operator*: it's the same) between two vector spaces V, W is a function $\mathcal{A} : V \to W$ such that

$$\begin{aligned} \mathcal{A}(\mathbf{v}_1 + \mathbf{v}_2) &= \mathcal{A}(\mathbf{v}_1) + \mathcal{A}(\mathbf{v}_2), \\ \mathcal{A}(\alpha \cdot \mathbf{v}) &= \alpha \cdot \mathcal{A}(\mathbf{v}) \end{aligned} \tag{1}$$

for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v} \in V$ and all $\alpha \in \mathbb{F}$. In words, a linear map is a function that preserves the vector space structure[a].

[a] Pay attention: the operations $+, \cdot$ in the left-hand sides of (1) are the operations of V, while the operations $+, \cdot$ in the right-hand sides are the operations of W.

Fact. The definition generalizes naturally to linear combinations:

$$\mathcal{A}(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_n\mathbf{v}_n) = \alpha_1\mathcal{A}(\mathbf{v}_1) + \alpha_2\mathcal{A}(\mathbf{v}_2) + \ldots + \alpha_n\mathcal{A}(\mathbf{v}_n),$$

that is, in symbolic notation:

$$\mathcal{A}\left(\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}\right) = \begin{bmatrix} \mathcal{A}(\mathbf{v}_1) & \cdots & \mathcal{A}(\mathbf{v}_n) \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}.$$

Examples:

- of course you already know that the multiplication of a matrix $A \in \mathbb{R}^{m\times n}$ by a column vector $v \in \mathbb{R}^n$ is the application of a linear map to a vector; in other words, a matrix can always be interpreted as a linear map;
- the *derivative operator* $\frac{d}{dt}$ is a linear map from $C^1[a,b]$ to $C^0[a,b]$;
- the *integral* $\mathrm{Int} : C^0[a,b] \to \mathbb{R}$,

  $$\mathrm{Int}(f) = \int_a^b f(t)\ dt,$$

  is a linear map;

- a *discrete-time linear system* $\Sigma : \mathcal{U} \to \mathcal{Y}$, where $\mathcal{U}$ and $\mathcal{Y}$ are two suitable vector spaces of "signals", i.e. *sequences of numbers* infinite in both directions, is a *continuous* linear map[5].

  Let B be the subspace (of both $\mathcal{U}$ and $\mathcal{Y}$) of *bounded signals.* If it is possible to restrict Σ to $B \subseteq \mathcal{U}$ in such a way that $\Sigma : B \to B$, that is, if Σ maps bounded signals to bounded signals, then it is called a "BIBO-stable" system.

1.4.2 Representation of a linear map (finite-dimensional spaces)

Let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis of the n-dimensional vector space V, let $\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m\}$ a basis of the m-dimensional vector space W, and $\mathcal{A} : V \to W$ a linear map. The behavior of $\mathcal{A}$ is captured by its action on the basis of V. Indeed,

$$\mathcal{A}(\mathbf{b}_1) = \mathbf{w}_1 = a_{11}\boldsymbol{\beta}_1 + a_{21}\boldsymbol{\beta}_2 + \cdots + a_{m1}\boldsymbol{\beta}_m = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

$$\mathcal{A}(\mathbf{b}_2) = \mathbf{w}_2 = a_{12}\boldsymbol{\beta}_1 + a_{22}\boldsymbol{\beta}_2 + \cdots + a_{m2}\boldsymbol{\beta}_m = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

$$\vdots$$

$$\mathcal{A}(\mathbf{b}_n) = \mathbf{w}_n = a_{1n}\boldsymbol{\beta}_1 + a_{2n}\boldsymbol{\beta}_2 + \cdots + a_{mn}\boldsymbol{\beta}_m = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

for suitable coefficients $a_{11}, \ldots, a_{mn}$; and stacking $\mathcal{A}(\mathbf{b}_1), \ldots, \mathcal{A}(\mathbf{b}_n)$ in a row is the same as stacking the columns with coefficients a_{ij} on the right hand sides:

$$\begin{bmatrix} \mathcal{A}(\mathbf{b}_1) & \mathcal{A}(\mathbf{b}_2) & \cdots & \mathcal{A}(\mathbf{b}_n) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Now, we know that any vector $\mathbf{v} \in V$ admits a unique representation as a linear combination of $\mathbf{b}_1, \ldots, \mathbf{b}_n$ (denote v_j the coefficients), and any $\mathbf{w} \in W$ admits a unique representation as a

[5] The fact that a linear map is *continuous* is almost a triviality if it is a function between finite-dimensional spaces; not so if it is a function between infinite-dimensional ones. A linear map between infinite-dimensional spaces may fail to be continuous. (Now recall that sequences, infinite in one direction or both, indeed form an infinite-dimensional space.) This is a rather subtle point, so don't worry about it too much

linear combination of $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$ (denote w_i the coefficients); therefore

$$\begin{aligned}
\mathcal{A}(\mathbf{v}) &= \mathcal{A}\left(\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) \\
&= \begin{bmatrix} \mathcal{A}(\mathbf{b}_1) & \mathcal{A}(\mathbf{b}_2) & \cdots & \mathcal{A}(\mathbf{b}_n) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\
&:= \begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \cdots & \boldsymbol{\beta}_m \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \mathbf{w}.
\end{aligned}$$

We conclude the following

Fact.

