NOTES FOR MATHEMATICS FOR DATA SCIENCE/BIOSTATISTICS

TRENTO, A.A. 2019/20

INSTRUCTORS: ANDREA CARANTI, SIMONE UGOLINI

Nota.

- Please let me know of any errors you may find!
- The original Italian version is being translated into English as the lectures proceed.

1. Linear equations

1.1. **One linear equation in one unknown.** The equation $2x = 3$ in the unknown x has the unique solution $x = \frac{3}{2}$.

Consider the equation $ax = b$ in the unknown x, where a, b are numbers. If $a \neq 0$, it has the unique solution $x = a^{-1}b$. If $a = 0$, there are two possibilities. Either $b \neq 0$, and then$0 = 0x = ax = b$ has no solution; or $b = 0$, and then any value of x yields a solution.

Here a, b were real numbers, say, and we were looking for the real numbers that made the equation an identity.

1.2. **One linear equation in two unknowns.** Consider now the equation

$$ax + by = c, \tag{1.1}$$

in the *two* unknowns x, y. As above, if $a = b = 0$ there are two possibilities:

(1) if $c \neq 0$ there are no solutions, while
(2) if $c = 0$ the equation becomes $0 = 0$, so every value of x, y yields a solution.

If a, b are not both zero, say $a \neq 0$, then we have

$$x = -a^{-1}by + a^{-1}c. \tag{1.2}$$

So it is enough to assign an *arbitrary* value to y to obtain infinite solutions.

As an example, the equation

$$x + y = 10$$

has all the solutions $x = 0, y = 10$, $x = 1, y = 9$, $x = 2, y = 8$ and more. In general, one can assign an *arbitrary* value to y and obtain a value of x which yields a solution.

Note that the harmless looking equation $ax = b$ can now hide a trick, because you need to know whether it is in the single unknown x, or maybe in the unknown $x, y, z, \dots$, who knows?

Trento, A.A. 2019/20.

1.3. **Two linear equations in two unknowns.** Let us now consider two equations in two unknowns

$$\begin{cases} 2x + 3y = 4 \\ 5x + 7y = 6. \end{cases} \tag{1.3}$$

Here we are lokking for all values of x, y, if any, which satisfy both equations at the same time.

One method consists in exploiting the case of a single equation. Thus the solutions of the first equation of (1.3) can be obtained by giving an arbitray value to y, and then obtaining x as

$$x = -\frac{3}{2}y + 2. \tag{1.4}$$

Substituting this value of x in the second equation of (1.3), we obtain an equation in one unknown

$$5(-\frac{3}{2}y + 2) + 7y + 10 = -\frac{1}{2}y + 10 = 6, \tag{1.5}$$

which yields $y = 8$. Substituting this value of y in (1.4) we obtain $x = -10$, so that $x = 10, y = 8$ is the unique solution of the system.

For later usage, another methods appears handier. Note that (1.3) can be rewritten as

$$\begin{cases} 2x + 3y - 4 = 0 \\ 5x + 7y - 6 = 0, \end{cases} \tag{1.6}$$

and then if x, y is a solution of (1.3), it follows it is also a solution of

$$0 = (5x + 7y - 6) - \frac{5}{2}(2x + 3y - 4) = -\frac{1}{2}y + 4.$$

Here the coefficient $-\frac{5}{2}$ has been chosen so that x vanishes. (As it often happens in Mathematics, this will become clearer when we deal with the *general* case.)

Thus if x, y are a solution of (1.3), they are also a solution of

$$\begin{array}{cccc} 2x & +3y & = & 4 \\ & \frac{1}{2}y & = & 4. \end{array} \tag{1.7}$$

Now the second equation of (1.7) immediately yields $y = 8$. Substituting this value in the first one, we obtain $2x + 24 = 4$, so that $x = -10$.

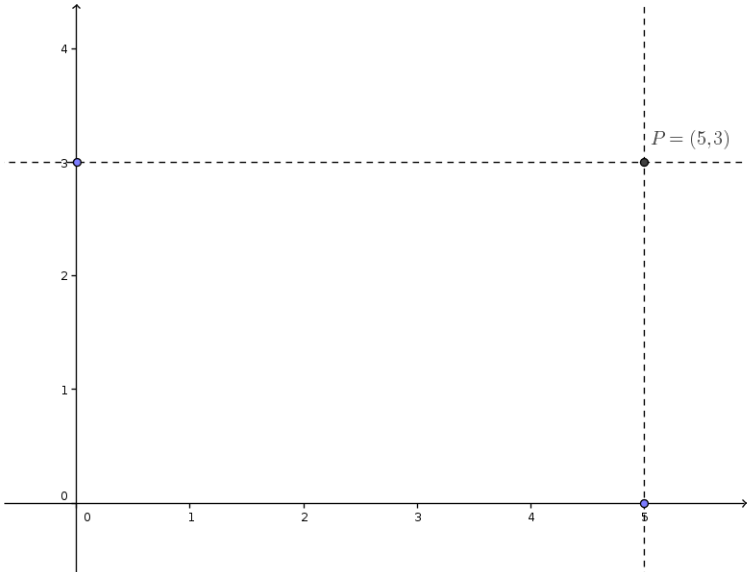
At this stage, the setting screams for a *geometrical intepretation*.

2. Lines and vectors in the plane

2.1. **The real line and the Cartesian plane.** Recall that the real numbers represent the points of a line, the *real line*.

The points of the plane can be represented through *ordered* pairs of real numbers, that is, elements of the Cartesian product of two copies of the set $\mathbf{R}$ of real numbers, $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R} = \{ (a, b) : a, b \in \mathbf{R} \}$. The basic property of ordered pairs states that $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. One can then say that a is the first coordinate of (a, b), and b the second one.

FIGURE 2.1. The point P has coordinates $(5, 3)$



For instance, in Figure 2.1 the point P has coordinates $(5, 3)$, because it lies on the vertical line intersecting the x axis in the point 5, and on the horizontal line intersecting the y axis in the point 3.

2.2. **Vettori.** If A, B are two points in $\mathbf{R}^2$, the *applied vector (in A)* $\overrightarrow{AB}$ is the line segment AB, in the direction from A to B. This is drawn as a line segment, with an arrow pointing from A to B.

If $O = (0, 0)$ is the origin (denoted as you see by a capital *letter* "O", while its coordinates are two zeroes), a vector $\overrightarrow{Ov}$, where $v \in \mathbf{R}^2$ applied in O is referred to simply as a vector, and denoted by its other end v, that is, a point in the plane.

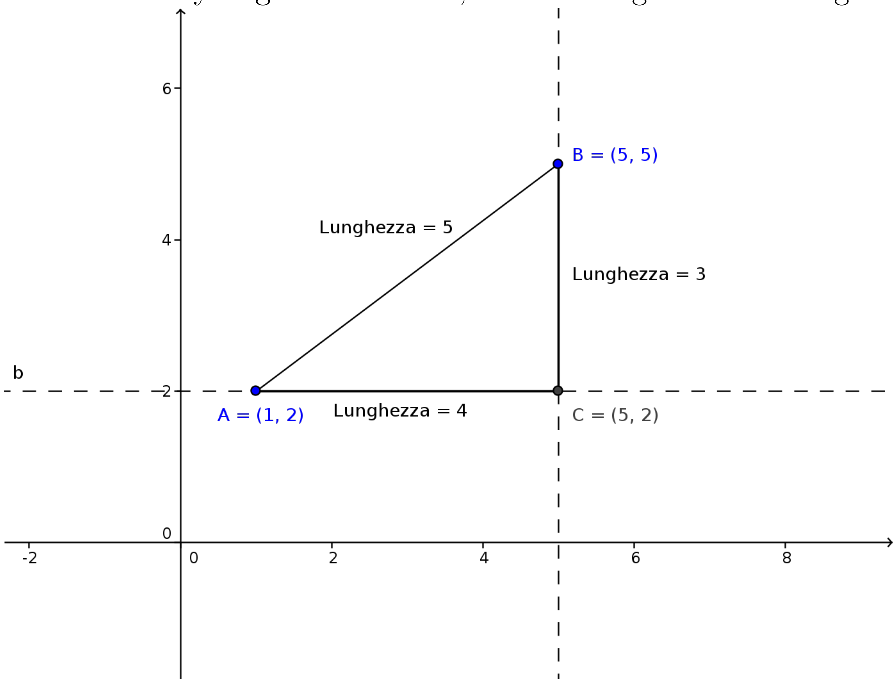
Let $A = (x_A, y_A)$, and $B = (x_B, y_B)$ be two points in $\mathbf{R}^2$, then to the applied vector $\overrightarrow{AB}$ one associates the vector (according to the definition we have just given) $v = (x_B - x_A, y_B - y_A)$. We claim

2.2.1. **Proposition.** *The four-sided polygon $OABv$ is a parallelogram.*

2.3. **Length of a line segment.** To prepare the proof of Proposition 2.2.1, let us see how one computes the length of a line segment AB. Let us refer to Figure 2.2. One the points $A = (x_A, y_A)$ and $B = (x_B, y_B)$ are given, consider the point $C = (x_B, y_A)$; we see that the triangle ABC is a right triangle in C. Now the side AC has length $x_B - x_A$ (in general, $|x_B - x_A|$, see the definition of the absolute value in 2.3), and the side BC has length $y_B - y_A$ (in general, $|y_B - y_A|$), so that the hypotenuse AB has length

$$\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}. \tag{2.1}$$

FIGURE 2.2. Pythagoras Theorem, and the length of a line segment



The length of the line segment AB is denoted by $\|AB\|$. In calculations, we will prefer to compute with $\|AB\|^2 = (x_B - x_A)^2 + (y_B - y_A)^2$, to get rid of the square root.

Proof of Proposition 2.2.1. Let us refer to Figure 2.3. Abbiamo $A = (x_A, y_A)$, $B = (x_B, y_B)$, e $v = (x_B - x_A, y_B - y_A)$, dunque

$$\|AB\|^2 = (x_B - x_A)^2 + (y_B - y_A)^2 = \|Ov\|^2, \qquad \|OA\|^2 = x_A^2 + y_A^2 = \|vB\|^2.$$

The opposite sides of the quadrilateral $OABv$ have the same length, so that $OABv$ is a parallelogram. In particular opposite sides are parallel. $\square$

From Proposition 2.2.1 it follows that the applied vector $\overrightarrow{AB}$ and the associated vector v have the same length, and are parallel. We will say that two appied vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are *equivalent* if they have the same associated vector. Thus two applied vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are equivalent if they have the same length, are parallel, and have the same direction.

2.4. **Operazioni su vettori.** Ora introduciamo due operazioni su vettori. La prima è la moltiplicazione di un vettore $v \in \mathbf{R}^2$ per uno *scalare*, cioè un $t \in \mathbf{R}$. Se $v = (x, y)$, questa è definita mediante

$$tv = t(x, y) = (tx, ty). \tag{2.2}$$

Come si vede nella Figura 2.3, illustrata per $t > 0$, i triangoli O, v, C e O, tv, D hanno il lati proporzionali, dunque sono simili, e gli angoli $\widehat{v, O, C}$ e $\widehat{tv, O, D}$ sono eguali a uno stesso angolo α. Dunque la situazione è quella della Figura 2.5, cioè

FIGURE 2.3. An applied vector AB and the associated vector v

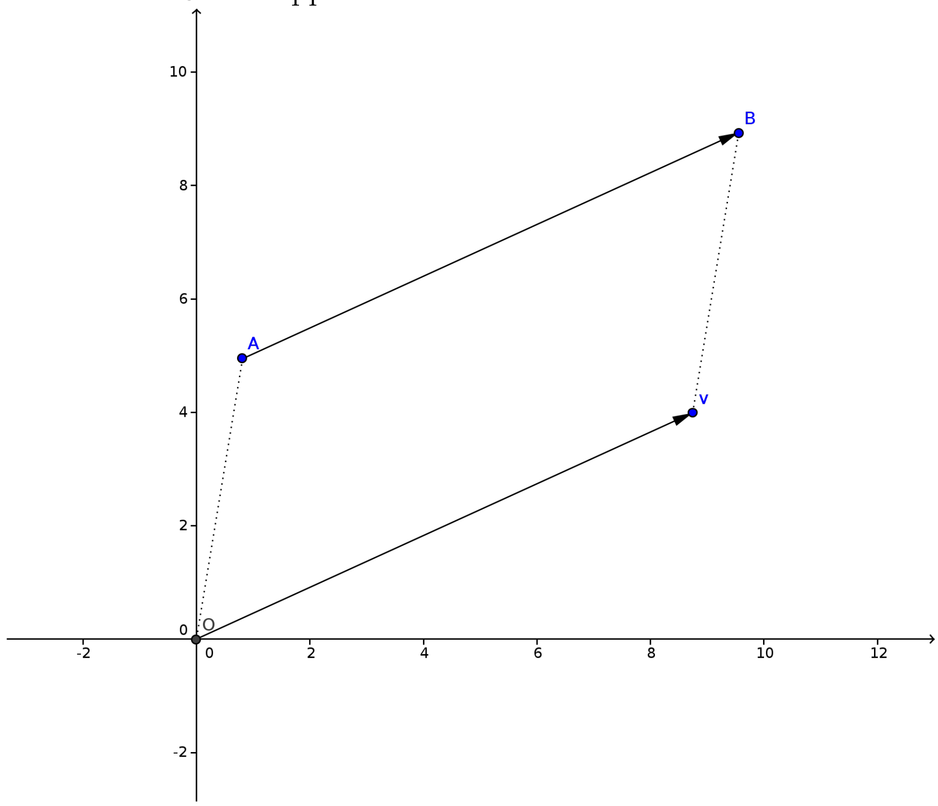
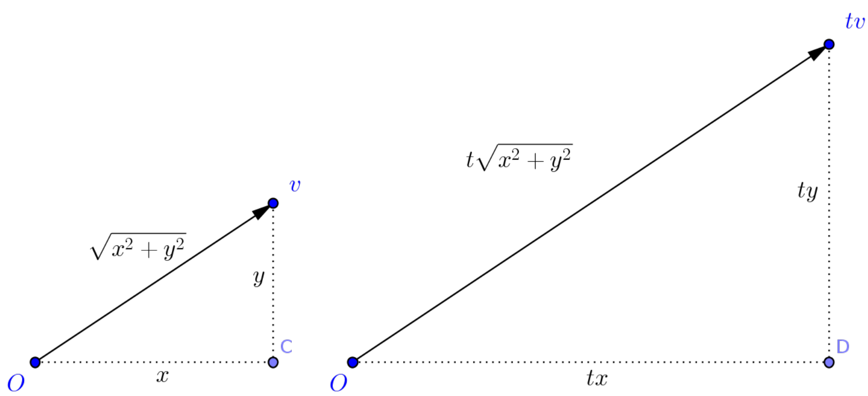


FIGURE 2.4. Product of a vector by a scalar



il vettore tv è parallelo a v, ha la stessa direzione, e ha lunghezza t volte quella di v.

Se $t = 0$ si ha ovviamente $tv = (0, 0) = O$. Se $t < 0$, allora tv è sempre parallelo a v, ha direzione opposta, e

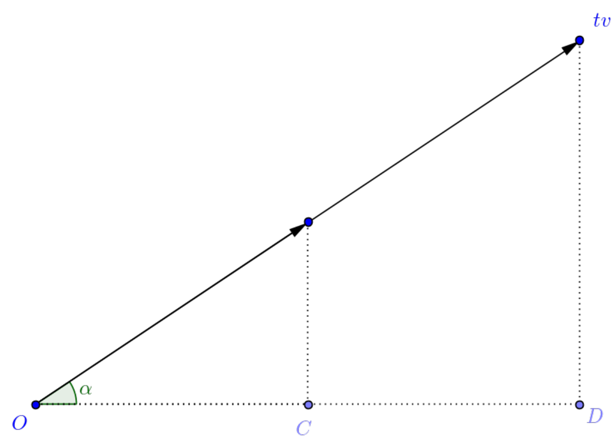
$$\|tv\| = |t| \cdot \|v\| \, ,$$

formula questa che vale per ogni valore di $t \in \mathbf{R}$, positivo, negativo o nullo. Qui $|t|$ indica il *valore assoluto* di t, definito da

$$|t| = \begin{cases} t & \text{se } t \geq 0, \\ -t & \text{se } t < 0, \end{cases} \tag{2.3}$$

e abbiamo usato il fatto che $\sqrt{t^2 a} = |t| \sqrt{a}$. Un caso particolare importante è notare che se $v = (x, y)$, allora $-v = (-1)v = (-x, -y)$.

FIGURE 2.5. Prodotto di un vettore per uno scalare



Definiamo poi la somma di due vettori $v = (v_1, v_2)$ e $w = (w_1, w_2)$ mediante

$$v + w = (v_1 + w_1, v_2 + w_2). \tag{2.4}$$

Notiamo subito che valgono le seguenti proprietà.

(1) La somma è associativa, cioè se u, v, w sono tre vettori, allora

$$(u + v) + w = u + (v + w).$$

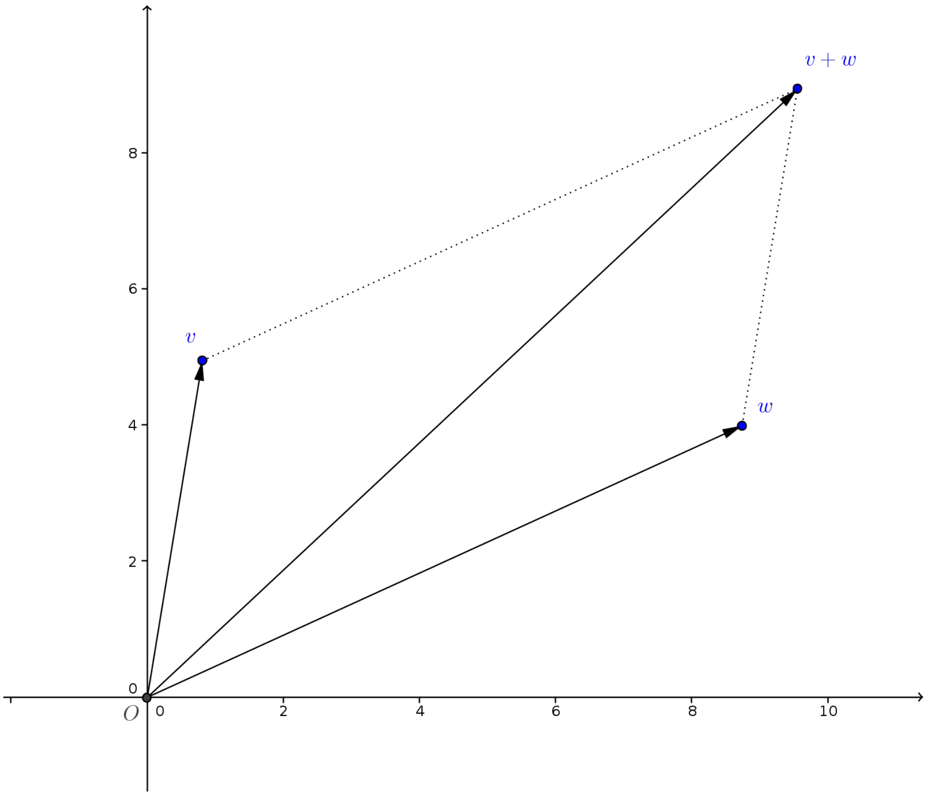
(2) La somma è commutativa, cioè se v, w sono due vettori, allora

$$v + w = w + v.$$

(3) Per ogni vettore v si ha $v + O = v$.

(4) per ogni vettore v si ha $v + (-v) = O$.

Figure 2.6. Somma di due vettori



Inoltre la somma ha un'importante interpretazione geometrica. Consideriamo la Figura 2.6. Con gli stessi argomenti della Proposizione 2.2.1 si vede che $O, v, v+w, w$ ha i lati opposti della stessa lunghezza, e dunque è un parallelogramma. Vale dunque

2.4.1. **Proposizione.**

(1) *Dati due vettori v, w, la loro somma $v+w$ è la diagonale di un parallelogramma di cui due lati sono v, w.*

(2) *Il vettore applicato*

$$\overrightarrow{v, v+w}$$

si ottiene dal vettore w traslandolo parallalemente mediante il vettore v.

2.5. **Equazione implicita (cartesiana) e parametrica della retta nel piano.** Torniamo adesso alle soluzioni dell'equazione (1.1). Abbiamo indicato negli Esercizi che se a, b non sono entrambi nulli, allora l'equazione si può risolvere nel modo seguente. Sia (x_0, y_0) una soluzione fissata. (Per trovarne una, se per esempio $a \neq 0$, basta dare un valore arbitrario y_0 a y, e ricavare il valore x_0 per x.) Se (x, y) è ora una soluzione qualsiasi, si ha

$$0 = (ax + by - c) - (ax_0 + by_0 - c) = a(x - x_0) + b(y - y_0). \tag{2.5}$$

Se $a \neq 0$, poniamo $t = \dfrac{y - y_0}{a}$, e sostituendo in (2.5) si ottiene $x - x_0 = -bt$. Dunque

$$\begin{cases} x = -bt + x_0 \\ y = at + y_0, \end{cases} \tag{2.6}$$

ove t è un parametro arbitrario. Si vede che si arriva alla stessa formula se $b \neq 0$, ponendo $t = -\dfrac{x - x_0}{b}$.

Ora consideriamo le equazioni (2.6) dapprima nel caso $(x_0, y_0) = (0, 0)$, in cui diventano

$$\begin{cases} x = -bt \\ y = at, \end{cases} \tag{2.7}$$

ove $t \in \mathbf{R}$. Possiamo scriverle come

$$(x, y) = t(-b, a),$$

dunque i vettori (x, y) sono tutti i multipli scalari del vettore $(-b, a) \neq (0, 0)$, quindi descrivono una *retta passante per l'origine* $O = (0, 0)$.

Adesso vediamo che (2.6) si scrive

$$(x, y) = t(-b, a) + (x_0, y_0),$$

dunque per quanto visto sulla somma di due vettori abbiamo traslato parallelamente la retta precedente per il vettore (x_0, y_0), ovvero abbiamo una retta per (x_0, y_0), di direzione $(-b, a)$.

(2.6) si dice *equazione parametrica* di una retta nel piano, nel senso che al variare del parametro t in $\mathbf{R}$, il punto (x, y) descrive tutti e soli i punti di una retta nel piano. Invece (1.1) si dice *equazione implicita* (o anche *cartesiana*) di una retta nel piano, perché sono le sue soluzioni che formano una retta. (Stiamo sempre assumendo che a e b non siano entrambi nulli.)

2.6. **Retta nel piano per due punti.** Da questa discussione segue che si può trovare l'equazione della retta per p e q molto semplicemente in forma parametrica. Infatti diciamo che cerco la retta per i due punti $p = (p_1, p_2)$ e $q = (q_1, q_2)$, con $p \neq q$. La retta sarà parallela al vettore applicato $\overrightarrow{pq}$, dunque al vettore ad esso associato $(q_1 - p_1, q_2 - p_2)$. Dunque la parallela che passa per l'origine è

$$\begin{cases} x = t(q_1 - p_1) \\ y = t(q_2 - p_2), \end{cases}$$

e tralsandola di p ottengo la retta cercata

$$\begin{cases} x = p_1 + t(q_1 - p_1) \\ y = p_2 + t(q_2 - p_2), \end{cases} \tag{2.8}$$

infatti per $t = 0$ ho $(x, y) = p$, e per $t = 1$ ho $(x, y) = q$. (Qui si vede che se $p = q$, non ottengo una retta, ma solo un punto.)

Per trovare la forma implicita, posso eliminare t in (2.8), usare (2.6), o risolvere il sistema

$$\begin{cases} ap_1 + bp_2 - c = 0 \\ aq_1 + bq_2 - c = 0 \end{cases}$$

nelle incognite a, b, c. Badate che si troveranno infinite soluzioni, perché, fissati a, b, c, con a e b non entrambi nulli, tutte le equazioni $s(ax+by-c) = sax+sby-sc = 0$, per ogni valore di $s \neq 0$ hanno le stesse soluzioni. (Che succede se $p = q$?)

2.7. **Intersezione di due rette.** Adesso si comprende la discussione che abbiamo fatto di una sistema come quello di (1.3). Le due equazioni rappresentano due rette, e dunque sto cercando i punti di intersezione di due rette nel piano. I casi sono tre

(1) Le due rette sono la stessa retta, e allora le soluzioni formano questa retta.
(2) Le due rette sono distinte, ma parallele, e allora non ci sono punti di intersezione.
(3) Se le due rette non sono parallele, allora si intersecano in un punto.

(Ricordarsi di riprendere il discorso della retta per due punti, sia spiegando come arrivo alla forma parametrica, che vedendo che succede cercandola direttamente in forma implicita.)

2.8. **Angoli e prodotto scalare.** Consideriamo due vettori non nulli $v = (v_1, v_2)$ e $w = (w_1, w_2)$ nel piano.

Abbiamo

$$\|v - w\|^2 = \|(v_1 - w_1, v_2 - w_2)\|^2 = \|v\|^2 + \|w\|^2 - 2(v_1 w_1 + v_2 w_2). \tag{2.9}$$

Il calcolo si può fare anche in un altro modo, illustrato nella Figura 2.7, da cui si vede, guardando il triangolo rettangolo wCv che

$$\begin{aligned} \|v - w\|^2 = \|w\|^2 \sin(\alpha)^2 + ((\|v\| - \|w\| \cos(\alpha))^2 = \\ = \|v\|^2 + \|w\|^2 - 2\,\|v\|\,\|w\| \cos(\alpha), \end{aligned} \tag{2.10}$$

da cui si deduce

$$\cos(\alpha) = \frac{v \cdot w}{\|v\|\,\|w\|}, \tag{2.11}$$

ove con $v \cdot w$ si indica il *prodotto scalare* fra i vettori v e w

$$v \cdot w = v_1 w_1 + v_2 w_2. \tag{2.12}$$

In particolare, v e w sono ortogonali se e solo se $v \cdot w = 0$. La formula

$$ax + by = 0 \tag{2.13}$$

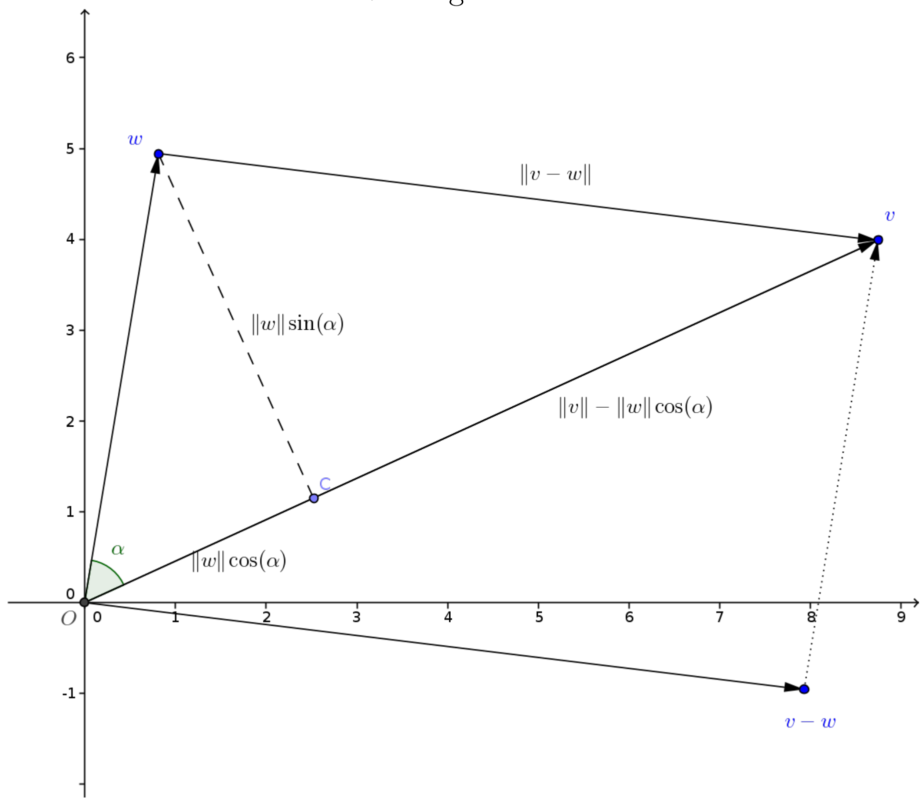
per una retta passante per l'origine si può riscrivere come

$$(x, y) \cdot (a, b) = 0, \tag{2.14}$$

dunque l'equazione (2.13) rappresenta tutti i vettori (x, y) ortogonali al vettore (a, b).

Notiamo

FIGURE 2.7. Angolo fra due vettori



2.8.1. **Proposizione** (Alcune proprietà del prodotto scalare)**.** *Se u, v, w sono vettori, e t è uno scalare, si hanno le relazioni seguenti.*

(1) $u \cdot v = v \cdot u$.
(2) $v \cdot v = \|v\|^2$.
(3) $(u + v) \cdot w = u \cdot w + v \cdot w$.
(4) $(tu) \cdot v = t(u \cdot v)$.

Le proprietà (3) e (4) esprimono il fatto che il prodotto scalare è *bilineare*, in un senso che chiariremo.

2.9. **Piani nello spazio.** Quanto fatto finora si può estendere allo spazio, cioè a

$$\mathbf{R}^3 = \{ (x, y, z) : x, y, z \in \mathbf{R} \} . \tag{2.15}$$

Se a, b, c non sono tutti nulli, allora le soluzioni dell'equazione

$$ax + by + cz = d \tag{2.16}$$

formano un *piano*.

In effetti, in analogia a quanto fatto con la retta, fissiamo una soluzione particolare (x_0, y_0, z_0) di (2.16). Abbiamo allora

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

se per esempio $a \neq 0$ pongo

$$s = \frac{y - y_0}{a}, \qquad t = \frac{z - z_0}{a},$$

e ricavo

$$\begin{cases} x = x_0 - bs - ct \\ y = y_0 + as \\ z = z_0 \qquad\quad + at. \end{cases} \tag{2.17}$$

Gli altri casi danno origine a equazioni parametriche della forma

$$\begin{cases} x = x_0 + u_1 s + v_1 t \\ y = y_0 + u_2 s + v_2 t \\ z = z_0 + u_3 s + v_3 t. \end{cases} \tag{2.18}$$

Come per il caso della retta nel piano, consideriamo il caso dell'equazione

$$ax + by + cz = 0, \tag{2.19}$$

fra le cui soluzioni c'è dunque l'origine O. Allora le equazioni parametriche (2.18) si possono mettere nella forma

$$\begin{cases} x = u_1 s + v_1 t \\ y = u_2 s + v_2 t \\ z = u_3 s + v_3 t. \end{cases} \tag{2.20}$$

Ora notiamo una cosa importante, che sarà ampiamente generalizzata nel seguito. Se uno dei due vettori $u = (u_1, u_2, u_3)$ e $v = (v_1, v_2, v_3)$ è multiplo dell'altro, ad esempio $u = kv$, allora posso riscrivere (2.20) come

$$(x, y, z) = \lambda v, \tag{2.21}$$

ove ora $\lambda = sk + t \in \mathbf{R}$ è un parametro. Ora esattamente come nel piano questa è una retta per l'origine, parallela al vettore v (ricordare l'interpretazione del prodotto di un vettore per uno scalare). Richiediamo allora che in (2.20) nessuno dei due vettori u, v sia un multiplo dell'altro. Adesso (2.20) rappresenta un piano, perché la somma $su + tv$ (per l'interpretazione della somma di due vettori) è un vettore complanare a u e v. (Notate che in (2.17) si ha $u = (ab, a, 0)$, $v = (ac, 0, a)$, con $a \neq 0$, e dunque u e v non sono uno multiplo dell'altro.)

Un altro modo di vedere che le soluzioni di (2.19) formano un piano per l'origine è notare che esse formano l'insieme dei vettori (x, y, z) ortogonali al vettore non nullo (a, b, c). In effetti la formula (2.11) vale anche per vettori nello spazio, ove se $v = (v_1, v_2, v_3)$ e $w = (w_1, w_2, w_3)$, si ha $\|v\|^2 = v_1^2 + v_2^2 + v_3^2$, $\|w\|^2 = w_1^2 + w_2^2 + w_3^2$. e $v \cdot w = v_1 w_1 + v_2 w_2 + v_3 w_3$.

Più in generale, le equazioni (2.16) e (2.18) descrivono un piano (passante per un punto (x_0, y_0, z_0)).

Anche la discussione di un sistema ad esempio di due o tre equazioni lineari in tre incognite si può fare sia algebricamente, sia considerando le possibili posizioni relative di due o tre piani dello spazio.

2.10. **Retta per un punto perpendicolare a un'altra retta nel piano.** Trovare la retta S per un punto $p = (p_1, p_2)$, perpendicolare a una retta data R.

Una volta messa la retta R in forma implicita $ax + bx = c$, basta fare la retta per p parallela al vettore (a, b), dunque la retta S ha equazioni parametriche

$$\begin{cases} x = p_0 + at \\ y = p_1 + bt. \end{cases}$$

Se la retta R è in forma parametrica, la si mette prima in forma implicita.

2.11. **Retta nello spazio per due punti.** L'equazione parametrica della retta passante per i punti *distinti* p, q è del tutto analoga a quella della retta nel piano,

$$\begin{cases} x = p_1 + (q_1 - p_1)t \\ y = p_2 + (q_2 - p_2)t \\ z = p_3 + (q_3 - p_3)t \end{cases}$$

ovvero

$$\begin{cases} x = p_1 + at \\ y = p_2 + bt \\ z = p_3 + ct \end{cases} \tag{2.22}$$

è la retta per il punto p, parallela al vettore $(a, b, c) \neq O$. Per metterla in forma cartesiana, dovrò scriverla come intersezione di due piani. L'idea è di eliminare il parametro t fra due equazioni di (2.22). Diciamo che sia $c \neq 0$, allora da (2.22) ottengo le due equazioni

$$\begin{array}{ccccccc} cx & - & & & az & = & cp_1 - ap_3 \\ & & cy & - & bz & = & cp_2 - bp_3 \end{array}$$

che sono le equazioni dei due piani cercati. Ad esempio, per la prima delle due equazioni si ha

$$cx - az = c(p_1 + at) - a(p_3 + ct) = cp_1 + act - ap_3 - act = cp_1 - ap_3.$$

2.12. **Retta nello spazio passante per un punto e perpendicolare a un piano.** Se il piano è dato da $ax + by + cz = d$, e il punto è p, è semplicemente la retta data da (2.22).