Once the basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of V and the basis $\{\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m\}$ of W are fixed, $\mathcal{A}$ is represented unambiguously by a matrix in $A \in \mathbb{R}^{m \times n}$, and the relation $\mathcal{A}(\mathbf{v}) = \mathbf{w}$ is represented unambiguously by a matrix multiplication:

$$Av = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = w.$$

Fact.

Let the maps $\mathcal{B} : U \to V$ and $\mathcal{A} : V \to W$ (between finite-dimensional spaces) be represented by the matrices A and B respectively. Then their composition $\mathcal{A} \circ \mathcal{B}$, defined by

$$(\mathcal{A} \circ \mathcal{B})(\mathbf{u}) = \mathcal{A}(\mathcal{B}(\mathbf{u})),$$

is represented by the matrix AB.

Indeed, fix bases $\{\mathbf{b}_i^{(u)}\}$, $\{\mathbf{b}_i^{(v)}\}$, $\{\mathbf{b}_i^{(w)}\}$ of the three spaces U, V, W respectively, and let u, v, w denote the column of coefficients (i.e. coordinates) of vectors $\mathbf{u} \in U$, $\mathbf{v} \in V$, and $\mathbf{w} \in W$

respectively. We have

$$
\begin{aligned}
\mathbf{w} = \mathcal{A}(\mathcal{B}(\mathbf{u})) &= \mathcal{A}\left(\mathcal{B}\left(\begin{bmatrix} \mathbf{b}_1^{(u)} & \cdots & \mathbf{b}_l^{(u)} \end{bmatrix}\begin{bmatrix} u_1 \\ \vdots \\ u_l \end{bmatrix}\right)\right) \\
&= \mathcal{A}\left(\begin{bmatrix} \mathbf{b}_1^{(v)} & \cdots & \mathbf{b}_n^{(v)} \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nl} \end{bmatrix}\begin{bmatrix} u_1 \\ \vdots \\ u_l \end{bmatrix}\right) \\
&= \begin{bmatrix} \mathbf{b}_1^{(w)} & \cdots & \mathbf{b}_m^{(w)} \end{bmatrix}\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nl} \end{bmatrix}\begin{bmatrix} u_1 \\ \vdots \\ u_l \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{b}_1^{(w)} & \cdots & \mathbf{b}_m^{(w)} \end{bmatrix}\begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix},
\end{aligned}
$$

that is, $w = ABu$ is the unambiguous representation of $\mathbf{w} = \mathcal{A}(\mathcal{B}(\mathbf{u}))$ in terms of matrices and columns of coordinates.

Definition.

The *identity map* over a finite-dimensional space V is the map $\mathcal{I} : V \to V$ defined by $\mathcal{I}(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$. Let $\mathcal{A} : V \to V$ be a linear map. If there exists a linear map $\mathcal{A}^{-1} : V \to V$ such that $\mathcal{A} \circ \mathcal{A}^{-1} = \mathcal{A}^{-1} \circ \mathcal{A} = \mathcal{I}$, then $\mathcal{A}^{-1}$ is called the *inverse* of $\mathcal{A}$.

Fact.

Assume that $\mathcal{A}$ is represented by the matrix A. If $\mathcal{A}^{-1}$ exists, then the inverse of A also exists, and $\mathcal{A}^{-1}$ is represented by A^{-1}.

Indeed, it is immediate to recognize that $\mathcal{I}$ is represented by the identity matrix I. If $\mathcal{A}^{-1}$ exists, then it is represented unambiguously by a matrix A', and it follows immediately that the identity $\mathcal{A} \circ \mathcal{A}^{-1} = \mathcal{A}^{-1} \circ \mathcal{A} = \mathcal{I}$ is represented by $AA' = A'A = I$: then $A' = A^{-1}$.

I hope that you are now fully convinced that, when dealing *only* with finite-dimensional spaces, working with linear maps and their properties is equivalent to work with matrices.

1.5 Euclidean geometry

1.5.1 Distances and norms

Definition.

A *distance* in a set V is a function $d : V \times V \to \mathbb{R}$ that satisfies the following properties (for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$:

- $d(\mathbf{v}, \mathbf{w}) \geq 0$, and $d(\mathbf{v}, \mathbf{w}) = 0$ if and only if $\mathbf{v} = \mathbf{w}$;
- $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v}) = 0$ (symmetry);
- $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$ (triangular inequality).

The set S, endowed with a distance, is called a *metric space*.

Example. Note that here I am not yet assuming that V is a vector space; indeed, attached to V there may be no notion of addition and multiplication by a scalar at all. For example, let V be a connected undirected graph: one can define the distance between two vertices $\mathbf{v}$, $\mathbf{w}$ as the minimum number of edges necessary to travel from $\mathbf{v}$ to $\mathbf{w}$ (and of course $d(\mathbf{v}, \mathbf{v}) = 0$ for all vertices $\mathbf{v}$), and this turns out to be a well-defined distance over V.

Example. The natural distance in $\mathbb{R}$ is $d(\mathbf{v}, \mathbf{w}) = |\mathbf{v} - \mathbf{w}|$.