2.13. **Distanza di un punto da una retta nel piano.** Diciamo che ho il punto $p = (p_1, p_2)$, e la retta $\mathfrak{r}$ data in forma implicita da $ax + by = c$, con a, b non entrambi nulli. Dunque

$$\begin{cases} x = p_1 + at \\ y = p_2 + bt \end{cases}$$

è una retta passante per p e perpendicolare a $\mathfrak{r}$. Faccio l'intersezione con $\mathfrak{r}$ e trovo il punto q su $\mathfrak{r}$ a minima distanza da p, dopodiché calcolo la distanza fra p e q.

2.14. **Distanza di un punto da una piano nello spazio.** Diciamo che ho il punto $p = (p_1, p_2, p_3)$, e il piano $\mathfrak{r}$ data in forma implicita da $ax + by + cz = d$, con a, b, c non tutti nulli. Dunque

$$\begin{cases} x = p_1 + at \\ y = p_2 + bt \\ z = p_3 + ct \end{cases}$$

è una retta passante per p e perpendicolare a $\mathfrak{r}$. Faccio l'intersezione con $\mathfrak{r}$ e trovo il punto q su $\mathfrak{r}$ a minima distanza da p, dopodiché calcolo la distanza fra p e q.

2.15. **Distanza di un punto da una retta nello spazio.** Diciamo che ho il punto $p = (p_1, p_2, p_3)$, e la retta $\mathfrak{r}$ data in forma parametrica da

$$\begin{cases} x = p_1 + at \\ y = p_2 + bt \\ z = p_3 + ct \end{cases}$$

Il piano $\mathfrak{p}$ di equazione

$$a(x - p_1) + b(y - p_2) + c(z - p_3) = 0$$

passa per p, ed è ortogonale a $\mathfrak{r}$. Faccio l'intersezione di $\mathfrak{r}$ con $\mathfrak{p}$, e trovo il punto q su $\mathfrak{r}$ a minima distanza da p.

3. Spazi vettoriali

3.1. **Oltre lo spazio.** Tutto quanto visto si può generalizzare allo spazio

$$\mathbf{R}^n = \{ a = (a_1, a_2, \dots, a_n) : a_i \in \mathbf{R} \}.$$

Delle operazioni parliamo fra un attimo, ma anche qui abbiamo un prodotto scalare

$$a \cdot b = \sum_{i=1}^{n} a_i b_i,$$

e una lunghezza

$$\|a\| = \sqrt{a \cdot a} = \sqrt{\sum_{i=1}^{n} a_i^2},$$

e vale la formula (2.11).

3.2. **Gruppi.** Un *gruppo commutativo* è un insieme V, dotato di un'operazione $+$, che soddisfa le proprietà seguenti.

(1) "+" è una operazione associativa, cioè per ogni $u, v, w \in V$ si ha $(u + v) + w = u + (v + w)$.
(2) $+$ è una operazione commutativa, cioè per ogni $u, v \in V$ si ha $u + v = v + u$.
(3) Esiste un elemento neutro $O \in V$, cioè con la proprietà che per ogni $v \in V$ si ha $v + O = O + v = v$.
(4) Ogni elemento ha un opposto, cioè per ogni $v \in V$ esiste $-v$ in V tale che $v + (-v) = (-v) + v = O$.

Più avanti vedremo anche gruppi *non commutativi*, cioè in cui non vale il secondo assioma.

Esempi di gruppi rispetto al "+" sono gli interi $\mathbf{Z}$, i razionali $\mathbf{Q}$, i reali $\mathbf{R}$. L'esempio per noi più importante è il piano cartesiano $\mathbf{R}^2$, con la somma di vettori sopra definita. Più in generale ci interessa lo spazio

$$\mathbf{R}^n = \{ (a_1, a_2, \dots, a_n) : a_i \in \mathbf{R} \}$$

delle n-ple ordinate di elementi di $\mathbf{R}$, con la somma

$$(a_1, a_2, \dots, a_n) + (b_1, b_2, \dots, b_n) = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n),$$

con elemento neutro

$$O = (0, 0, \dots, 0),$$

e opposto

$$-(a_1, a_2, \dots, a_n) = (-a_1, -a_2, \dots, -a_n).$$

Notate che per le n-ple ordinate vale quanto valeva per le coppie ordinate, cioè

$$(a_1, a_2, \dots, a_n) = (b_1, b_2, \dots, b_n)$$

se e solo se $a_i = b_i$ per ogni i.

Ma la struttura che più ci interessa su $\mathbf{R}^n$ è quella che segue.

3.3. **Spazi vettoriali.** Uno spazio vettoriale (reale) è un insieme V, dotato di due operazioni

- una operazione "+" rispetto alla quale è un gruppo commutativo,
- una operazione di prodotto di un vettore per uno scalare, che associa a un numero reale $t \in \mathbf{R}$ e un elemento $v \in V$ un elemento $t \cdot v \in V$.

Queste operazioni devono soddisfare, oltre agli assiomi di gruppo visti in 3.2, i seguenti assiomi.

(1) per ogni $u \in V$ si ha $1u = u$.
(2) per ogni $s, t \in \mathbf{R}$ e $u \in V$ si ha $(s + t)u = su + tu$.
(3) per ogni $s, t \in \mathbf{R}$ e $u \in V$ si ha $(st)u = s(tu)$.
(4) per ogni $t \in \mathbf{R}$ e $u, v \in V$ si ha $t(u + v) = tu + tv$.

Notiamo subito alcune conseguenze.

- Per ogni $u \in V$ si ha $0u = O$. Infatti $0u = (0+0)u = 0u+0u$, e aggiungendo $-(0u)$ da entrambe le parti si trova $0u = O$.
- Per ogni $t \in \mathbf{R}$ si ha $tO = O$. Infatti $tO = t(O + O) = tO + tO$, e aggiungendo $-(tO)$ da entrambe le parti si ottiene $tO = O$.

3.3.1. **Teorema.** *$V = \mathbf{R}^n$ è uno spazio vettoriale, con la somma sopra definita, e il prodotto di un vettore per uno scalare definito da*

$$t(a_1, a_2, \dots, a_n) = (ta_1, ta_2, \dots, ta_n).$$

Enunceremo la maggior parte dei risultati per uno spazio vettoriale qualsiasi, ma tenderemo a dimostrarli per $\mathbf{R}^n$.

3.4. **Combinazione lineare.** E' comodo mischiare somma di vettori e prodotto di un vettore per uno scalare nella seguente

3.4.1. **Definizione.** Sia V uno spazio vettoriale, e siano $v_1, v_2, \dots, v_n \in V$ vettori, e $a_1, a_2, \dots, a_n \in \mathbf{R}$ scalari.

La *combinazione lineare* dei v_i con cofficienti a_i è il vettore

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n.$$

4. Matrici

4.1. **Matrici.** Quando abbiamo a che fare ad esempio con un sistema di equazioni lineari omogene

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = 0 \\ \ddots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = 0 \end{cases} \tag{4.1}$$

è utile operare direttamente sui suoi coefficienti a_{ij}, senza portarsi appresso anche le incognite. Si considera una tabella con m righe e n colonne

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \tag{4.2}$$

che viene detta la *matrice* $m \times n$ dei coefficienti del sistema.

Un vettore $v = (v_1, v_2, \dots, v_n) \in \mathbf{R}^n$ può anche essere visto come una matrice, addirittura in due modi,

(1) come vettore riga $\begin{bmatrix} v_1 v_2 \dots v_n \end{bmatrix}$, dunque come una matrice $1 \times n$, o
(2) come vettore colonna

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},$$

dunque come una matrice $n \times 1$.

4.2. **Le matrici come spazio vettoriale.** Fissati m, n, consideriamo l'insieme V di tutte le matrici $m \times n$ a coefficienti reali.

Allora V diventa uno spazio vettoriale, con la somma definita mediante

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} & \dots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \dots & b_{2n} \\ & \ddots & & & \\ b_{m1} & b_{m2} & b_{m3} & \dots & b_{mn} \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} & \dots & a_{2n}+b_{2n} \\ & & \ddots & & \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & a_{m3}+b_{m3} & \dots & a_{mn}+b_{mn} \end{bmatrix}$$

e il prodotto per uno scalare t definito mediante

$$t \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ta_{11} & ta_{12} & ta_{13} & \dots & ta_{1n} \\ ta_{21} & ta_{22} & ta_{23} & \dots & ta_{2n} \\ & \ddots & & & \\ ta_{m1} & ta_{m2} & ta_{m3} & \dots & ta_{mn} \end{bmatrix}$$

4.3. **Prodotto fra matrici.** Supponiamo di avere A, una matrice $m \times h$, e B, una matrice $k \times n$. Se $h = k$ (cioè il numero di colonne di A è eguale al numero di righe di B), si può definire un prodotto fra matrici AB, che vedremo tornerà utilissimo.

Cominciamo col definire il prodotto di un vettore riga $\begin{bmatrix} u_1 u_2 \dots u_n \end{bmatrix}$ per un vettore colonna

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Questo è una matrice 1×1 (dunque un numero fra parentesi quadre), ed è il prodotto scalare $u \cdot v$, dunque

$$\begin{bmatrix} u_1, u_2, \dots, u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 v_1 + u_2 v_2 + \dots + u_n v_n \end{bmatrix}.$$

Adesso se A è una matrice $m \times k$, e B è una matrice $k \times n$, la matrice prodotto AB è una matrice $m \times n$, in cui nel posto (i, j) (dunque all'incrocio della i-sima riga con la j-sima colonna) c'è il prodotto (scalare) della i-sima riga di A per la j-sima colonna di B.

Vediamo alcuni esempi.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ -1 & 3 \end{bmatrix}.$$

Dato che A è 2×3, e B è 3×2, si può fare il prodotto, e viene la matrice 2×2

$$\begin{bmatrix} 1 \cdot 2 + 2 \cdot 1 + 3 \cdot (-1) & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 3 \\ 0 \cdot 2 + 1 \cdot 1 + (-1) \cdot (-1) & 0 \cdot 0 + 1 \cdot 1 + (-1) \cdot 3 \end{bmatrix} = \begin{bmatrix} 1 & 11 \\ 2 & -2 \end{bmatrix}.$$

Ma si può fare anche il prodotto BA, e viene la matrice 3×3

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 3 & 2 \\ -1 & 1 & -6 \end{bmatrix}.$$

Dunque AB e BA possono esistere entrambe, ma avere dimensioni diverse.

Ma anche quando AB e BA esistono entrambe, e hanno la stessa dimensione, possono essere diverse, come si vede dall'esempio seguente.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad BA = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

dunque $AB \neq BA$. Inoltre abbiamo visto che il prodotto di due matrici diverse da zero può ben essere zero.

Se

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

dato che A è 2×2, e x è 2×1, si può fare solo il prodotto Ax, e viene la matrice 2×1

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}. \tag{4.3}$$

4.4. **Legami fra le operazioni.** Si può vedere che le operazione di somma, prodotto, e prodotto per scalare che abbiamo introdotto godono delle seguenti proprietà.

4.4.1. **Proposizione.** *Siano A, B, C matrici per cui le operazioni qui sotto abbiano senso, e siano $s, t \in \mathbf{R}$ scalari.*

1. $(sA + tB)C = sAC + tBC$.
2. $A(sB + tC) = sAB + tAC$.
3. $(AB)C = A(BC)$.

Le prime due proprietà dicono che il prodotto è bilineare; esse seguono dalle corrispondenti proprietà del prodotto scalare. L'ultima proprietà è semplicemente quella associativa.

5. Sistemi lineari e matrici

5.1. **Sistemi e sistemi omogenei.** Consideriamo un sistema di m equazioni elementari in n incognite

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2 \\ \ddots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m \end{cases}. \tag{5.1}$$

Notiamo che il sistema si può riscrivere nella forma matriciale

$$Ax = b, \tag{5.2}$$

ove A è

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}.$$

Se x_0 è una soluzione fissata di (5.2) (sempre ammesso che il sistema ammetta soluzione), e x una soluzione qualsiasi, allora si ha

$$A(x - x_0) = Ax - Ax_0 = b - b = O.$$

Dunque una soluzione del sistema è della forma $x = x_0 + y$, ove x_0 è una soluzione fissata, e y è una soluzione del *sistema omogeneo associato*

$$Ax = O. \tag{5.3}$$

Interessiamoci quindi dapprima ai sistemi omogenei.

5.2. **Sottospazi.**

5.2.1. **Definizione.** Sia V uno spazio vettoriale, U un suo sottoinsieme.

U si dice un sottospazio (vettoriale) di V se

(1) $O \in U$,
(2) per ogni $u, w \in U$ si ha $u + w \in U$, e
(3) per ogni $u \in U$ e $t \in \mathbf{R}$ si ha $tu \in U$.

Le condizioni (2) e (3) si possono unificare in una sola,

(2') per ogni $u, w \in U$, e $s, t \in \mathbf{R}$, la combinazione lineare $su + tw$ sta ancora in U.

Se U è un sottospazio di V, allora anche U è uno spazio vettoriale, con le stesse operazioni di V.

5.2.2. **Proposizione.** *Sia V uno spazio vettoriale, $v_1, v_2, \dots, v_m \in V$. Allora l'insieme*

$$S = \{\, a_1 v_1 + a_2 v_2 + \dots + a_m v_m : a_i \in \mathbf{R} \,\} \tag{5.4}$$

di tutte le combinazioni lineari dei v_i è un sottospazio vettoriale di V.

Proof. O si ottiene quando tutti gli a_i sono zero.

E poi, se $u, w \in S$

$$u = a_1 v_1 + a_2 v_2 + \dots + a_m v_m, \qquad w = b_1 v_1 + b_2 v_2 + \dots + b_m v_m,$$

e $s, t \in \mathbf{R}$, allora

$$su + tv = (sa_1 + tb_1)v_1 + (sa_2 + tb_2)v_2 + \dots + (sa_1 + tb_1)v_m \in S.$$

□

Ora vogliamo vedere come i sottospazi saltano fuori in modo naturale quando risolviamo i sistemi di equazioni lineari.

5.3. **Sistemi omogenei.**

5.3.1. **Proposizione.**
L'insieme $\mathcal{U}$ *l'insieme delle soluzioni di* (4.1) *(ovvero* (5.3)*) è un sottospazio vettoriale di* $\mathbf{R}^n$*.*

Proof. E' chiaro che $AO = O$, dunque $O \in \mathcal{U}$. E poi, se u, v sono soluzioni, e $s, t \in \mathbf{R}$, si ha

$$A(su + tv) = s(Au) + t(Av) = sO + tO = O.$$

□

5.4. **Basi.**

5.4.1. **Definizione.** Una *base* (finita) di uno spazio vettoriale V è un sistema $v_1, \dots, v_n$ di vettori tali che *ogni elemento di* V *si scriva in modo unico come combinazione lineare dei* v_i, cioè nella forma

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n. \tag{5.5}$$

L'unicità vuol dire che se posso scrivere

$$a_1 v_1 + a_2 v_2 + \dots + a_n v_n = b_1 v_1 + b_2 v_2 + \dots + b_n v_n,$$

con $a_i, b_i \in \mathbf{R}$, allora $a_i = b_i$ per ogni i.

Vi sono alcuni fatti importanti da sapere subito, di cui, almeno per ora, omettiamo la dimostrazione.

5.4.2. **Teorema.**

(1) *Ogni spazio vettoriale* V *ha una base.*
(2) *Se* V *ha due basi, una di* m *elementi, e una di* n *elementi, allora* $m = n$.

Il numero di elementi di una base di V *si dirà* dimensione *di* V*.*

$\mathbf{R}^n$ ha dimensione n, perché una base (detta *base standard*) è data dai vettori

$$\begin{cases} e_1 = (1, 0, 0, \dots, 0), \\ e_2 = (0, 1, 0, \dots, 0), \\ e_3 = (0, 0, 1, \dots, 0), \\ \dots, \\ e_n = (0, 0, 0, \dots, 1). \end{cases} \tag{5.6}$$

Infatti se prendo $v = (v_1, v_2, \dots, v_n) \in \mathbf{R}^2$, e

$$v = a_1 e_1 + a_2 e_2 + \dots + a_n e_n,$$

allora

$$(v_1, v_2, \dots, v_n) = a_1 e_1 + a_2 e_2 + \dots + a_n e_n = (a_1, a_2, \dots, a_n),$$

da cui $a_i = v_i$ per ogni i.

Ma, attenzione, $\mathbf{R}^n$ di basi ne ha tantissime! Solo per fare un esempio, $e_1 + e_2, e_2$ è un'altra base di $\mathbf{R}^2$, perché per ogni $v = (v_1, v_2)$, se

$$v = (v_1, v_2) = a_1(e_1 + e_2) + a_2 e_2 = a_1 e_1 + (a_1 + a_2) e_2,$$

allora

$$\begin{cases} a_1 \qquad = v_1 \\ a_1 + a_2 = v_2, \end{cases}$$

che visto come un sistema nelle incognite a_1, a_2 ha l'unica soluzione $a_1 = v_1$ e $a_2 = v_2 - v_1$.

5.5. **Sistemi di generatori e sistemi linearmente indipendenti.** Qui spezziamo in qualche modo in due la definizione di *base*.

5.5.1. **Definizione.** Se V è uno spazio vettoriale, i suoi elementi $v_1, \dots, v_n$ si dicono

- un *sistema di generatori* di V se ogni elemento di V si scrive come combinazione lineare dei v_i,
- *linearmente indipendenti* se per ogni scelta di $a_i \in \mathbf{R}$, se la combinazione lineare $a_1 v_1 + a_2 v_2 + \dots + a_n v_n$ è zero, allora ogni a_i è zero.

Attenzione! La definizione di lineare indipendenza

- è difficile, anche se magari non sembra,
- dovete saperla, e
- ve la chiedo sicuramente nel compito.

Vediamo due esempi.

Consideriamo i vettori $v_1 = (1, 2, 1), v_2 = (1, 1, 1), v_3 = (1, 1, 0) \in \mathbf{R}^3$. Siano $a_i \in \mathbf{R}$ tali che

$$a_1 v_1 + a_2 v_2 + a_3 v_3 = (a_1 + a_2 + a_3, 2a_1 + a_2 + a_3, a_1 + a_2) = O.$$

Risolviamo il sistema di tre equazioni nelle tre incognite a_1, a_2, a_3, col metodo già visto per due equazioni in due incognite, e che rivedremo in generale nel seguito

$$\begin{cases} a_1 + a_2 + a_3 = 0 \\ 2a_1 + a_2 + a_3 = 0 \\ a_1 + a_2 = 0 \end{cases}$$

Sottraiamo dalla terza equazione la prima, e dalle seconda due volte la prima, per ottenere il sistema

$$\begin{array}{ccccccc} a_1 & + & a_2 & + & a_3 & = & 0 \\ & - & a_2 & - & a_3 & = & 0 \\ & & & & a_3 & = & 0 \end{array}$$

che ha le stesse soluzioni. Leggendo dal basso in alto ho nell'ordine $a_3 = a_2 = a_1 = 0$, dunque questi tre vettori sono linearmente indipendenti.

Consideriamo ora i vettori $v_1 = (1, 2, 1), v_2 = (1, 1, 2), v_3 = (0, 1, -1) \in \mathbf{R}^3$. Stavolta il sistema da risolvere è

$$\begin{cases} a_1 + a_2 \qquad = 0 \\ 2a_1 + a_2 + a_3 = 0 \\ a_1 + 2a_2 - a_3 = 0 \end{cases}$$

Sottraiamo come prima dalla terza equazione la prima, e dalle seconda due volte la prima, per ottenere

$$\begin{array}{ccccccc} a_1 & + & a_2 & + & & = & 0 \\ & - & a_2 & + & a_3 & = & 0 \\ & & a_2 & - & a_3 & = & 0 \end{array}$$

Ora sottraiamo dalla terza equazione la seconda, ottenendo

$$\begin{array}{ccccccc} a_1 & + & a_2 & + & & = & 0 \\ & - & a_2 & + & a_3 & = & 0 \\ & & & & 0 & = & 0 \end{array}$$

Dunque la terza equazione è scomparsa. Allora posso scegliere ad esempio $a_3 = 1$, e ottenere $a_2 = 1$ e $a_1 = -1$. (Più avanti chiariremo quali sono tutte le possibilità di scelta in una simile situazione.) Attenzione ora! Abbiamo trovato una combinazione lineare dei tre vettori che fa zero

$$(-1)v_1 + 1 \cdot v_2 + 1 \cdot v_3 = O.$$

Siccome i tre coefficienti della combinazione lineare $-1, 1, 1$ non sono tutti e tre nulli, questo vuol dire che i tre vettori *non* sono linearmente indipendenti. (E allora si dice, con una certa inventiva, che sono *linearmente dipendenti.*)

Dunque un altro modo di enunciare il fatto che i vettori $v_1, v_2, \dots, v_n$ sono linearmente indipendenti è dire che sono equivalenti

- $v_1, v_2, \dots, v_n$ sono linearmente indipendenti, e
- l'equazione

(5.7) $$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = 0$$

  nella incognite $a_1, a_2, \dots, a_n$ ha come unica soluzione $a_1 = a_2 = \cdots = a_n = 0$.

Se i vettori v_i sono in $\mathbf{R}^k$, abbiamo appena visto negli esempi che l'equazione (5.7) equivale a un sistema di k equazioni lineari omogenee.

5.5.2. **Proposizione.** *Siano $v_1, v_2, \dots, v_n$ vettori linearmente indipendenti. Se $a_i, b_i \in \mathbf{R}$, e*

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = b_1 v_1 + b_2 v_2 + \cdots + b_n v_n,$$

allora $a_i = b_i$ per ogni i.

Gli elementi di una base sono senz'altro un sistema di generatori (questo è immediato), e sono anche linearmente indipendenti. Infatti se $v_1, v_2, \dots, v_n$ è una base, e $a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = O = 0 v_1 + 0 v_2 + \cdots + 0 v_n$, per definizione di base si deve avere $a_i = 0$ per ogni i. Abbiamo quindi la prima metà del

5.5.3. **Teorema.**

(1) *Una base è un sistema di generatori linearmente indipendente.*
(2) *Un sistema di generatori linearmente indipendente è una base.*

Proof. Per il secondo punto, resta solo da notare che se $v_1, v_2, \dots, v_m$ sono linearmente indipendenti, e

$$a_1 v_1 + a_2 v_2 + \cdots + a_m v_m = b_1 v_1 + b_2 v_2 + \cdots + b_m v_m,$$

allora

$$(a_1 - b_1) v_1 + (a_2 - b_2) v_2 + \cdots + (a_m - b_m) v_m = 0,$$

e dunque, per la lineare indipendenza si ha $a_i - b_i = 0$, ovvero $a_i = b_i$ per ogni i. □

5.6. **Risolvere un sistema.**

5.7. **Operazioni elementari.** Sia A una matrice $m \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ & \ddots & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}.$$

Consideriamo alcune *operazioni elementari* su di essa.

(1) Scambiare fra loro due righe. (Indicheremo con S_{ij} l'operazione che scambia fra loro le righe i-sima e j-sima.)
(2) Aggiungere a una riga un multiplo di un'altra riga. (Indicheremo con $E_{ij}(c)$ l'operazione che alla riga i-sima aggiunge la riga j-sima, moltiplicata per lo scalare $c \in \mathbf{R}$.)
(3) Moltiplicare una riga per un numero diverso da zero. (Indicheremo con $D_i(c)$ l'operazione che moltiplica la riga i-sima per lo scalare $c \neq 0$.)

Da fare: Interpretazione in termini di prodotto per matrici.

Con le sole prime due operazioni possiamo mettere la matrice in una forma (anzi, due) che ci sarà particolarmente utile.

(1) Consideriamo la prima colonna di A. Se tutti i suoi elementi sono zero, saltiamo al passo (4), eliminiamo la prima colonna (o meglio, ignoriamola), e ripetiamo il procedimento dall'inizio.
(2) Se non tutti i suoi elementi sono zero, facciamo uno scambio di righe, in modo che sia $a_{11} \neq 0$. Questo elemento lo chiamiamo un *pivot*.
(3) Poi aggiungiamo a ogni riga a partire dalla seconda un opportuno multiplo della prima, in modo da annullare il primo elemento della riga. Dunque alla riga i-sima, aggiungiamo la prima riga moltiplicata per $-a_{11}^{-1} a_{i1}$. Dopo questa operazione, e continuando a denotare la matrice risultante come A in (4.2), avremo $a_{i1} = 0$ per ogni $i > 1$.
(4) A questo punto ritorniamo al passo (1), applicandolo alla matrice che si ottiene eliminando (o meglio, ignorando) da A la prima riga e la prima colonna.

Vale forse la pena dire che l'algoritmo termina quando mi ritrovo la matrice vuota, cioè delle parentesi quadre senza nulla dentro.

Alla fine avremo ottenuto una matrice *a scalini*. Questo vuol dire che presa una riga qualsiasi, diciamo la i-sima,

- o essa è fatta tutta di zeri,
- oppure se a_{it} è il primo elemento diverso da zero, allora tutti gli elementi a_{jk} per $j > i$ e $k \geq t$ sono zero.

In altre parole, e più semplicemente, *il numero di elementi di una riga, che precedono il primo elemento non nullo della riga stessa, cresce da una riga alla successiva.*

Una volta ottenuta la matrice a scalini, possiamo fare due ulteriori trasformazioni. La prima consiste nell'annullare gli elementi della colonna *sopra* ogni pivot. Se un pivot è a_{ij}, questo si ottiene aggiungendo alla riga k-sima, per $k < i$, la riga i-sima, moltiplicata per $a_{ij}^{-1} a_{kj}$. Questo ci porta a una forma che chiamiamo (il nome è mio) *qRREF*, che sta per *quasi Row Reduced Echelon Form*, ovvero *forma a scalini quasi ridotta per righe.*

Infine posso moltiplicare una riga non nulla per l'inverso del suo pivot, ottenendo quella che si chiama la *Row Reduced Echelon Form* (*forma a scalini ridotta per righe*), che è come la qRREF, ma in più il primo elemento non nullo su ciascuna riga non nulla è 1. Vale l'importante

5.7.1. **Teorema.** *La forma RREF è unica.*

La dimostrazione di questo risultato è fuori dagli scopi del corso, ma ne faremo un uso essenziale nel seguito.

5.8. **Risolvere un sistema omogeneo.** Due sistemi omogenei si dicono *equivalenti* se hanno lo stesso insieme di soluzioni.

Ora è chiaro che l'insieme delle soluzioni di un sistema non cambia se scambio fra loro due equazioni. Questo corrisponde a fare un'operazione di scambio di righe nella matrice dei coefficienti.

L'insieme delle soluzioni non cambia neanche se moltiplico una equazione per un numero diverso da zero. Questo corrisponde a moltiplicare una riga della matrice per quel numero (diverso da zero).

Infine, se due delle equazioni sono $P = 0$ e $Q = 0$, l'insieme delle soluzioni non cambia se sostituisco l'equazione $P = 0$ con $P + cQ = 0$, ove $c \in \mathbf{R}$. Questo perché se $P = 0$ e $Q = 0$, allora $P + cQ = 0 + c0 = 0$, e viceversa se $P + cQ0$ e $Q = 0$, allora $P = (P + cQ) - cQ = 0 - c0 = 0$. Questo corrisponde a sostituire un'operazione del tipo $E_{ij}(c)$ nella matrice.

L'idea allora è di usare queste operazioni, sul sistema o sulla matrice ad esso associata, in modo da ridurlo a un sistema equivalente, ma molto più semplice. Facciamo un esempio. Consideriamo il sistema

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ x_1 - 2x_2 - x_3 = 0 \\ 2x_1 - x_2 + 3x_3 = 0. \end{cases} \tag{5.8}$$

Operiamo sulla matrice associata

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & -1 \\ 2 & -1 & 3 \end{bmatrix}$$

nel modo seguente

$$\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & -1\\ 2 & -1 & 3\end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 2 & -1 & 3\end{bmatrix} \xrightarrow{E_{31}(-2)}$$

$$\xrightarrow{E_{31}(-2)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 0 & -3 & 1\end{bmatrix} \xrightarrow{E_{32}(-1)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 0 & 0 & 3\end{bmatrix}.$$

Dunque il sistema (5.8) equivale al *sistema a gradini*

$$\begin{cases} x_1 & + & x_2 & + & x_3 & = & 0\\ & & -3x_2 & - & 2x_3 & = & 0\\ & & & & 3x_3 & = & 0.\end{cases}$$

che si risolve facilmente leggendo le equazioni dal basso verso l'alto. La terza mi dice $x_3 = 0$. Sostituendo $x_3 = 0$ nella seconda ottengo $x_2 = 0$. Sostituendo $x_2 = x_3 = 0$ nella prima ottengo $x_1 = 0$, che è dunque l'unica soluzione.

Consideriamo ora il sistema

$$\begin{cases} x_1 + x_2 + x_3 = 0\\ x_1 - 2x_2 - x_3 = 0\\ 3x_1 - 3x_2 - x_3 = 0.\end{cases} \tag{5.9}$$

Stavolta abbiamo la matrice associata

$$\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & -1\\ 3 & -3 & -1\end{bmatrix}$$

e procediamo come segue

$$\begin{bmatrix}1 & 1 & 1\\ 1 & -2 & -1\\ 3 & -3 & -1\end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 3 & -3 & -1\end{bmatrix} \xrightarrow{E_{31}(-3)}$$

$$\xrightarrow{E_{31}(-3)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 0 & -6 & -4\end{bmatrix} \xrightarrow{E_{32}(-2)} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -2\\ 0 & 0 & 0\end{bmatrix}.$$

Dunque otteniamo il sistema equivalente

$$\begin{cases} x_1 & + & x_2 & + & x_3 & = & 0\\ & & -3x_2 & - & 2x_3 & = & 0\end{cases} \tag{5.10}$$

dato che l'ultima equazione è semplicemente $0 = 0$, che è soddisfatta per qualunque valore di x_1, x_2, x_3. (Notiamo quindi che il numero di equazioni di un sistema non è di per sé significativo, conta invece quante ne rimangono dopo questo procedimento.) Per risolvere (5.9) conviene operare un'ulteriore riduzione. Consideriamo la matrice dei coefficienti di(5.9)

$$\begin{bmatrix}\mathbf{1} & 1 & 1\\ 0 & -\mathbf{3} & -2\end{bmatrix}$$

Il primo elemento non nullo di una riga si chiama un *pivot*. In questo caso i pivot sono quelli segnati in neretto. Cominciamo con l'eliminare ogni elemento diverso da zero in ogni colonna che contiene un pivot. Questo si ottiene facendo

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & -2 \end{bmatrix} \xrightarrow{E_{12}(1/3)} \begin{bmatrix} 1 & 0 & 1/3 \\ 0 & -3 & -2 \end{bmatrix}.$$

Poi moltiplichiamo ogni riga per l'inverso del suo pivot

$$\begin{bmatrix} 1 & 0 & 1/3 \\ 0 & -3 & -2 \end{bmatrix} \xrightarrow{D_2(-1/3)} \begin{bmatrix} 1 & 0 & 1/3 \\ 0 & 1 & 2/3 \end{bmatrix},$$

ottenendo la forma RREF. A questo punto il sistema corrispondente è

$$\begin{cases} x_1 \;+ & & \frac{1}{3}x_3 & = & 0 \\ & x_2 \;+ & \frac{2}{3}x_3 & = & 0 \end{cases} \tag{5.11}$$

ovvero

$$\begin{cases} x_1 & & = & -\frac{1}{3}x_3 \\ & x_2 & = & -\frac{2}{3}x_3 \end{cases} \tag{5.12}$$

Ora in una qualsiasi soluzione x_3 un valore dovrà averlo, giusto? Qui vediamo che se diamo a x_3 in (5.12) un valore *arbitrario* t, da (5.12) ricaviamo valori $x_1 = -\frac{1}{3}t$ e $x_2 = -\frac{2}{3}t$. Dunque le soluzioni del sistema (5.9) sono tutte e sole della forma

$$(x_1, x_2, x_3) = t\left(-\frac{1}{3}, -\frac{2}{3}, 1\right),$$

una retta nello spazio passante per l'origine O.

Si dice che le variabili corrispondenti a colonne dove c'è un pivot sono *dipendenti*, mentre le altre sono *libere*. In questo caso x_3 è libera, e in effetti le abbiamo dato un valore arbitrario, mentre x_1, x_2 sono dipendenti, nel senso che il loro valore dipende da quello di x_3.

Vediamo un altro esempio

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 - 2x_2 - x_3 - 3x_4 = 0 \\ 3x_1 - 3x_2 - x_3 - 5x_4 = 0. \end{cases} \tag{5.13}$$

Abbiamo

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -2 & -1 & -3 \\ 3 & -3 & -1 & -5 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 \\ 3 & -3 & -1 & -5 \end{bmatrix} \xrightarrow{E_{31}(-3)}$$

$$\xrightarrow{E_{31}(-3)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 \\ 0 & -6 & -4 & -8 \end{bmatrix} \xrightarrow{E_{32}(2)} \begin{bmatrix} \mathbf{1} & 1 & 1 & 1 \\ 0 & \mathbf{-3} & -2 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{E_{12}(1/3)}$$

$$\xrightarrow{E_{12}(1/3)} \begin{bmatrix} 1 & 0 & 1/3 & -1/3 \\ 0 & -3 & -2 & -4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{D_2(-1/3)} \begin{bmatrix} 1 & 0 & 1/3 & -1/3 \\ 0 & 1 & 2/3 & 4/3 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

dove ho messo in neretto i pivot. Dunque x_1, x_2 sono le variabili dipendenti, mentre x_3, x_4 sono libere. Otteniamo il sistema equivalente

$$\begin{cases} x_1 = -\frac{1}{3}x_3 + \frac{1}{3}x_4 \\ x_2 = -\frac{2}{3}x_3 - \frac{4}{3}x_4 \end{cases} \tag{5.14}$$

Dando valori arbitrari $x_3 = s$, $x_4 = t$, otteniamo che le soluzioni sono della forma

$$(x_1, x_2, x_3, x_4) = s\left(-\frac{1}{3}, -\frac{2}{3}, 1, 0\right) + t\left(\frac{1}{3}, -\frac{4}{3}, 0, 1\right).$$

Vediamo un altro esempio, dove capita la possibilità considerata al punto (1) dell'algoritmo.

$$\begin{cases} x_1 + x_2 - x_3 + 2x_4 + 3x_5 = 0 \\ x_1 + 4x_2 + 3x_4 + 4x_5 = 0 \\ 3x_1 + 6x_2 - 2x_3 + 9x_4 + 11x_5 = 0 \\ 2x_1 + 2x_2 - 2x_3 + 4x_4 + 5x_5 = 0. \end{cases} \tag{5.15}$$

Abbiamo

$$\begin{bmatrix} 1 & 1 & -1 & 2 & 3 \\ 1 & 4 & 0 & 3 & 4 \\ 3 & 6 & -2 & 9 & 11 \\ 2 & 2 & -2 & 4 & 5 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 1 & -1 & 2 & 3 \\ 0 & 3 & 1 & 1 & 1 \\ 3 & 6 & -2 & 9 & 11 \\ 2 & 2 & -2 & 4 & 5 \end{bmatrix} \xrightarrow{E_{31}(-3)}$$

$$\xrightarrow{E_{31}(-3)} \begin{bmatrix} 1 & 1 & -1 & 2 & 3 \\ 0 & 3 & 1 & 1 & 1 \\ 0 & 3 & 1 & 3 & 2 \\ 2 & 2 & -2 & 4 & 5 \end{bmatrix} \xrightarrow{E_{41}(-2)} \begin{bmatrix} 1 & 1 & -1 & 2 & 3 \\ 0 & 3 & 1 & 1 & 1 \\ 0 & 3 & 1 & 3 & 2 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} \xrightarrow{E_{32}(-1)}$$

$$\xrightarrow{E_{32}(-1)} \begin{bmatrix} \mathbf{1} & 1 & -1 & 2 & 3 \\ 0 & \mathbf{3} & 1 & 1 & 1 \\ 0 & 0 & 0 & \mathbf{2} & 1 \\ 0 & 0 & 0 & 0 & -\mathbf{1} \end{bmatrix}$$

dove ho messo in neretto i pivot. Qui parto dalla matrice

$$\begin{bmatrix} 1 & 1 & -1 & 2 & 3 \\ 1 & 4 & 0 & 3 & 4 \\ 3 & 6 & -2 & 9 & 11 \\ 2 & 2 & -2 & 4 & 5 \end{bmatrix}$$

col passo (4) dell'algoritmo, cancellando prima riga e prima colonna dopo aver messo a zero tutti gli elementi della prima colonna, mi sono ridotto a

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 3 & 1 & 3 & 2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

poi metto di nuovo a zero tutti gli elementi della prima colonna tranne il primo, cancello prima riga e prima colonna e trovo

$$\begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

dove la prima colonna è zero, e quindi applico il passo (1), e la cancello semplicemente, e ho

$$\begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$$

poi

$$\begin{bmatrix} -1 \end{bmatrix}$$

e qui applicando il passo (4) la matrice è vuota, e ho finito.