Example. A trivial distance (over any non-empty set) is defined by $d(\mathbf{v}, \mathbf{w}) = 0$ if $\mathbf{v} = \mathbf{w}$, and $d(\mathbf{v}, \mathbf{w}) = 1$ otherwise.

Definition.

A *norm* over a vector space V (real or complex) is a function $\| \cdot \| : V \to \mathbb{R}$ that satisfies these properties:

- $\|\mathbf{v}\| \geq 0$ for all $\mathbf{v} \in V$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = 0$;
- $\|\alpha\mathbf{v}\| = |\alpha| \, \|\mathbf{v}\|$ for all $\mathbf{v} \in V$ and $\alpha \in \mathbb{R}$;
- $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$ (triangular inequality).

Examples:

- $\|\mathbf{v}\| = |\mathbf{v}|$ is a norm over $\mathbb{R}$ (considered as a vector space over itself);
- for any p, $1 \leq p < \infty$, the function

  $$\|v\|_p = \sqrt[p]{\sum_{i=1}^{n} |v_i|^p}$$

  is a norm over $\mathbb{R}^n$;
- the function

  $$\|v\|_\infty = \max_{i=1,\dots,n} |v_i|$$

  is also a norm over $\mathbb{R}^n$, called the Chebycheff norm;

- if V is any n-dimensional vector space where a basis has been fixed and we let $v \in \mathbb{R}^n$ contain the coordinates of a vector $\mathbf{v}$ with respect to that basis, then $\|\mathbf{v}\|_p = \|v\|_p$ is a norm over V (for any p, $1 \leq p \leq \infty$).

Examples (infinite-dimensional spaces):

- let V be the space of *bounded* sequences $\mathbf{u} = (u_i)_{i=-\infty}^{\infty}$; then the function
$$\|\mathbf{u}\|_\infty = \min \{K \geq 0 \ : \ |u_i| \leq K \text{ for all } i \in \mathbb{Z}\}$$
is a norm over V;
- the function
$$\|f\|_1 = \int_a^b |f(t)| \, dt$$
is a norm over $C^0[a, b]$.

Fact.

If $\|\cdot\| : V \to \mathbb{R}$ is a norm, then the function defined by $d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$ is a distance over V.

Indeed, the first property of a distance is obvious; symmetry holds because
$$d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\| = \|(-1)(\mathbf{w} - \mathbf{v})\| = |-1| \ \|\mathbf{w} - \mathbf{v}\| = d(\mathbf{w}, \mathbf{v});$$
and the triangular inequality holds because
$$d(\mathbf{u}, \mathbf{w}) = \|\mathbf{u} - \mathbf{w}\| = \|\mathbf{u} - \mathbf{v} + \mathbf{v} - \mathbf{w}\| \leq \|\mathbf{u} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{w}\| = d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}).$$

1.5.2 Scalar products

Definition.

Let V be a vector space over the field $\mathbb{R}$. A *scalar product*[a] in V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies the following properties (for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha, \beta \in \mathbb{R}$):

- $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$ (positive definiteness);
- $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ (symmetry);
- $\langle \mathbf{u}, \ \alpha\mathbf{v} + \beta\mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle + \beta \langle \mathbf{u}, \mathbf{w} \rangle$ (linearity).

[a] Distinguish carefully between "scalar product", which is a product between vectors, and "product by a scalar", which is the second operation of the vector space.

It is quite easy to check that, because of symmetry, linearity holds also with respect to the *first* argument, i.e. $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \ \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$.

Fact.

(*Cauchy-Schwarz inequality*). Define the quantity

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}. \tag{2}$$

For all $\mathbf{v}, \mathbf{w} \in V$, it holds

$$|\langle \mathbf{v}, \mathbf{w}\rangle| \leq \|\mathbf{v}\|\ \|\mathbf{w}\|.$$

Proof. The inequality is obvious $(0 = 0)$ if $\mathbf{w} = 0$. Thus, assume that $\mathbf{w} \neq 0$ and let $\alpha = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{w}\|^2}$. It follows:

$$\begin{aligned} 0 \leq \|\mathbf{v} - \alpha\mathbf{w}\|^2 &= \langle \mathbf{v} - \alpha\mathbf{w},\ \mathbf{v} - \alpha\mathbf{w}\rangle \\ &= \langle \mathbf{v}, \mathbf{v}\rangle - \alpha \langle \mathbf{v}, \mathbf{w}\rangle - \alpha \langle \mathbf{w}, \mathbf{v}\rangle + \alpha^2 \langle \mathbf{w}, \mathbf{w}\rangle \\ &= \|\mathbf{v}\|^2 - 2\alpha \langle \mathbf{v}, \mathbf{w}\rangle + \alpha^2\|\mathbf{w}\|^2 \\ &= \|\mathbf{v}\|^2 - 2\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{w}\|^2}\ \langle \mathbf{v}, \mathbf{w}\rangle + \frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\|\mathbf{w}\|^4}\ \|\mathbf{w}\|^2 \\ &= \|\mathbf{v}\|^2 - \frac{\langle \mathbf{v}, \mathbf{w}\rangle^2}{\|\mathbf{w}\|^2}; \end{aligned}$$

therefore $\langle \mathbf{v}, \mathbf{w}\rangle^2 \leq \|\mathbf{v}\|^2\ \|\mathbf{w}\|^2$, and the claim follows.