In generale, la riduzione sopra effettuata permette di raggiungere dapprima una forma a gradini, in cui come già detto chiamiamo *pivot* i primi elementi non zero di ogni riga che sia non zero. Il numero di pivot si dice *rango* del sistema o della matrice a lui associata. Ora le variabili delle colonne in cui ci sono i pivot si dicono *dipendenti*, le altre si dicono *libere*. Facendo la riduzione completa in forma RREF, un sistema si mette quindi in una forma del tipo

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & b_{1,k+1} & \dots & b_{1n} \\ 0 & 1 & 0 & \dots & 0 & b_{2,k+1} & \dots & b_{2n} \\ 0 & 0 & 1 & \dots & 0 & b_{3,k+1} & \dots & b_{3n} \\ & & & \ddots & & & & \\ 0 & 0 & 0 & \dots & 1 & b_{k,k+1} & \dots & b_{kn} \end{bmatrix} \tag{5.16}$$

Qui k è il rango, e per semplicità ho assunto che le variabili libere siano le ultime $n-k$ (cioè $x_{k+1}, \dots, x_n$), mentre quelle dipendenti sono le prime k (cioè $x_1, \dots, x_k$).

Ora notiamo una cosa importante. La matrice (5.16) corrisponde al sistema

$$\begin{cases} x_1 & = -b_{1,k+1}x_{k+1} - \dots - b_{1n}x_n \\ \quad x_2 & = -b_{2,k+1}x_{k+1} - \dots - b_{2n}x_n \\ \qquad x_3 & = -b_{3,k+1}x_{k+1} - \dots - b_{3n}x_n \\ \qquad\quad \ddots & \\ \qquad\qquad x_k & = -b_{k,k+1}x_{k+1} - \dots - b_{kn} \end{cases} \tag{5.17}$$

Consideriamo le $n-k$ soluzioni $v_1, v_2, \dots, v_{n-k}$ che si ottengono dando al vettore $(x_{k+1}, x_{k+2}, \dots, x_n)$ valori nella base

$$(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)$$

di $\mathbf{R}^{n-k}$. Se

$$v = (\dots, t_1, t_2, \dots, t_{n-k})$$

è una soluzione qualsiasi del sistema allora, dato che le soluzioni formano un sottospazio vettoriale, anche

$$v - t_1 v_1 - t_2 v_2 - \dots - t_{n-k} v_{n-k}$$

è una soluzione. E si vede subito che per questa soluzione i valori di $x_{k+1}, \dots, x_n$ sono zero. Dunque per (5.17) anche i valori di $x_1, \dots, x_k$ sono zero, ovvero

$$v = t_1 v_1 - t_2 v_2 - \dots - t_{n-k} v_{n-k}.$$

5.8.1. **Definizione.** Sia data una matrice A di dimensione $m \times n$. Il numero k di pivot in una sua forma a gradini si dice *rango*. Il numero $n-k$ si dice *nullità*.

Dunque se consideriamo il sistema omogeneo $AX = O$ associato ad A, il rango è il numero di variabili dipendenti, mentre la nullità è quello delle variabili libere. Possiamo ora formulare il seguente importante

5.8.2. **Teorema.** *Sia dato un sistema di rango k e nullità $n-k$.*

Allora il sottospazio delle soluzioni ha dimensione pari alla nullità $n-k$, ovvero il sottospazio delle soluzioni ha una base di $n-k$ elementi.

Una base dello spazio delle soluzioni si ottiene dando alle variabili libere i valori in una base di $\mathbf{R}^{n-k}$, e ricavando le variabili dipendenti.

Il concetto di *dimensione* formalizza quello di *numero di parametri* da cui dipendono le soluzioni.

5.9. **Sistemi non omogenei.** Con i sistemi non omogenei, si procede come sopra, ma considerando la *matrice completa dei coefficienti.* Ad esempio se parto da

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 - 2x_2 - x_3 = 2 \\ 2x_1 - x_2 + 3x_3 = 3. \end{cases} \tag{5.18}$$

Operiamo sulla matrice associata

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -2 & -1 & 2 \\ 2 & -1 & 3 & 3 \end{bmatrix}$$

nel modo seguente

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -2 & -1 & 2 \\ 2 & -1 & 3 & 3 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & 1 \\ 2 & -1 & 3 & 3 \end{bmatrix} \xrightarrow{E_{31}(-2)}$$
$$\xrightarrow{E_{31}(-2)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & 1 \\ 0 & -3 & 1 & 1 \end{bmatrix} \xrightarrow{E_{32}(-1)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & 1 \\ 0 & 0 & 3 & 0 \end{bmatrix}.$$

Dunque il sistema (5.18) equivale al *sistema a gradini*

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ -3x_2 - 2x_3 = 1 \\ 3x_3 = 0. \end{cases}$$

che si risolve facilmente leggendo le equazioni dal basso verso l'alto. Ho dapprima $x_3 = 0$, poi $-3x_2 = 1$, dunque $x_2 = -1/3$, e infine $x_1 = 4/3$.

Ma guardate cosa succede in quest'altro esempio.

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 1 \\ x_1 - 2x_2 - x_3 - 3x_4 = 1 \\ 3x_1 - 3x_2 - x_3 - 5x_4 = 0. \end{cases} \tag{5.19}$$

Abbiamo come sopra per la matrice completa dei coefficienti

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & -2 & -1 & -3 & 1 \\ 3 & -3 & -1 & -5 & 0 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 & 0 \\ 3 & -3 & -1 & -5 & 0 \end{bmatrix} \xrightarrow{E_{31}(-3)}$$

$$\xrightarrow{E_{31}(-3)} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 & 0 \\ 0 & -6 & -4 & -8 & -3 \end{bmatrix} \xrightarrow{E_{32}(-2)} \begin{bmatrix} \mathbf{1} & 1 & 1 & 1 & 1 \\ 0 & -\mathbf{3} & -2 & -4 & 0 \\ 0 & 0 & 0 & 0 & -\mathbf{3} \end{bmatrix}.$$

dove i pivot sono in neretto. Vedete che c'è un pivot nell'ultima colonna? Questo significa che il sistema corrispondente ha l'ultima equazione della forma

$$0 = -1,$$

e dunque il sistema non ha soluzione. Abbiamo ottenuto la prima parte dell'importante

5.9.1. **Teorema.**

(1) *Un sistema non omogeneo non ha soluzione se c'è un pivot nell'ultima colonna.*

(2) *Se non ci sono pivot nell'ultima colonna, una soluzione particolare si può trovare dando ad esempio valori zero alle variabili libere. Una soluzione qualsiasi sarà la somma di questa soluzione particolare e di una soluzione del sistema omogeno associato.*

Come esempio di sistema non omogeneo che ha soluzioni, consideriamo

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 1 \\ x_1 - 2x_2 - x_3 - 3x_4 = 1 \\ 3x_1 - 3x_2 - x_3 - 5x_4 = 3. \end{cases} \tag{5.20}$$

Abbiamo come sopra

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & -2 & -1 & -3 & 1 \\ 3 & -3 & -1 & -5 & 3 \end{bmatrix} \xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 & 0 \\ 3 & -3 & -1 & -5 & 3 \end{bmatrix} \xrightarrow{E_{31}(-3)}$$

$$\xrightarrow{E_{31}(-3)} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -3 & -2 & -4 & 0 \\ 0 & -6 & -4 & -8 & 0 \end{bmatrix} \xrightarrow{E_{32}(-2)} \begin{bmatrix} \mathbf{1} & 1 & 1 & 1 & 1 \\ 0 & -\mathbf{3} & -2 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{E_{12}(1/3)}$$

$$\xrightarrow{E_{12}(1/3)} \begin{bmatrix} 1 & 0 & 1/3 & -1/3 & 1 \\ 0 & -3 & -2 & -4 & 0 \end{bmatrix} \xrightarrow{D_2(-1/3)} \begin{bmatrix} 1 & 0 & 1/3 & -1/3 & 1 \\ 0 & 1 & 2/3 & 4/3 & 0 \end{bmatrix}.$$

Dunque una soluzione particolare la ottengo dando ad esempio valore zero alle variabile libere $x_3 = x_4 = 0$, e ricavando $x_1 = 1, x_2 = 0$. Poi dando valori arbitrari $x_3 = s$ e $x_4 = t$ trovo tutte le soluzioni nella forma

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \cdot \begin{bmatrix} -1/3 \\ -2/3 \\ 1 \\ 0 \end{bmatrix} + t \cdot \begin{bmatrix} 1/3 \\ -4/3 \\ 0 \\ 1 \end{bmatrix}.$$

5.10. **Di nuovo piani nello spazio.** Riprendiamo gli argomenti della sezione 2.9. Se a_1, a_2, a_3 non sono tutti nulli, sappiamo che le soluzioni dell'equazione

$$a_1 x_1 + a_2 x_2 + a_3 x_3 = b$$

formano un *piano*. Supponiamo per semplicità che $a_1 \neq 0$ (e dunque a_1 è l'unico pivot, x_1 è l'unica variabile dipendente, e x_2, x_3 sono libere) allora possiamo riscrivere l'equazione nella forma, usando un caso particolarissimo del procedimento appena visto,

$$x_1 = -a_1^{-1} a_2 x_2 - a_1^{-1} a_3 x^3 - a_1^{-1} b.$$

Dunque una soluzione particolare p sarà ad esempio quella che si ottiene dando valore zero alle variabili libere x_2, x_3, ottendo

$$p = (a_1^{-1} b, 0, 0).$$

Ora una soluzione qualsiasi sarà della forma $p+v$, ove v è una soluzione del sistema (beh, è una sola equazione in realtà) omogeneo associato

$$a_1 x_1 + a_2 x_2 + a_3 x_3 = 0,$$

che secondo quanto visto si ottengono dando a (x_2, x_3) i valori $(1, 0), (0, 1)$, e ricavando x_1, dunque trovando le due soluzioni

$$v_1 = (-a_1^{-1} a_2, 1, 0), \qquad v_2 = (-a_1^{-1} a_3, 0, 1),$$

e poi una soluzione si scrive in modo unico nella forma

$$t_1 v_1 + t_2 v_2,$$

ovvero le soluzioni sono

$$\begin{cases} x_1 = a_1^{-1} b + t_1(-a_1^{-1} a_2) + t_2(-a_1^{-1} a_3) \\ x_2 = \phantom{a_1^{-1} b + {}} t_1 \\ x_3 = \phantom{a_1^{-1} b + t_1(-a_1^{-1} a_2) + {}} t_2 \end{cases} \tag{5.21}$$

Se fosse anche $a_2 \neq 0$, potremmo in maniera del tutto analoga trovare le soluzioni nella forma

$$\begin{cases} x_1 = \phantom{a_2^{-1} b + {}} s_1 \\ x_2 = a_2^{-1} b + s_1(-a_2^{-1} a_1) + s_2(-a_2^{-1} a_3) \\ x_3 = \phantom{a_2^{-1} b + s_1(-a_2^{-1} a_1) + {}} s_2 \end{cases} \tag{5.22}$$

Consideriamo il caso $b = 0$, dunque le soluzioni formano uno spazio vettoriale, di cui abbiamo trovato due basi diverse, nel primo caso

$$v_1 = \begin{bmatrix} -a_1^{-1}a_2 \\ 1 \\ 0 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} -a_1^{-1}a_3 \\ 0 \\ 1 \end{bmatrix},$$

nel secondo caso

$$w_1 = \begin{bmatrix} 1 \\ -a_2^{-1}a_1 \\ 0 \end{bmatrix}, \qquad w_2 = \begin{bmatrix} 0 \\ -a_2^{-1}a_3 \\ 1 \end{bmatrix},$$

Le due basi si ottengono facilmente una dall'altra. Si ha infatti

$$w_1 = a_2^{-1}a_1v_1, \qquad w_2 = -a_2^{-1}a_3v_1 + v_2.$$

(Basta scrivere ad esempio $w_2 = xv_1 + yv_2$ e risolvere il sistema.)

In 6.3 vedremo come si fa in generale un cambio di base.

6. Basi

6.1. **Isomorfismo.** Esistono spazi vettoriali di dimensione infinita, cioè in cui ogni base ha un numero infinito di elementi. Un esempio sono i polinomi, che hanno base $1, x, \dots, x^n, \dots$.

Qui considereremo solo spazi vettoriali di dimensione finita. Supponiamo che lo spazio vettoriale abbia dimensione finita n, e fissiamo una base $v_1, \dots, v_n$. Allora possiamo considerare la funzione $f : \mathbf{R}^n \to V$ che associa a $(a_1, \dots, a_n) \in \mathbf{R}^n$ il vettore

$$a_1v_1 + \dots + a_nv_n$$

di V. Questo ci fornisce una funzione

$$\begin{aligned} f : \mathbf{R}^n &\to V \\ (a_1, \dots, a_n) &\mapsto a_1v_1 + \dots + a_nv_n. \end{aligned}$$

Questa funzione è *biiettiva.* Infatti è *iniettiva*, dato che se

$$a_1v_1 + \dots + a_nv_n = b_1v_1 + \dots + b_nv_n,$$

allora

$$(a_1, \dots, a_n) = (b_1, \dots, b_n)$$

perché i v_i sono una base, e dunque linearmente indipendenti. Inoltre è *suriettiva*, perché ogni $v \in V$ si scrive nella forma $v = a_1v_1 + \dots + a_nv_n$, perché i v_i sono una base, e dunque sono un sistema di generatori, e quindi si ha $f(v) = (a_1, \dots, a_n)$.

Dunque, *una volta fissata una base* di V, ogni elemento di V corrisponde a una n-pla di numeri reali, cioè a un elemento di $\mathbf{R}^n$. Di più, questa corrispondenza, data dalla funzione f, rispetta le operazioni, nel senso che se $f(a) = v$ e $f(b) = w$, e $s, t \in \mathbf{R}$ allora si vede subito che

$$f(sa + tb) = sv + tw = sf(a) + tf(b).$$

(In matematica si dice che f è un *isomorfismo* di spazi vettoriali.)

Per questa ragione, ci limiteremo tendenzialmente a considerare solo gli spazi vettoriali $\mathbf{R}^n$, e i loro sottospazi.

Parlare di sottospazi affini.

6.2. **Ancora basi.** Elenchiamo alcune importanti proprietà degli spazi vettoriali e delle loro basi. Tutte queste si potrebbero dimostrare, e svolgere algoritmicamente, con le riduzioni che abbiamo appreso.

6.2.1. **Teorema.** *Sia V uno spazio vettoriale.*

(1) *V ha una base.*
(2) *Due basi di V hanno lo stesso numero di elementi. (Questo numero viene detto* dimensione *di V, e indicato con* $\dim(V)$*.)*
(3) *Sia $n = \dim(V)$. Se $v_1, \dots, v_k \in V$ sono linearmente indipendenti, allora esistono $v_{k+1}, \dots, v_n$ tali che $v_1, \dots, v_n$ sia una base di V.*
(4) *Se $v_1, \dots, v_k$ sono un sistema di generatori di V, allora fra di essi si può scegliere una base di V.*

6.3. **Cambio di basi.** Sia V uno spazio vettoriale, e consideriamo due basi

$$b_1, \dots, b_n$$

e

$$c_1, \dots, c_n.$$

Allora ogni elemento della prima base si scriverà in termini della seconda, e viceversa. Dunque esisteranno s_{ij} e t_{ij} tali che

$$b_i = \sum_{j=1}^{n} s_{ji} c_j, \qquad c_i = \sum_{j=1}^{n} t_{ji} b_j.$$

Tutto questo si scrive meglio in termini di matrici. Introduco i vettori di base

$$\mathcal{B} = [b_1, \dots, b_n], \qquad \mathcal{C} = [c_1, \dots, c_n],$$

e matrici $n \times n$ $S = [s_{ij}]$ e $T = [t_{ij}]$, e avrò

$$\mathcal{B} = \mathcal{C}S, \qquad \mathcal{C} = \mathcal{B}T.$$

Avrò

$$\mathcal{B} = \mathcal{C}S = \mathcal{B}TS, \qquad \mathcal{C} = \mathcal{B}T = \mathcal{C}ST,$$

e dunque *per la definizione di base* $ST = TS = I$, ove $I = I_n$ è la *matrice identica* (o *matrice identità*) $n \times n$,

$$I = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Notate che I fa le funzioni di 1 nei numeri reali, cioè $AI = A = IA$ per ogni matrice $n \times n$ A. Dunque S, T sono una l'inversa dell'altra, e posso scrivere $T = S^{-1}$ e $S = T^{-1}$.

Sia ora $v \in V$, e scriviamo v rispetto a entrambe le basi

$$v = \mathcal{B}X = \mathcal{C}Y,$$

ove X e Y sono vettori colonna $1 \times n$, dunque

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Qui sto semplicemente dicendo in forma matriciale che

$$v = \sum_{i=1}^{n} x_i b_i = \sum_{i=1}^{n} y_i c_i.$$

Allora

$$v = \mathcal{B}X = \mathcal{C}Y = \mathcal{B}TY, \qquad v = \mathcal{C}Y = \mathcal{B}X = \mathcal{C}SX,$$

da cui $X = TY$ e $Y = SX$.

6.4. **Funzioni lineari.** In diverse applicazioni si trovano funzioni fra spazi vettoriali che soddisfano la seguente condizione.

6.4.1. **Definizione.** Siano V, W spazi vettoriali. Una funzione $f : V \to W$ si dice *lineare* se per ogni $v_1, v_2 \in V$ e ogni $t_1, t_2 \in \mathbf{R}$ si ha

$$f(t_1 v_1 + t_2 v_2) = t_1 f(v_1) + t_2 f(v_2).$$

6.4.2. **Proposizione.** *Una funzione lineare è determinata completamente dai suoi valori su una base, che possono essere scelti arbitrariamente.*

Proof. Sia $v_1, \dots, v_n$ una base di V. Dunque ogni elemento $v \in V$ si scrive in modo unico come

$$v = t_1 v_1 + \cdots + t_n v_n.$$

Allora

$$f(v) = t_1 f(v_1) + \cdots + t_n f(v_n).$$

E viceversa, se fisso n vettori $w_1, \dots, w_n \in W$, allora la funzione definita da

$$f(t_1 v_1 + \cdots + t_n v_n) = t_1 w_1 + \cdots + t_n w_n$$

è lineare, con $f(v_i) = w_i$. □

Sia ora $w_1, \dots, w_m$ una base di W, e scriviamo ogni $f(v_j)$ come combinazione lineare dei w_i, dunque

$$f(v_j) = a_{1j} w_1 + a_{2j} w_2 + \cdots + a_{mj} w_m,$$

ovvero

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i. \tag{6.1}$$

Consideriamo la matrice $A = [a_{ij}]$, che dunque *determina completamente la funzione lineare.* Allora

$$f(v) = \sum_{j=1}^{n} t_j f(v_j) = \sum_{j=1}^{n} t_j \left(\sum_{i=1}^{m} a_{ij} w_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} t_j \right) w_j.$$

In altre parole, se le coordinate di v rispetto alla base v_j sono date dal vettore colonna

$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix},$$

allora le coordinate di $f(v)$ rispetto alla base w_i sono date dal vettore colonna

$$\begin{bmatrix} \sum_{j=1}^{n} a_{1j} t_j \\ \sum_{j=1}^{n} a_{2j} t_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} t_j \end{bmatrix} = \begin{bmatrix} a_{11}t_1 + a_{12}t_2 + \cdots + a_{1n}t_n \\ a_{21}t_1 + a_{22}t_2 + \cdots + a_{2n}t_n \\ \vdots \\ a_{m1}t_1 + a_{m2}t_2 + \cdots + a_{mn}t_n \end{bmatrix} =$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}.$$

La matrice

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

si dice *matrice associata* alla funzione lineare f. Dunque una funzione lineare da $\mathbf{R}^n$ a $\mathbf{R}^m$ si può rappresentare mediante una matrice $m \times n$, dunque mediante mn numeri reali. Questa ragione di *economicità della rappresentazione* fa sí che ovunque possibile, ad esempio nella implementazione di funzioni su hardware limitato, tipo una smart card, si cerchi di usare il più possibile funzioni lineari.

Ho ripreso la corrispondenza fra uno spazio vettoriale V di dimensione n, di cui sia fissata una base $v_1, \dots, v_n$, e $\mathbf{R}^n$. Questa corrispondenza è data dalla funzione lineare biiettiva che associa a $v \in V$ quell'unica n-pla $(a_1, \dots a_n)$ tale che $v = a_1 v_1 + \cdots + a_n v_n$.

Ho rivisto come, data una funzione lineare $f : V \to W$, ove V è uno spazio vettoriale di dimensione n, e W è uno spazio vettoriale di dimensione m, e fissate basi di V e W, posso rappresentare f mediante una matrice $m \times n$.

Ho visto come le operazioni elementari su una matrice si possono rappresentare mediante la moltiplicazione a sinistra per una matrice invertibile.

Per esempio $D_1(t)$ si rappresenta mediante la moltiplicazione DA, ove

$$D = \begin{bmatrix} t & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

e notate che se $t \neq 0$, allora D è invertibile, con inversa

$$D^{-1} = \begin{bmatrix} t^{-1} & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Allo stesso modo ad esempio $E_{12}(t)$ si rappresenta mediante EA, ove

$$E = \begin{bmatrix} 1 & t & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

e si ha

$$E^{-1} = \begin{bmatrix} 1 & -t & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Infine S_{12} si rappresenta mediante PA, ove

$$P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

e si ha

$$P^{-1} = P.$$

Ho fatto vedere come da questo segua che se parto da una matrice $m \times n$ A, e considero la sua forma a gradini S, o anche la sua RREF T, allora il sottospazio di $\mathbf{R}^n$ generato dalle righe di A sia lo stesso di quello generato dalle righe di S, o dalle righe di T.

Proseguendo con la notazione appena introdotta, il *rango* di una matrice A è il numero di pivot nella sua forma a gradini S, dunque il numero di righe non nulle di S o anche di T. Dunque è la dimensione dello spazio generato dalle righe di A, S, o T.

Si può vedere che questo rango è anche la dimensione dello spazio generato dalle colonne di A. (Il primo rango si chiama *rango per righe*, il secondo *rango per colonne.*)

6.4.3. **Teorema.** *Sia A una matrice $m \times n$.*
Sono equali

- *la dimensione dello spazio generato dalle righe di A, e*
- *la dimensione dello spazio generato dalle colonne di A.*

Ora, data una funzione lineare $f : V \to W$, rappresentiamo come la funzione $\mathbf{R}^n \to \mathbf{R}^m$ data da $Y = AX$, ove X è un vettore colonna $n \times 1$, Y è un vettore colonna $m \times 1$, e A è una matrice $m \times n$. Si dice *immagine* di f l'insieme

$$I(f) = \{ f(v) : v \in V \} = \{ AX : X \in \mathbf{R}^n \} \subseteq \mathbf{R}^m,$$

e *nucleo* di f l'insieme

$$N(f) = \{ v \in V : f(v) = 0 \} = \{ X \in \mathbf{R}^n : AX = O \} \subseteq \mathbf{R}^n.$$

Si ha (poi aggiungo i dettagli della dimostrazione)

6.4.4. **Teorema** (**Nullità più rango:** il nome si riferisce all'ultimo punto)**.**

(1) *$I(f)$ è un sottospazio di W, ovvero di $\mathbf{R}^m$.*
(2) *$N(f)$ è un sottospazio di V, ovvero di $\mathbf{R}^n$.*
(3) *f è iniettiva se e solo se $N(f) = \{ O \}$.*
(4) *La dimensione di $I(f)$ è eguale al rango di A.*
(5) *La dimensione di $N(f)$ è la dimensione dello spazio delle soluzioni del sistema omogeneo $AX = O$, dunque è eguale a $n - \operatorname{rango}(f)$.*
(6) $\dim(I(f)) + \dim(N(f)) = n$.

Proof.

(1) Siano $x, y \in I(f)$, dunque $x = f(u)$ e $y = f(v)$ per certi $u, v \in V$. Siano $s, t \in \mathbf{R}$. Allora $sx + ty = sf(u) + tf(v) = f(su + tv) \in I(f)$.
(2) $N(f) = \{ X \in \mathbf{R}^n : AX = O \}$ è l'insieme delle soluzioni di un sistema omogeneo, e abbiamo visto che questo è un sottospazio.
(3) Se f è iniettiva, allora $f(v) = O = f(O)$ implica $v = O$, dunque $N(f) = \{ O \}$. Viceversa, se $N(f) = \{ O \}$ e $f(u) = f(v)$, allora $O = f(u) - f(v) = f(u - v)$ e dunque $u - v = O$, ovvero $u = v$.
(4) $I(f)$ è dato dall'insieme

$$AX = A \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{bmatrix} + \dots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{bmatrix}$$

delle combinazioni lineari delle colonne di A. Dunque $I(f)$ è lo spazio generato dalle colonne di A, e quindi la sua dimensione è il rango di A.
(5) In effetti $N(f) = \{ X : AX = O \}$, e sappiamo che la dimensione di questo spazio è la nullità $n - \operatorname{rango}(f)$.
(6) Segue subito dai due punti precedenti.

□

Si può vedere che la composizione di funzioni lineari corrisponde al prodotto fra matrici. Dunque vale la seguente

6.4.5. **Proposizione.** *Siano* $f : \mathbf{R}^n \to \mathbf{R}^m$ *e* $g : \mathbf{R}^k \to \mathbf{R}^n$ *funzioni lineari.*

Sia A *la matrice* $m \times n$ *associata a* f*, e* B *la matrice* $n \times k$ *associata a* g*.*

Allora la funzione composta $f \circ g : \mathbf{R}^k \to \mathbf{R}^m$ *è lineare, e la matrice associata è la matrice prodotto* AB*, che è una matrice* $m \times k$*.*

Qui la funzione $f \circ g$ è la funzione che su $v \in \mathbf{R}^k$ vale $f(g(v))$. In effetti $g(v) \in \mathbf{R}^n$, e dunque a $g(v)$ posso applicare f.

Se vi dessi la scelta fra calcolare la decima potenza di una delle seguenti due matrici, quale sceglieresti?

$$B = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix},$$

Risposta: B. Questo perché

$$B^{10} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}^{10} = \begin{bmatrix} (-1)^{10} & 0 \\ 0 & 2^{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1024 \end{bmatrix}.$$

la possibilità di fare il calcolo in questo modo dipende dal fatto che

$$Be_1 = -e_1, \qquad Be_2 = 2e_2,$$

e quindi

$$B^2 e_1 = B(Be_1) = B(-e_1) = (-1)^2 e_1, \text{ecc.}$$

Qui

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Ci chiediamo se esistono v_1, v_2 tali che

$$Cv_1 = -v_1, \qquad Cv_2 = 2v_2.$$

Proviamo. Se $v_1 = \begin{bmatrix} x \\ y \end{bmatrix}$, allora ho da risolvere

$$\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = - \begin{bmatrix} x \\ y \end{bmatrix},$$

che è il solito sistema lineare omogeneo

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

che ha soluzioni $x = -2t, y = t$. Prendo ad esempio $t = 1$, e dunque $v_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

Faccio la stessa cosa con v_2, e trovo

$$\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix},$$

che è il sistema lineare omogeneo

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

che ha soluzioni $x = t, y = t$. Prendo ad esempio $t = 1$, e dunque $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Ora rispetto alla base v_1, v_2 la funzione lineare data da C ha matrice B! Questo si può vedere anche nel modo seguente, che sarà chiarito fra poco, ma sarebbe comunque già contenuto nella teoria dei cambi di base vista prima. Scriviamo la matrice S che ha per colonne v_1, v_2 (scritti, ma è implicito, rispetto alla base e_1, e_2).

$$S = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Questa matrice ha un'inversa S^{-1}, che è data (poi vediamo perché, ma qui basta verificare che $SS^{-1} = S^{-1}S = I$, ove I è la matrice identica 2×2) da

$$S^{-1} = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Adesso calcoliamo e vediamo che

$$S^{-1}CS = S^{-1} \cdot \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 2 & 2 \\ -1 & 2 \end{bmatrix} =$$

$$= \begin{bmatrix} -\frac{3}{3} & \frac{-2+2}{3} \\ \frac{2-2}{3} & \frac{4+2}{3} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} = B.$$

Dunque $C = SBS^{-1}$, da cui

$$C^2 = SBS^{-1}SBS^{-1} = SBIBS^{-1} = SB^2S^{-1},$$

e da qui

$$C^{10} = SB^{10}S^{-1} = \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1024 \end{bmatrix} \cdot S^{-1} =$$

$$= \begin{bmatrix} -2 & 1024 \\ 1 & 1024 \end{bmatrix} \begin{bmatrix} -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1026}{3} & \frac{2046}{3} \\ \frac{1023}{3} & \frac{2049}{3} \end{bmatrix} = \begin{bmatrix} 342 & 682 \\ 341 & 683 \end{bmatrix}.$$

6.5. **Determinante.** Consideriamo l'insieme $\mathcal{M} = \mathcal{M}_n$ delle matrici quadrate $n \times n$, per un n fissato. Esiste una funzione da $\mathcal{M}$ a $\mathbf{R}$, detta *determinante* e indicata con $\det(A)$, per $A \in \mathcal{M}$, che gode di alcune utili proprietà. Forse le più importanti sono le seguenti.

6.5.1. **Teorema.** *Let $A, B \in \mathcal{M}_n$.*

1. $\det(AB) = \det(A)\det(B)$, *and* $\det(I) = 1$.
2. *A has an inverse A^{-1} iff* $\det(A) \neq 0$.
3. *The system $AX = b$ has a unique solution iff* $\det(A) \neq 0$, *and in this case the solution is $X = A^{-1}b$.*

Proof. Dimostreremo solo alcune di queste affermazioni nel seguito. Per ora notiamo che dal metodo di calcolo sotto riportato segue subito che $\det(I) = 1$. Dunque se A è invertibile, cioè esiste B tale che $AB = BA = I$, allora $1 = \det(I) = \det(A)\det(B)$, e dunque $\det(A) \neq 0$.

Inoltre, se A è invertibile, allora da $Ax = b$ si ottiene moltiplicando a sinistra per A^{-1} che $A^{-1}b = A^{-1}Ax = Ix = x$, e dunque la soluzione è unica. □

Cominciamo a definire un metodo di calcolo per il determinante. Riduciamo a scalini la matrice A. Se ha n pivot a_1, a_2, a_n, e nel corso della riduzione ho fatto s scambi di righe, allora il determinante è

$$(-1)^s a_1 \cdot a_2 \cdot \dots \cdot a_n,$$

altrimenti è zero.

Come immediata conseguenza, abbiamo che il determinante di una matrice diagonale

$$\begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & a_3 & \dots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & a_n \end{bmatrix}$$

è il prodotto $a_1 a_2 \cdots a_n$, in particolare $\det(I) = 1$. Più in generale, si può vedere che il determinante di una matrice *triangolare*

$$\begin{bmatrix} a_1 & b_{12} & b_{13} & \dots & b_{1n} \\ 0 & a_2 & b_{23} & \dots & b_{2n} \\ 0 & 0 & a_3 & \dots & b_{3n} \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & a_n \end{bmatrix}$$

è il prodotto $a_1 a_2 \cdots a_n$.

Dunque se abbiamo una matrice $A = [a]$ che sia 1×1, si ha chiaramente $\det(A) = a$.

Consideriamo il caso di una matrice 2×2, sia

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Supponiamo senza perdita di generalità che sia $a \neq 0$, allora il primo passo nella riduzione sarà

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{E_{21}(-a^{-1}c)} \begin{bmatrix} a & b \\ 0 & d - a^{-1}cb \end{bmatrix},$$

e dunque il determinante sarà

$$\det(A) = a(d - a^{-1}cb) = ad - bc.$$

Si può vedere che vale il seguente metodo di calcolo generale per una matrice $n \times n$ (con $n > 1$) $A = [a_{ij}]$.

6.5.2. **Proposizione** (Sviluppo di un determinante rispetto a una riga)**.** *Fissiamo un indice i. Allora*
$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$
ove A_{ij} è la matrice che si ottiene da A rimuovendo la i-sima riga e la j-sima colonna.

Ad esempio per una matrice 3×3 ho, scegliendo $i = 1$,
$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} =$$
$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) =$$
$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32},$$
da spiegare a voce (magari faccio una figura per spiegarla).