Fact.

If $\langle\cdot,\cdot\rangle$ is a scalar product in V, then the function $\|\cdot\|$ defined by (2) is a norm on V. It is called the *Euclidean norm* induced by the scalar product.

Indeed, the first property in the definition of norm is obvious; the second one holds because, by linearity with respect to both arguments,

$$\|\alpha\mathbf{v}\| = \sqrt{\langle \alpha\mathbf{v}, \alpha\mathbf{v}\rangle} = \sqrt{\alpha^2 \langle \mathbf{v}, \mathbf{v}\rangle} = |\alpha|\sqrt{\langle \mathbf{v}, \mathbf{v}\rangle} = |\alpha|\ \|\mathbf{v}\|;$$

the triangular inequality holds because of Cauchy-Schwarz inequality. Indeed:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w},\ \mathbf{v} + \mathbf{w}\rangle \\ &= \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle + \langle \mathbf{w}, \mathbf{v}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle \\ &= \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w}\rangle + \|\mathbf{w}\|^2 \\ &\leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\|\ \|\mathbf{w}\| + \|\mathbf{w}\|^2 \\ &= (\|\mathbf{v}\| + \|\mathbf{w}\|)^2. \end{aligned}$$

Explanation. The Cauchy-Schwarz inequality (along its implication: $\|\cdot\|$ is a norm) is of paramount importance in all branches of mathematics. Often it is employed in the following form:

$$\frac{|\langle \mathbf{v}, \mathbf{w}\rangle|}{\|\mathbf{v}\|\ \|\mathbf{w}\|} \leq 1, \qquad \text{that is,} \qquad -1 \leq \frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{v}\|\ \|\mathbf{w}\|} \leq 1.$$

In this forms it allows us to define the *angle between two vectors* $\mathbf{v}, \mathbf{w}$ as

$$\text{angle between } \mathbf{v} \text{ and } \mathbf{w} := \arccos\left(\frac{\langle \mathbf{v}, \mathbf{w}\rangle}{\|\mathbf{v}\|\ \|\mathbf{w}\|}\right)$$

and bring geometry in territories where you would not normally expect it (because all this section holds in full generality): if we define scalar products between *real functions* or between *real random variables*, as one normally does in signal theory and probability, then we recover such things as *angles between functions* and *angles between random variables.* Besides, the above definition is consistent with the notion of "scalar product" that you learned in basic geometry and/or physics:

$$\langle \mathbf{v}, \mathbf{w}\rangle = \|\mathbf{v}\| \cdot \|\mathbf{w}\| \cdot \cos(\text{angle between } \mathbf{v} \text{ and } \mathbf{w}).$$

Definition.

(Scalar product, complex case.) Let V be a vector space over the field $\mathbb{C}$. A scalar product in V is a function $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{C}$ that satisfies the following properties (for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha, \beta \in \mathbb{C}$):

- $\langle \mathbf{v}, \mathbf{v}\rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v}\rangle = 0$ if and only if $\mathbf{v} = 0$ (positive definiteness);
- $\langle \mathbf{v}, \mathbf{w}\rangle = \overline{\langle \mathbf{w}, \mathbf{v}\rangle}$ (conjugate symmetry);
- $\langle \mathbf{u},\ \alpha\mathbf{v} + \beta\mathbf{w}\rangle = \alpha \langle \mathbf{u}, \mathbf{v}\rangle + \beta \langle \mathbf{u}, \mathbf{w}\rangle$ (linearity in the second parameter).

The fundamental differences with the real case are the conjugation in the "symmetry" property and its main consequence: the scalar product is *not* linear in the first argument. namely, what would be linearity in the real case now reads: $\langle \alpha\mathbf{u} + \beta\mathbf{v},\ \mathbf{w}\rangle = \bar{\alpha} \langle \mathbf{u}, \mathbf{w}\rangle + \bar{\beta} \langle \mathbf{v}, \mathbf{w}\rangle$. Everything else, i.e. Cauchy-Schwarz's inequality and the definition of norm, remain intact.

1.5.3 Representation of the scalar product (finite-dimensional spaces)

My purpose here is to reconnect the abstract definition of scalar product with the usual one that you know in good ol' $\mathbb{R}^n$. In this case we must get our hands a bit dirty with indices.

Suppose that a finite-dimensional vector space V over $\mathbb{R}$ is equipped with a scalar product $\langle \cdot, \cdot\rangle$, and assume that we have fixed a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \subseteq V$. Splitting sums and bringing scalars out, the scalar product between two vectors can be written as

$$\begin{aligned}
\langle \mathbf{v}, \mathbf{w}\rangle &= \langle v_1\mathbf{b}_1 + \ldots + v_n\mathbf{b}_n,\quad w_1\mathbf{b}_1 + \ldots + w_n\mathbf{b}_n\rangle \\
&= \sum_{i,j=1}^{n} v_i w_j\ \ \langle \mathbf{b}_i, \mathbf{b}_j\rangle \\
&= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1\rangle & \langle \mathbf{b}_1, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_1, \mathbf{b}_n\rangle \\ \langle \mathbf{b}_2, \mathbf{b}_1\rangle & \langle \mathbf{b}_2, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_2, \mathbf{b}_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{b}_n, \mathbf{b}_1\rangle & \langle \mathbf{b}_n, \mathbf{b}_2\rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n\rangle \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}
\end{aligned} \tag{3}$$

where the coefficients v_i, w_j are uniquely determined by the basis. At this stage I confess that I am tempted to rewrite (3) as

$$\begin{aligned}
&\left\langle \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \right\rangle \\
&= \left\langle \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \right\rangle \\
&= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}
\end{aligned}$$

and pretend that the sporting club maths taking place in the inner column·row product works nicely because the products between $\mathbf{b}_i$ and $\mathbf{b}_j$ *do* make sense (as scalar products, of course); but let's say I will resist the temptation.