6.6. **Come si calcola la matrice inversa.** Si debba calcolare la matrice inversa di A, una matrice $n \times n$ che abbia $\det(A) \neq 0$. Si costruisce la matrice ausiliaria $n \times 2n$ a blocchi
$$\begin{bmatrix} A & | & I \end{bmatrix},$$
ove I è la matrice identica $n \times n$. Si applica l'algoritmo per portare la matrice in forma RREF (questo è possibile perché $\det(A) \neq 0$, ovvero nessun pivot è zero)
$$\begin{bmatrix} I & | & B \end{bmatrix}.$$
Allora sarà $B = A^{-1}$. Facciamo un esempio
$$A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$
Allora
$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & -1 & 0 & 1 \end{bmatrix} \xrightarrow{E_{21}(-3)} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & -7 & -3 & 1 \end{bmatrix} \xrightarrow{D_2(-1/7)}$$
$$\xrightarrow{D_2(-1/7)} \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 3/7 & -1/7 \end{bmatrix} \xrightarrow{E_{12}(-2)} \begin{bmatrix} 1 & 0 & 1/7 & 2/7 \\ 0 & 1 & 3/7 & -1/7 \end{bmatrix}$$
e in effetti
$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1/7 & 2/7 \\ 3/7 & -1/7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
La ragione per cui tutto questo funziona è che le operazioni del nostro algoritmo su una matrice A si possono rappresentare, come visto più sopra, mediante la moltiplicazione di A a sinistra per una matrice. Allora se parto da
$$\begin{bmatrix} A & | & I \end{bmatrix},$$
e la porto in forma RREF, avrò
$$\begin{bmatrix} I & | & B \end{bmatrix} = F \begin{bmatrix} A & | & I \end{bmatrix},$$

per una certa matrice invertibile F. Ora per matrici *a blocchi* si può fare il prodotto per blocchi, dunque

$$F\left[A \mid I\right] = \left[FA \mid F\right],$$

da cui

$$I = FA, \qquad B = FI = F,$$

e quindi

$$BA = I,$$

ovvero $B = A^{-1}$.

6.7. **L'inversa di una matrice due per due.** Supponiamo $a \neq 0$, e denotiamo $\Delta = ad - bc = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) \neq 0$.

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \xrightarrow{E_{21}(-a^1 c)} \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & a^{-1}\Delta & -a^{-1}c & 1 \end{array}\right] \xrightarrow{D_1(a^{-1})}$$
$$\xrightarrow{D_1(a^{-1})} \left[\begin{array}{cc|cc} 1 & a^{-1}b & a^{-1} & 0 \\ 0 & a^{-1}\Delta & -a^{-1}c & 1 \end{array}\right] \xrightarrow{D_2(a\Delta^{-1})}$$
$$\xrightarrow{D_2(a\Delta^{-1})} \left[\begin{array}{cc|cc} 1 & a^{-1}b & a^{-1} & 0 \\ 0 & 1 & -c\Delta^{-1} & a\Delta^{-1} \end{array}\right] \xrightarrow{E_{12}(-a^{-1}b)}$$
$$\xrightarrow{E_{12}(-a^{-1}b)} \left[\begin{array}{cc|cc} 1 & 0 & a^{-1} + (a^{-1}bc)\Delta^{-1} & -b\Delta^{-1} \\ 0 & 1 & -c\Delta^{-1} & a\Delta^{-1} \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 0 & d\Delta^{-1} & -b\Delta^{-1} \\ 0 & 1 & -c\Delta^{-1} & a\Delta^{-1} \end{array}\right],$$

where we have used

$$a^{-1} + (a^{-1}bc)\Delta^{-1} = a^{-1}\Delta^{-1}\Delta + a^{-1}bc\Delta^{-1} = a^{-1}\Delta^{-1}(\Delta + bc) =$$
$$= a^{-1}\Delta^{-1}(ad - bc + bc) = a^{-1}\Delta^{-1}ad = d\Delta^{-1}.$$

Dunque

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

La matrice 3×3. L'esercizio 9.1, che potrei cercare di scrivere.

Un risultato analogo alla Proposizione 6.5.2 vale scambiando righe e colonne:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}),$$

ove j è un indice fissato.

Si può vedere che il determinante gode di alcune proprietà, che seguono dal fatto che $\det(AB) = \det(A)\det(B)$, e dall'altro fatto che le operazioni descritte si realizzano moltiplicando a destra per una matrice opportuna.

6.7.1. **Proposizione.** *Sia A una matrice $n \times n$.*

(1) *Se B è la matrice ottenuta da A scambiando due righe, o due colonne, allora* $\det(B) = -\det(A)$.
(2) *Se B è la matrice ottenuta da A moltiplicando una riga (o una colonna) per t, allora* $\det(B) = t\det(A)$.

(3) *Segue dal punto precedente che se si moltiplica* ogni riga *per t, allora* $\det(tA) = t^n \det(A)$.
(4) *Se B è la matrice ottenuta da A aggiungendo alla riga i-sima di A un multiplo della riga j-sima, per $j \neq i$, allora* $\det(B) = \det(A)$.

Fra le conseguenze ho

- Se una matrice A ha due righe eguali, allora $\det(A) = 0$. Si usa il punto 1 della Proposizione 6.7.1: infatti scambiando fra loro le due righe la matrice non cambia, mentre il determinante dovrebbe cambiare segno, allora $\det(A) = -\det(A)$, dunque $\det(A) = 0$.
- Se una matrice ha una riga nulla, allora $\det(A) = 0$. Si usa il punto 2 della Proposizione 6.7.1: se la prima riga, per dire, è nulla, allora A si ottiene da A stessa moltiplicando la prima riga per $t = 0$. Ne segue $\det(A) = 0 \cdot \det(A) = 0$.

Questi due fatti si possono anche ottenere considerando la riduzione a gradini. Notate che i due fatti sono quasi la stessa cosa, infatti se ho una matrice con due righe eguali, sottraendo una riga dall'altra ottengo una riga nulla. E se una riga è nulla, aggiungendoci un'altra riga ottengo due righe eguali.

6.8. **Un importante caso particolare.** Quando abbiamo una funzione lineare f da uno spazio V allo stesso spazio V, è importante vedere come cambia la matrice associata, in termini di una singola base che scelgo per V.

Supponiamo che rispetto a una certa base $v_1, \dots, v_n$ la funzione lineare f venga rappresentata dalla matrice (quadrata, $n \times n$) A. Dunque se

$$\mathcal{V} = [v_1, \dots, v_n]$$

è questa base, e

$$\mathcal{W} = [w_1, \dots, w_n]$$

è un'altra base, vogliamo confrontare le matrici A, B di f, rispetto alle basi $\mathcal{V}, \mathcal{W}$. Se X sono le coordinate di un vettore rispetto a $\mathcal{V}$, e X' rispetto a $\mathcal{W}$, abbiamo visto che si ha $X' = SX$, ove $\mathcal{V} = \mathcal{W}S$. Qui dunque S è la matrice del cambio di base.

Dunque se $Y = AX$, si ha

$$Y' = SY = SAX = SAS^{-1}X',$$

ovvero $B = SAS^{-1}$.

Abbiamo visto al dimostrazione di entrambi i versi del punto 3 del Teorema 6.5.1, che magari cerco di scrivere.

Sia V uno spazio vettoriale di dimensione n, e $f : V \to V$ una funzione lineare. Supponiamo che la matrice A di f rispetto alla base $\mathcal{B} = [b_1, b_2, \dots, b_n]$ sia diagonale

$$A = \begin{bmatrix} a_1 & 0 & 0 & \dots & 0 \\ 0 & a_2 & 0 & \dots & 0 \\ 0 & 0 & a_3 & \dots & 0 \\ & & & \ddots & \\ 0 & & 0 & \dots & a_n \end{bmatrix}.$$

Questo significa che l'immagine $f(b_i)$ di b_i è la combinazione lineare dei b_i con i coefficienti della i-sima colonna di A:

$$f(b_i) = 0b_1 + 0b_2 + \cdots + a_i b_i + \cdots + 0b_n = a_i b_n.$$

Ne deriva la seguente

6.8.1. **Definizione** (Autovalori e autovettori)**.** Sia V uno spazio vettoriale di dimensione n, e $f : V \to V$ una funzione lineare.

Sia $O \neq v \in V$.

Si dice che v è un *autovettore* di f se esiste $a \in \mathbf{R}$ tale che

$$f(v) = av.$$

In tal caso si dice che a è un *autovalore* di f rispetto all'autovettore v.

Un paio di commenti.

- Se v è un autovettore, ogni suo multiplo non nullo è ancora un autovettore. Infatti se $f(v) = av$, allora $f(tv) = tf(v) = tav = a(tv)$.
- E' essenziale richiedere che $v \neq O$, Infatti $f(O) = O = a \cdot O$ per ogni $a \in \mathbf{R}$, dunque se ammettessi l'autovettore nullo, allora ogni numero reale sarebbe un autovalore.
- L'autovalore è automaticamente determinato dall'autovettore $v \neq O$. Infatti se $f(v) = av = bv$, con $a, b \in \mathbf{R}$, allora $(a - b)v = 0$, e dunque $a - b = 0$, dato che $v \neq O$.

Supponiamo che la matrice $n \times n$ A abbia un autovalore a rispetto all'autovettore $v \neq O$, dunque

$$Av = av.$$

Ne segue

$$O = Av - av = Av - aIv = (A - aI)v.$$

Dunque l'equazione

$$(A - aI)v = O$$

ha una soluzione $v \neq O$, dunque *non* ha un'unica soluzione, dunque per il punto 3 del Teorema 6.5.1 si ha $\det(A - aI) = O$. Abbiamo ottenuto una parte di

6.8.2. **Teorema** (Polinomio caratteristico)**.** *Sia A una matrice $n \times n$.*

Gli autovalori di A sono le radici del polinomio di grado n

$$\det(A - \lambda I) \in \mathbf{R}[\lambda],$$

che viene detto polinomio caratteristico di A.

Proof. In effetti ci si potrebbe convincere facilmente che $f(\lambda) = \det(A - \lambda I)$ è un polinomio in λ di grado n, e abbiamo appena visto che gli autovalori sono radici di questo polinomio.

Viceversa, se a è una radice di $f(\lambda)$, allora $\det(A - aI) = O$, e dunque per il punto 3 del Teorema 6.5.1 esiste $O \neq v \in V$ tale che $(A - aI)v = O$, e dunque a è un autovalore rispetto all'autovettore v. □

Notate che sembrerebbe che il polinomio caratteristico dipenda dalla matrice, e non dalla funzione lineare. Ma se cambio base, e la matrice A diventa SAS^{-1}, allora

$$\begin{aligned}
\det(SAS^{-1} - \lambda I) &= \det(SAS^{-1} - S\lambda I S^{-1}) \\
&= \det(S(A - \lambda I)S^{-1}) \\
&= \det(S)\det(A - \lambda I)\det(S^{-1}) \\
&= \det(S)\det(S)^{-1}\det(A - \lambda I) \\
&= \det(A - \lambda I).
\end{aligned}$$

Per esempio, consideriamo la matrice

$$A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$$

già vista in precedenza. Il suo polinomio caratteristico è

$$\begin{aligned}
\det(A - \lambda I) &= \det\left(\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} -\lambda & 2 \\ 1 & 1-\lambda \end{bmatrix}\right) \\
&= -\lambda(1-\lambda) - 2 \\
&= \lambda^2 - \lambda - 2 \\
&= (\lambda - 2)(\lambda + 1),
\end{aligned}$$

dunque gli autovalori sono $-1, 2$, e notate che eravamo riusciti a mettere A in forma diagonale

$$\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}.$$

6.8.3. **Definizione.** Una matrice si dice *diagonalizzabile* se esiste una base fatta interamente di autovettori.

Ci sono almeno due problemi. Consideriamo la matrice

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

Si vede subito che il suo polinomio caratteristico è $\lambda^2 + 1$, che non ha radici reali, le due radici sono i numeri complessi ± 1. Ricordiamo

6.8.4. **Teorema** (Teorema fondamentale dell'algebra)**.** *Sia*

$$f(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0 \in \mathbf{C}[\lambda]$$

un polinomio a coefficienti complessi di grado n. Allora f ha tutte le sue radici in $\mathbf{C}$, nel senso che esistono $z_1, z_2, \ldots, z_n \in \mathbf{C}$ tali che

$$f(\lambda) = (\lambda - z_1)(\lambda - z_2)\cdots(\lambda - z_n).$$

Qui vale la pena di ricordare

6.8.5. **Lemma** (Regola di Ruffini)**.** *Sia $f(\lambda)$ un polinomio, reale o complesso, e z un numero reale o complesso. Sono equivalenti*

- *z è una radice di f, ovvero $f(z) = 0$, e*
- *$\lambda - z$ divide $f(\lambda)$.*

Il secondo problema si vede considerando la matrice

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Il suo polinomio caratteristico è $\lambda^2 - 2\lambda + 1 = (\lambda - 1)^2$, dunque le sue radici sono $1, 1$. Ma non c'è una base fatta di autovettori (dunque A non è diagonalizzabile), perché se

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix},$$

allora

$$\begin{cases} x + y = 0 \\ \qquad y = 0 \end{cases},$$

da cui segue che gli autovettori sono $x \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ che non bastano per fare una base.

Vediamo qualcosa in positivo.

6.8.6. **Lemma.** *Autovettori relativi ad autovalori distinti sono linearmente indipendenti.*

Proof. Siano

$$v_1, v_2, \dots, v_k$$

autovettori relativi agli autovalori

$$a_1, a_2, \dots, a_k.$$

Supponiamo esistano $x_1, x_2, \dots, x_n$ tali che

$$x_1 v_1 + x_2 v_2 + \dots + x_n v_n = O, \tag{6.2}$$

e dunque devo vedere che tutti gli x_i sono zero. Supponiamo il contrario, cioè che non tutti gli x_i siano zero.

Posso supporre che 6.2 sia la relazione *più breve* che esiste fra i v_i, nel senso che se esiste una relazione

$$x_1 v_1 + x_2 v_2 + \dots + x_k v_k = O,$$

con $k < n$, allora tutti gli x_i sono nulli. (Altrimenti si può procedere per induzione, se sapete cos'è.)

Applico A ad (6.2), ottenendo

$$x_1 a_1 v_1 + x_2 a_2 v_2 + \dots + x_n a_n v_n = O. \tag{6.3}$$

Ma da (6.2) ottengo

$$x_n v_n = -x_1 v_1 - x_2 v_2 - \dots - x_{n-1} v_{n-1},$$

e sostituendo in (6.3) ottengo

$$x_1(a_1 - a_n) + x_2(a_2 - a_n) + \cdots + x_{n-1}(a_{n-1} - a_n) = O. \tag{6.4}$$

Ora notate che in (6.2) non può essere $x_1 = \cdots = x_{n-1} = 0$, altrimenti $x_n v_n = 0$, e dunque anche $x_n = 0$, dato che $v_n \neq O$. Inoltre $a_1 - a_n \neq 0$ per $i < n$. Dunque (6.4) sarebbe una relazione più breve, il che è una contraddizione. □

6.8.7. **Proposizione.** *Se il polinomio caratteristico ha radici distinte, allora la matrice è diagonalizzabile.*

Il viceversa non vale, nel senso che le matrici

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

hanno entrambe come polinomio caratteristico $(\lambda - 1)^2$, che dunque *non* ha radici distinte, ma la prima è diagona(lizzabil)e, mentre abbiamo visto che la seconda non lo è.

Vediamo alcuni esempi un pochino più complessi.

Consideriamo dapprima la matrice

$$A = \begin{bmatrix} 5/2 & -1/2 & -1 \\ 1/2 & 3/2 & -1 \\ -1/2 & -1/2 & 2 \end{bmatrix}$$

Calcoliamo il polinomio caratteristico

$$\begin{aligned}
\det \begin{bmatrix} 5/2 - x & -1/2 & -1 \\ 1/2 & 3/2 - x & -1 \\ -1/2 & -1/2 & 2 - x \end{bmatrix} &= (\frac{5}{2} - x)(\frac{3}{2} - x)(2 - x) - \frac{1}{4} + \frac{1}{4} \\
&\quad - \frac{1}{2}(\frac{3}{2} - x) - \frac{1}{2}(\frac{5}{2} - x) + \frac{1}{4}(2 - x) \\
&= (\frac{15}{4} - 4x + x^2)(\frac{3}{2} - x)(2 - x) \\
&\quad - \frac{1}{2}(\frac{3}{2} - x) - \frac{1}{2}(\frac{5}{2} - x) + \frac{1}{4}(2 - x) \\
&= \frac{15}{2} - \frac{15}{4}x - 8x + 4x^2 + 2x^2 - x^3 \\
&\quad - \frac{3}{4} + \frac{1}{2}x - \frac{5}{4} + \frac{1}{2}x + \frac{1}{2} - \frac{1}{4}x \\
&= -x^3 + 6x^2 - 11x + 6.
\end{aligned}$$

Ora usiamo senza dimostrazione il facile

6.8.8. **Teorema.** *Sia*

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

un polinomio a coefficienti interi.

Se $\alpha \in \mathbf{Z}$ è una radice di $f(x)$, allora α divide a_0.

Dunque se ci sono radici intere, sono i divisori di 6. Proviamo con 1, e vediamo che è radice, e poi dividendo

$$-x^3 + 6x^2 - 11x + 6 = -(x-1)(x^2 - 5x + 6),$$

per cui le radici sono $1, 2, 3$. Qui abbiamo usato la Regola di Ruffini.

Dato che le radici del polinomio caratteristico sono distinte, gli autovettori relativi sono indipendenti, e dunque la matrice è diagonalizzabile. Risolviamo dunque $AX = aX$ per $a = 1, 2, 3$. Vediamo che

$$AX = \begin{bmatrix} 5/2 & -1/2 & -1 \\ 1/2 & 3/2 & -1 \\ -1/2 & -1/2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5/2x - 1/2y - z \\ 1/2x + 3/2y - z \\ -1/2x - 1/2y + 2z \end{bmatrix}.$$

Dunque $AX = X$ mi dà il sistema

$$\begin{cases} 3/2x - 1/2y - z = 0 \\ 1/2x + 1/2y - z = 0 \\ -1/2x - 1/2y + z = 0 \end{cases},$$

che ha nullità 1, e ammette soluzione $(1, 1, 1)$.