Anyway the square matrix in (3) has very nice properties: it is symmetric ($P = P^\top$) and it satisfies

$$\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle & \langle \mathbf{b}_1, \mathbf{b}_2 \rangle & \cdots & \langle \mathbf{b}_1, \mathbf{b}_n \rangle \\ \langle \mathbf{b}_2, \mathbf{b}_1 \rangle & \langle \mathbf{b}_2, \mathbf{b}_2 \rangle & \cdots & \langle \mathbf{b}_2, \mathbf{b}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{b}_n, \mathbf{b}_1 \rangle & \langle \mathbf{b}_n, \mathbf{b}_2 \rangle & \cdots & \langle \mathbf{b}_n, \mathbf{b}_n \rangle \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2 > 0$$

whenever $\mathbf{v} \neq 0$ or, which is the same (linear independence of the basis!), when the vector of coordinates $v \neq 0$. Such a matrix is called *positive definite*.

Definition.

A symmetric matrix $P \in \mathbb{R}^{n \times n}$ is called:

- *positive semi-definite*, if $v^\top P v \geq 0$ for all $v \in \mathbb{R}^n$; this fact is denoted $P \geq 0$;
- *positive definite*, if $v^\top P v > 0$ for all $v \in \mathbb{R}^n, v \neq 0$; this fact is denoted $P > 0$;

of course a positive definite matrix is also positive semi-definite.

Fact.

Suppose that V is a vector space over $\mathbb{R}$ with $\dim V = n$. A scalar product $\langle \cdot, \cdot \rangle$ over V and a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\} \subseteq V$ uniquely identify a matrix $P \in \mathbb{R}^{n \times n}$, $P > 0$, such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = v^\top P w$$

for all $\mathbf{v}, \mathbf{w} \in V$. Vice versa, to any choice $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $P > 0$ there corresponds a well-defined scalar product in V.

Now the following should come from your heart:

Fact.

Any matrix $P \in \mathbb{R}^{n\times n}$, $P > 0$ defines a scalar product in $\mathbb{R}^n$:

$$\langle v, w\rangle := v^\top P w$$

(note: we are implicitly assuming that $\mathbb{R}^n$ comes along with the canonical basis); vice versa, any scalar product in $\mathbb{R}^n$ is represented by a matrix $P > 0$.

The simplest positive definite matrix that one can come up with is the identity matrix ($v^\top I v = v^\top v = \sum_{i=1}^n v_i^2 > 0$ for all $v \neq 0$); and indeed the prototypical scalar product in $\mathbb{R}^n$ is

$$\langle v, w\rangle = v^\top w. \tag{4}$$

The space $\mathbb{R}^n$ (with the implicit canonical basis), equipped with the scalar product (4), the norm $\|v\|_2 = \sqrt{\langle v, v\rangle}$, and the distance $d(v, w) = \|v - w\|_2$ is called the *n-dimensional Euclidean space.*

All these concepts must be reworked a bit for *complex* vector spaces, but the substance remains the same.

Definition.

A matrix $P \in \mathbb{C}^{n\times n}$ is called *Hermitian* if $P = P^*$ ($*$ denotes transpose and conjugate). Assuming that V is a vector space over $\mathbb{C}$, the computations until (3) remain the same, but the matrix in (3) is complex and, in general, *not symmetric*; it is, however, Hermitian. A Hermitian matrix $P \in \mathbb{C}^{n\times n}$ is called

- *positive semi-definite* ($P \geq 0$), if $v^* P v \geq 0$ for all $v \in \mathbb{C}^n$;
- *positive definite* ($P > 0$), if $v^* P v > 0$ for all $v \in \mathbb{C}^n, v \neq 0$.

Fact.

- Suppose that V is a vector space over $\mathbb{C}$ and $\dim V = n$. A scalar product $\langle \cdot, \cdot \rangle$ over V and a basis $\{\mathbf{b}_1, \dots, \mathbf{b}_n\} \subseteq V$ uniquely identify a Hermitian matrix $P \in \mathbb{C}^{n \times n}$, $P > 0$, such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = v^* P w$$

  for all $\mathbf{v}, \mathbf{w} \in V$. Vice versa, to any choice $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ and $P > 0$ there corresponds a well-defined scalar product in V.

- Any Hermitian matrix $P \in \mathbb{C}^{n \times n}$, $P > 0$, defines a scalar product in $\mathbb{C}^n$ trough

$$\langle v, w \rangle := v^* P w$$

  (again, we are implicitly assuming that $\mathbb{C}^n$ comes with the canonical basis); vice versa, any scalar product in $\mathbb{C}^n$ is represented by a Hermitian and positive definite matrix P.

- $P = I$ yields the canonical scalar product in $\mathbb{C}^n$:

$$\langle v, w \rangle = v^* w.$$

1.5.4 Orthogonality

From now on, unless otherwise stated, I will restrict the discussion to *finite-dimensional* vector spaces over $\mathbb{R}$. (Vector spaces over $\mathbb{C}$ are interesting, but we don't really use them in the DDSM course; and in infinite-dimensional spaces things tend to get more involved.) So, unless otherwise stated, in the rest of the document V and W will denote finite-dimensional, real vector spaces equipped with a scalar product, a norm, and a distance.

Definition.

Two vectors $\mathbf{v}, \mathbf{w} \in V$ are called *orthogonal* if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. This is denoted $\mathbf{v} \perp \mathbf{w}$.

Fact.

(*Pythagoras's theorem.*) If $\mathbf{v} \perp \mathbf{w}$, then

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

Indeed:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2. \end{aligned}$$

Fact.

If the nonzero vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ are orthogonal to each other, they are linearly independent.

Indeed, let $\mathbf{v}_i \perp \mathbf{v}_j$ for all $i, j = 1 \ldots n, i \neq j$. Suppose that

$$\alpha_1 \mathbf{v}_1 + \ldots + \alpha_i \mathbf{v}_i + \ldots + \alpha_n \mathbf{v}_n = 0.$$

Then, taking the scalar product with $\mathbf{v}_i$,

$$\begin{aligned} 0 &= \alpha_1 \langle \mathbf{v}_1, \mathbf{v}_i \rangle + \ldots + \alpha_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle + \ldots + \alpha_n \langle \mathbf{v}_n, \mathbf{v}_i \rangle \\ &= \alpha_1 \cdot 0 + \ldots + \alpha_i \cdot \|\mathbf{v}_i\|^2 + \ldots + \alpha_n \cdot 0 = \alpha_i \|\mathbf{v}_i\|^2. \end{aligned}$$

Since $\mathbf{v}_i \neq 0$, it must be $\alpha_i = 0$. Repeating for $i = 1, \cdots, n$ we get $\alpha_1 = \cdots = \alpha_n = 0$, which meets the definition of linear independence.

Definition.

Given a subset $S \subset V$, the *orthogonal complement* of S in V is the set

$$S^\perp = \{\mathbf{v} \in V \ : \ \mathbf{v} \perp \mathbf{w} \text{ for all } \mathbf{w} \in S\}.$$

Whatever set is S, $S^\perp$ is a subspace of V; indeed if $\mathbf{v}_1, \mathbf{v}_2 \in S^\perp$, then for all $\mathbf{w} \in S$ it holds $\langle \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2, \ \mathbf{w} \rangle = \alpha_1 \langle \mathbf{v}_1, \mathbf{w} \rangle + \alpha_2 \langle \mathbf{v}_2, \mathbf{w} \rangle = 0$, and hence $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \in S^\perp$ as well.

Fact.

Let W be a subspace of V. Then[a]

$$\left(W^\perp\right)^\perp = W.$$

[a] The claim is false, in general, if V is infinite-dimensional. What remains true is that $\left(W^\perp\right)^\perp =$ the closure of W. However, any finite-dimensional space or subspace is automatically closed (topological notions like "closure", "open set", "closed set" and so on are inherited naturally from the distance).

Fact.

Let W be a subspace of V. Then every vector $\mathbf{v} \in V$ can be expressed in an unique way as

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$

where $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$. The vector $\mathbf{w}$ is called the *orthogonal projection* of $\mathbf{v}$ on the subspace W (similarly, $\mathbf{w}^\perp$ is the orthogonal projection of $\mathbf{v}$ on the subspace $W^\perp$).

1.6 Range and null space

1.6.1 General definitions

Let $\mathcal{A} : V \to W$ be a linear map.

Definition.

The *range* of $\mathcal{A}$ is the set

$$\text{range } \mathcal{A} = \{\mathbf{w} \in W \ : \ \mathbf{w} = \mathcal{A}(\mathbf{v}) \text{ for some } \mathbf{v} \in V\} .$$

Fact.

range $\mathcal{A}$ is a subspace of W.

Indeed, if $\mathbf{w}_1, \mathbf{w}_2 \in \text{range } \mathcal{A}$, then $\mathbf{w}_1 = \mathcal{A}(\mathbf{v}_1)$ and $\mathbf{w}_2 = \mathcal{A}(\mathbf{v}_2)$ for some vectors $\mathbf{v}_1, \mathbf{v}_2$. But then

$$\alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 = \alpha_1 \mathcal{A}(\mathbf{v}_1) + \alpha_2 \mathcal{A}(\mathbf{v}_2) = \mathcal{A}(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2),$$

so that also $\alpha_1 \mathbf{w}_1 + \alpha_2 \mathbf{w}_2 \in \text{range } \mathcal{A}$.

Definition.

The *null space* of $\mathcal{A}$ is the set[a]

$$\text{null } \mathcal{A} = \{\mathbf{v} \in V \ : \ \mathcal{A}(\mathbf{v}) = 0\} .$$

[a] In the literature, the null space of $\mathcal{A}$ is also called the *kernel* of $\mathcal{A}$, and denoted Ker $\mathcal{A}$.

Fact.

null $\mathcal{A}$ is a subspace of V.

Indeed, if $\mathbf{v}_1, \mathbf{v}_2 \in \text{null } \mathcal{A}$, then $\mathcal{A}(\mathbf{v}_1) = 0$ and $\mathcal{A}(\mathbf{v}_2) = 0$; but then

$$\mathcal{A}(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2) = \alpha_1 \mathcal{A}(\mathbf{v}_1) + \alpha_2 \mathcal{A}(\mathbf{v}_2) = 0,$$

so that $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \in \text{null } \mathcal{A}$ as well.

Fact.

A <u>*linear equation*</u>

$$\mathcal{A}(\mathbf{v}) = \mathbf{b} \tag{5}$$

admits a solution if and only if $\mathbf{b} \in \text{range } \mathcal{A}$. If $\bar{\mathbf{b}}$ is a "particular" solution of (5), then the set of all its solutions, that is the so-called "general" solution, is

$$\bar{\mathbf{b}} + \text{null } \mathcal{A} = \left\{\bar{\mathbf{b}} + \mathbf{k} \ : \ \mathbf{k} \in \text{null } \mathcal{A}\right\} .$$

In particular, the solution $\bar{\mathbf{b}}$ is unique if and only if null $\mathcal{A} = \{0\}$.

<u>Examples</u>:

- Let W be a subspace of V, and let $\mathcal{A} : V \to V$ defined by $\mathcal{A}(\mathbf{v})$ =the component $\mathbf{w}$ in the orthogonal decomposition $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$. It is not difficult to show that $\mathcal{A}$ is a linear map, and of course it holds range $\mathcal{A} = W$.

- For the last time, I will indulge in examples involving infinite-dimensional spaces. (Don't worry about mathematical subtleties.) Consider the *derivative operator* $\frac{d}{dt} : C^1[a,b] \to C^0[a,b]$; we have

$$\text{null}\ \frac{d}{dt} = \left\{ f \in C^1[a,b] \ : \ \frac{df(t)}{dt} \equiv 0 \right\} = \{\text{constant functions}\};$$

- consider the linear differential equation

$$\frac{d^n f(t)}{dt^n} + a_{n-1}\frac{d^{n-1}f(t)}{dt^{n-1}} + \ldots + a_1\frac{df(t)}{dt} + a_0 f(t) = g(t).$$

  It can be written as

$$\left(\frac{d^n}{dt^n} + a_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1\frac{d}{dt} + a_0\right) f = g,$$

  where $\left(\frac{d^n}{dt^n} + a_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1\frac{d}{dt} + a_0\right)$ is a linear "differential operator" defined over the space of sufficiently smooth functions. If a particular solution $\bar{f}(t)$ of the equation is known, then the general solution of the differential equation is

$$\bar{f} + \text{null}\ \left(\frac{d^n}{dt^n} + a_{n-1}\frac{d^{n-1}}{dt^{n-1}} + \ldots + a_1\frac{d}{dt} + a_0\right);$$

  stated otherwise, any solution of the differential equation has the form $\bar{f}(t) + k(t)$, where $k(t)$ is a solution of the associated "homogeneous" equation

$$\frac{d^n k(t)}{dt^n} + a_{n-1}\frac{d^{n-1}k(t)}{dt^{n-1}} + \ldots + a_1\frac{dk(t)}{dt} + a_0 k(t) = 0.$$

1.7 Range and null space of a matrix

The definitions above will become clearer if we apply them to $\mathbb{R}^n$ and $\mathbb{R}^m$, understood as "column vectors". Let therefore $A \in \mathbb{R}^{m\times n}$; without worrying too much about abuse of notation, understand A as a linear map: $A : \mathbb{R}^n \to \mathbb{R}^m$. I'll repeat the two definitions of the previous section, and add new fundamental facts:

Definition.

The *range* of A is the set

$$\text{range}\ A = \{w \in \mathbb{R}^m \ : \ w = Av \text{ for some } v \in \mathbb{R}^n\}.$$

(range A is a subspace of $\mathbb{R}^m$.)

Explanation. Visualize A as a row of columns:

$$Av = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Then

$$\begin{aligned}
\text{range } A &= \{\text{all vectors } w \in \mathbb{R}^m \text{ that can be obtained as } w = Av\} \\
&= \left\{ \text{all vectors } w \in \mathbb{R}^m \text{ that can be obtained as } \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \right\} \\
&= \{\text{all vectors } w \in \mathbb{R}^m \text{ that are linear combinations of the columns of } A\} \\
&= \text{span } \{\text{columns of } A\}\,.
\end{aligned}$$

Definition.

The *null space* of A is the set

$$\text{null } A = \{v \in \mathbb{R}^n \;:\; Av = 0\}\,.$$

(null A is a subspace of $\mathbb{R}^n$.)

Explanation. Visualize A as a column of rows:

$$A = \begin{bmatrix} — & r_1 & — \\ & \vdots & \\ — & r_m & — \end{bmatrix}, \qquad Av = \begin{bmatrix} — & r_1 & — \\ & \vdots & \\ — & r_m & — \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix} = \begin{bmatrix} r_1 v \\ r_2 v \\ \vdots \\ r_n v \end{bmatrix},$$

and note that the rows of A happen to be the transposes of the columns of $A^\top$:

$$\begin{aligned}
A^\top &= \begin{bmatrix} | & & | \\ r_1^\top & \cdots & r_m^\top \\ | & & | \end{bmatrix} := \begin{bmatrix} | & & | \\ \bar{c}_1 & \cdots & \bar{c}_m \\ | & & | \end{bmatrix}, \\
Av &= \begin{bmatrix} — & \bar{c}_1^\top & — \\ & \vdots & \\ — & \bar{c}_m^\top & — \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix} = \begin{bmatrix} \bar{c}_1^\top v \\ \vdots \\ \bar{c}_m^\top v \end{bmatrix} = \begin{bmatrix} \langle \bar{c}_1, v \rangle \\ \vdots \\ \langle \bar{c}_m, v \rangle \end{bmatrix}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{null } A &= \{\text{all vectors } v \in \mathbb{R}^n \text{ such that } Av = 0\} \\
&= \left\{ \text{all vectors } v \in \mathbb{R}^n \text{ such that } \begin{bmatrix} \langle \bar{c}_1, v \rangle \\ \vdots \\ \langle \bar{c}_m, v \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \right\} \\
&= \left\{ \text{all vectors } v \in \mathbb{R}^n \text{ that are orthogonal to the columns of } A^\top \right\} \\
&= \left\{ \text{all vectors } v \in \mathbb{R}^n \text{ that are orthogonal to } \textit{linear combinations} \text{ of columns of } A^\top \right\} \\
&= \left(\text{span } \left\{ \text{columns of } A^\top \right\} \right)^\perp \\
&= \left(\text{range } A^\top \right)^\perp.
\end{aligned}$$

We have proven the first claim of the following

Fact.

Let $A \in \mathbb{R}^{m\times n}$, understood as a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then

- $\text{null } A = \left(\text{range } A^\top\right)^\perp$;
- $\text{range } A^\top = (\text{null } A)^\perp$ (to prove this one, take another orthogonal complement);
- $\text{null } A^\top = (\text{range } A)^\perp$ (to prove this one, substitute $A^\top$ in place of A);
- $\text{range } A = \left(\text{null } A^\top\right)^\perp$ (same story).

And here is a corollary:

Fact.

For any matrix $A \in \mathbb{R}^{m\times n}$,

$$\text{range } A = \text{range } AA^\top.$$

Proof. Suppose that $v \in \text{null } A^\top$. This means $A^\top v = 0$, hence also $AA^\top v = 0$ and $v \in \text{null } AA^\top$. Suppose, on the other hand, that $v \in \text{null } AA^\top$. Then $AA^\top v = 0$, hence also $\|A^\top v\|_2^2 = (A^\top v)^\top A^\top v = v^\top AA^\top v = 0$. This implies that $A^\top v = 0$ and $v \in \text{null } A^\top$. Hence $\text{null } A^\top = \text{null } AA^\top$, and by the previous "fact",

$$\text{range } A = (\text{null } A^\top)^\perp = (\text{null } AA^\top)^\perp = \text{range } AA^\top.$$

Note: we could provide equivalent definitions and similar "facts" for *arbitrary* vector spaces (complex, infinite-dimensional). The issue would be to find a decent counterpart of $A^\top$. What should we take, instead of "the transpose", when dealing with a linear map $\mathcal{A}$?[6] If you are interested in this kind of mathematics, I invite you to refer to the *excellent* textbook: David G. Luenberger, *Optimization by vector space methods*.

1.8 Rank of a matrix

Let $A \in \mathbb{R}^{m\times n}$. Probably you remember the following definition:

Definition. The *rank* of A is the maximum dimension of a square matrix, obtained from A by suppressing some rows and/or columns, with nonzero determinant.

Letting aside the fact that here we don't yet care about determinants, the above definition seems pointless to me; it conveys no intuition at all. The following characterizations, instead, are way more intuitive and useful:

[6] We should take the so-called *adjoint* of $\mathcal{A}$.

Fact.

The rank of A is equal to the dimension of the subspace of $\mathbb{R}^m$ generated by its columns:

$$\begin{aligned} \operatorname{rank} A &= \dim \operatorname{span} \{\text{columns of } A\} \\ &= \dim \operatorname{range} A. \end{aligned}$$

Fact.

(Quite remarkable, I would say.) The rank of A is equal to the rank of $A^\top$. In a sense, then,

$$\begin{aligned} \operatorname{rank} A &= \dim \operatorname{range} A^\top \\ &= \dim \operatorname{span} \{\text{columns of } A^\top\} \\ &= \dim \operatorname{span} \{\text{rows of } A\}. \end{aligned}$$

If $A \in \mathbb{R}^{m \times n}$, where $m \geq n$ ("tall" matrix, i.e. more rows than columns), we say that A has *full rank* if $\operatorname{rank} A = n =$ number of columns. In this case, A has full rank if and only if the columns of A are linearly independent, and the subspace of $\mathbb{R}^m$ generated by them has dimension n (the maximum possible).

Conversely, if $m \leq n$ ("flat" matrix, i.e. more columns than rows), we say that A has *full rank* if $\operatorname{rank} A = m =$ number of rows. Then A has full rank if and only if the *rows* of A are linearly independent, and their span has dimension m.