Poi $AX = 2X$ mi dà il sistema

$$\begin{cases} 1/2x - 1/2y - z = 0 \\ 1/2x - 1/2y - z = 0 \\ -1/2x - 1/2y = 0 \end{cases},$$

che ha nullità 1, e ammette soluzione $(1, -1, 1)$, $AX = 3X$ mi dà il sistema

$$\begin{cases} -1/2x - 1/2y - z = 0 \\ 1/2x - 3/2y - z = 0 \\ -1/2x - 1/2y - z = 0 \end{cases},$$

che ha nullità 1, e ammette soluzione $(1, 1, -1)$. Per diagonalizzarla, considero la matrice

$$S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

che ha per colonne i vettori della base di autovalori, e ottengo

$$S^{-1}AS = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \tag{6.5}$$

dunque l'ho diagonalizzata. Notiamo che

$$S^{-1} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & -1/2 & 0 \\ 1/2 & 0 & -1/2 \end{bmatrix} \tag{6.6}$$

In realtà possiamo risparmiarci i conti di (6.5) e (6.6), ragionando come segue. La matrice S manda il vettore e_i della base canonica nella sua colonna i-sima, che è un autovettore v_i rispetto all'autovalore a_i. (Qui scriviamo a_1, a_2, a_3 per gli autovalori di A, dunque in questo caso particolare si ha $a_1 = 1, a_2 = 2, a_3 = 3$.)

Ora dato che questi v_i sono autovettori, si ha $Av_i = a_i v_i$. Ora S^{-1} fa il contrario di quello che faceva S, dunque manda ogni v_i in e_i. La morale è che

$$S^{-1}ASe_i = S^{-1}Av_i = S^{-1}a_i v_i = a_i e_i,$$

dunque gli e_i sono una base di autovettori per $S^{-1}AS$, dunque quest'ultima matrice è diagonale.

Consideriamo ora la matrice

$$M = \begin{bmatrix} 3/2 & 0 & -1/2 \\ 1/2 & 1 & -1/2 \\ -1/2 & 0 & 3/2 \end{bmatrix}.$$

Calcoliamo il polinomio caratteristico

$$\begin{aligned} \det \begin{bmatrix} 3/2-x & 0 & -1/2 \\ 1/2 & 1-x & -1/2 \\ -1/2 & 0 & 3/2-x \end{bmatrix} &= (\frac{3}{2}-x)(1-x)(\frac{3}{2}-x) - \frac{1}{4}(1-x) \\ &= (\frac{3}{2} - \frac{3}{2}x - x + x^2)(\frac{3}{2}-x) - \frac{1}{4}(1-x) \\ &= (\frac{3}{2} - \frac{5}{2}x + x^2)(\frac{3}{2}-x) - \frac{1}{4}(1-x) \\ &= \frac{9}{4} - \frac{3}{2}x - \frac{15}{4}x + \frac{5}{2}x^2 + \frac{3}{2}x^2 - x^3 - \frac{1}{4} + \frac{1}{4}x \\ &= -x^3 + 4x^2 - 5x + 2. \end{aligned}$$

Allora le radici di $f(x) = -x^3 + 4x^2 - 5x + 2 = 0$ sono da ricercare fra i divisori di -2, dunque $1, -1, 2, -2$. Proviamo con $\alpha = 1$, e in effetti $-1+4-5+2 = 0$. Allora dividiamo $f(x)$ per $x-1$, ottenendo

$$f(x) = (x-1)(-x^2+3x-2).$$

Ora le radici di $-x^2+3x-2$ si vedono subito essere $1, 2$, dunque

$$f(x) = -(x-1)(x-1)(x-2),$$

e dunque la radici di $f(x)$ sono $1, 1, 2$. Dato che non sono distinte, non siamo sicuri che la matrice sia diagonalizzabile.

Cerchiamo quindi gli autovettori relativi all'autovalore 1 risolvendo

$$MX = X,$$

cioè

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

ovvero

$$\begin{cases} \frac{1}{2}x & -\frac{1}{2}z & = & 0 \\ \frac{1}{2}x & -\frac{1}{2}z & = & 0 \\ -\frac{1}{2}x & +\frac{1}{2}z & = & 0 \end{cases},$$

che si vede subito dire $x = z$, ovvero il sistema equivale a

$$\begin{cases} x & -z & = & 0 \end{cases}.$$

La nullità è 2, dunque lo spazio delle soluzioni è quindi di dimensione 2 dato da

$$s(1,0,1)+t(0,1,0)=(s,t,s).$$

Cerchiamo ora gli autovettori relativi all'autovalore 2 risolvendo

$$MX=2X,$$

cioè

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

ovvero

$$\begin{cases} -\frac{1}{2}x & & -\frac{1}{2}z & = & 0 \\ \frac{1}{2}x & -y & -\frac{1}{2}z & = & 0 \\ -\frac{1}{2}x & & -\frac{1}{2}z & = & 0 \end{cases},$$

che si vede subito equivalere a

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

La nullità è 1, dunque lo spazio delle soluzioni è quindi di dimensione 1 dato da

$$s(-1,-1,1)=(-s,-s,s).$$

Abbiamo trovato una base

$$(1,0,1),(0,1,0),(-1,-1,1)$$

di autovalori, dunque la matrice è diagonalizzabile. Per diagonalizzarla, considero la matrice

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

che ha per colonne i vettori della base di autovalori, e ottengo

$$S^{-1}MS=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 2\end{bmatrix},$$

dunque l'ho diagonalizzata. Notiamo che

$$S^{-1}=\begin{bmatrix}1/2 & 0 & 1/2\\ -1/2 & 1 & 1/2\\ -1/2 & 0 & 1/2\end{bmatrix}$$

Anche qui si può evitare l'ultimo calcolo con il ragionamento fatto nell'esempio precedente.

Consideriamo ora

$$N=\begin{bmatrix}2 & -1/2 & -1/2\\ 1 & 1/2 & -1/2\\ 0 & -1/2 & 3/2\end{bmatrix}.$$

Calcoliamo il polinomio caratteristico

$$\begin{aligned}\det\begin{bmatrix}2-x & -1/2 & -1/2\\ 1 & 1/2-x & -1/2\\ 0 & -1/2 & 3/2-x\end{bmatrix} &= (2-x)(\frac{1}{2}-x)(\frac{3}{2}-x)+\frac{1}{4}\\ &\quad -\frac{1}{4}(2-x)+\frac{1}{2}(\frac{3}{2}-x)\\ &= (1-\frac{5}{2}x+x^2)(\frac{3}{2}-x)+\frac{1}{4}\\ &\quad -\frac{1}{4}(2-x)+\frac{1}{2}(\frac{3}{2}-x)\\ &= \frac{3}{2}-x-\frac{15}{4}x+\frac{5}{2}x^2+\frac{3}{2}x^2-x^3+\frac{1}{4}\\ &\quad -\frac{1}{2}+\frac{1}{4}x+\frac{3}{4}-\frac{1}{2}x\\ &= -x^3+4x^2-5x+2,\end{aligned}$$

lo stesso polinomio di prima.

Stavolta però quando risolviamo

$$NX = X,$$

troviamo

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

ovvero

$$\begin{cases}x & -\frac{1}{2}y & -\frac{1}{2}z & = & 0\\ x & -\frac{1}{2}y & -\frac{1}{2}z & = & 0\\ & -\frac{1}{2}y & +\frac{1}{2}z & = & 0\end{cases},$$

cioè

$$\begin{cases}x & & -z & = & 0\\ & y & -z & = & 0\end{cases},$$

ma qui la nullità è solo 1, e le soluzioni sono

$$s(1,1,1).$$

Ora consideriamo

$$NX = 2X,$$

cioè

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

ovvero

$$\begin{cases} & -\frac{1}{2}y & -\frac{1}{2}z & = & 0\\ x & -\frac{3}{2}y & -\frac{1}{2}z & = & 0\\ & -\frac{1}{2}y & -\frac{1}{2}z & = & 0\end{cases},$$

cioè

$$\begin{cases} x & & +z & = & 0 \\ & y & +z & = & 0 \end{cases},$$

la nullità è 1, e le soluzioni sono

$$s(-1, -1, 1).$$

Dunque *non* abbiamo una base di autovettori, e la matrice *non* è diagonalizzabile. Quello che si potrebbe vedere è che rispetto a una base opportuna la matrice si può mettere nella forma

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Vale poi questo importante risultato, che contiene un paio di definizioni.

6.8.9. **Teorema.** *Sia V uno spazio vettoriale di dimensione n, e sia $f : V \to V$ una funzione lineare, che supponiamo abbia tutti autovalori reali. Sia $g(x) = \det(A - xI)$ il polinomio caratteristico di f, ove A è la matrice di f rispetto a una base opportuna.*

Se a è un autovalore, definiamo la sua molteplicità algebrica *di a come il più grande k tale che $(x - a)^k$. Definiamo la* molteplicità geometrica *di a come la dimensione del sottospazio degli autovettori relativi ad a, cioè*

$$\dim \{ v \in V : f(v) = av \}.$$

Si ha

- *La molteplicità geometrica di un autovalore è sempre minore o eguale alla sua molteplicità algebrica.*
- *Inoltre sono equivalenti:*
  (1) *f è diagonalizzabile, e*
  (2) *la molteplicità algebrica e la molteplicità geometrica di ogni autovalore coincidono.*

Non staremo a dimostrare questo risultato, ma notiamo che per la matrice M qui sopra la condizione (2) del teorema è soddisfatta, dato che il polinomio caratteristico è $(x - 1)^2(x - 2)$, dunque gli autovalori sono 1 (con molteplicità algebrica 2) e 2 (con molteplicità algebrica 1). Abbiamo visto che

$$\dim \{ v \in V : Mv = v \}$$

ha dimensione 2, e

$$\dim \{ v \in V : Mv = 2v \}.$$

ha dimensione 1, dunque le molteplicità geometriche coincidono con quelle algebriche, e la matrice si può diagonalizzare.

Invece per la matrice N le molteplicità algebriche sono le stesse, ma

$$\dim \{ v \in V : Nv = v \}$$

ha solo dimensione 1, dunque la molteplicità geometrica dell'autovalore 1 è solo 1, cioè diversa (inferiore) a quella algebrica, che è 2, e in effetti la matrice non si è lasciata diagonalizzare.

Data una matrice A, la sua *trasposta* A^T è quella matrice che si ottiene da A scambiando gli elementi rispetto alla diagonale principale. In altre parole, l'elemento di posto (i, j) (i-sima riga, j-sima colonna) di A^T è l'elemento di posto (j, i) di A. Ovvero le colonne di A sono le righe di A^T e viceversa. Esempio:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \\ -2 & 4 & -3 \end{bmatrix}, \qquad A^T = \begin{bmatrix} 1 & 1 & -2 \\ 2 & -1 & 4 \\ 3 & 0 & -3 \end{bmatrix}.$$

Si può vedere che vale $(AB)^T = B^T A^T$. Questo è simile al fatto che

6.8.10. **Proposizione.** *Se A, B sono matrici quadrate invertibili, allora AB è invertibile, e si ha*

$$(AB)^{-1} = B^{-1} A^{-1}.$$

Ricordate che in generale il prodotto di matrici non è commutativo, quindi non scambiate i fattori!

Una matrice quadrata si dice *simmetrica* se $A^T = A$. Vale il seguente

6.8.11. **Teorema** (Teorema spettrale)**.** *Una matrice simmetrica, a coefficienti reali, ha tutti i suoi autovalori reali, ed è diagonalizzabile.*

Vediamo solo un esempio

$$B = \begin{bmatrix} 7 & 1 & 1 \\ 1 & 7 & 1 \\ 1 & 1 & 7 \end{bmatrix}.$$

Calcoliamo il polinomio caratteristico

$$\det \begin{bmatrix} 7-x & 1 & 1 \\ 1 & 7-x & 1 \\ 1 & 1 & 7-x \end{bmatrix} = (7-x)^3 + 2 - 3(7-x) = -x^3 + 21x^2 - 144x + 324.$$

Però qui conviene usare il fatto che il determinante non cambia facendo operazioni elementari del tipo "somma a una riga il multiplo di un'altra colonna". Dunque il determinante è lo stesso di

$$\begin{bmatrix} 7-x & 1 & 1 \\ 1 & 7-x & 1 \\ 1 & 1 & 7-x \end{bmatrix} \xrightarrow{E_{32}(-1)} \begin{bmatrix} 7-x & 1 & 1 \\ 1 & 7-x & 1 \\ 0 & x-6 & 6-x \end{bmatrix}$$

e ora posso sviluppare rispetto all'ultima riga per trovare che il determinante è

$$\begin{aligned} -(x-6)\det \begin{bmatrix} 7-x & 1 \\ 1 & 1 \end{bmatrix} + (6-x)\det \begin{bmatrix} 7-x & 1 \\ 1 & 7-x \end{bmatrix} &= \\ = (6-x)(6-x+(7-x)^2-1) = (6-x)(x^2-15x+54) &= \\ &= -(x-6)^2(x-9). \end{aligned}$$

Notate che gli autovalori sono dunque $6, 6, 9$, non distinti. Ma il fatto che la matrice sia simmetrica mi garantisce di poterla diagonalizzare. In effetti risolvendo

$BX = 6X$ trovo il sistema

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

che chiaramente ha rango 1 e nullità 2, una base dello spazio delle soluzioni è

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \qquad \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Invece da $BX = 9X$ ottengo

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

Col nostro solito algoritmo

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \xrightarrow{S_{13}} \begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ -2 & 1 & 1 \end{bmatrix} \xrightarrow{E_{31}(-2)} \begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 0 & -3 & 3 \end{bmatrix} \xrightarrow{E_{21}(-1)}$$

$$\xrightarrow{E_{21}(-1)} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 3 & -3 \\ 0 & -3 & 3 \end{bmatrix} \xrightarrow{E_{32}(1)} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{D_2(1/3)} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{E_{12}(2)}$$

$$\xrightarrow{E_{12}(2)} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Ne tiriamo fuori una soluzione

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

e quindi abbiamo trovato una base fatta di autovettori.

6.9. **Basi ortonormali.** In diverse situazioni conviene considerare basi di $\mathbf{R}^n$ che siano *ortonormali*.

6.9.1. **Definizione.** Una base $v_1, \dots, v_n$ di $\mathbf{R}^n$ si dice *ortonormale* se si ha

$$v_i \cdot v_j = 0$$

per $i \neq j$, mentre

$$\|v_i\| = 1$$

per ogni i.

In altre parole la base $v_1, \dots, v_n$ di $\mathbf{R}^n$ è ortonormale se

$$v_i \cdot v_j = \begin{cases} 0 & \text{se } i \neq j, \\ 1 & \text{se } i = j. \end{cases}$$

Per esempio la base canonica e_i è ortonormale, ma ad esempio anche la base

$$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \qquad \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

di $\mathbf{R}^2$ lo è.

Una delle ragioni che rende conveniente l'usod elle basi ortonormali, è che per una simile base è particolarmente facile scrivere un vettore come combinaizone lineare. In effetti se $b_1, \dots, b_n$ è una base ortonormale, e $v \in V$, allora per calcolare gli x_i tali che

$$v = x_1 b_1 + \dots + x_n b_n.$$

è sufficiente calcolare i prodotti scalari

$$v \cdot b_i = x_1(b_1 \cdot b_i) + \dots + x_n(b_n \cdot b_i) = x_i(b_i \cdot b_i) = x_i.$$

6.10. **Gram-Schmidt.** Esiste un algoritmo, detto di Gram-Schmidt, che data una base qualsiasi mi permette di costruire una base ortonormale. Questo vuol dire quanto segue. Sia V uno spazio vettoriale di dimensione finita n. Supponiamo che sia definita una funzione

$$\begin{aligned} V \times V &\to \mathbf{R} \\ (x, y) &\mapsto x \cdot y \end{aligned}$$

che soddisfa le proprietà del prodotto scalare. Dunque la funzione è simmetrica

$$x \cdot y = y \cdot x.,$$

è lineare nel primo argomento, cioè

$$(sx + ty) \cdot z = s(x \cdot z) + t(y \cdot z),$$

e quindi per la simmetria è lineare in entrambi gli argomenti, e inoltre

$$\begin{cases} x \cdot x \geq 0, \text{ e} \\ \text{se } x \cdot x = 0, \text{ allora } x = O. \end{cases}$$

Allora l'algoritmo di Gram-Schmidt permette di costruire, a partire da una base $v_1, v_2, \dots, v_n$ qualsiasi di V, una base ortonormale $b_1, b_2, \dots, b_n$.

Se le colonne di una matrice $n \times n$ formano una base ortonormale, allora si vede che la matrice P soddisfa

$$P^T P = I, \quad \text{ovvero} \quad P^T = P^{-1}.$$

Una tale matrice si dice *ortogonale*. Si può vedere, a complemento del Teorema 6.8.11, che se B è una matrice simmetrica, allora esiste una matrice ortogonale P tale che $P^{-1}BP$ è diagonale.

Vediamo l'algoritmo di Gram-Schmidt. Sia $v_1, v_2, \dots, v_n$ una base qualsiasi, e cominciamo a scrivere

$$b_1 = \frac{v_1}{\|v_1\|},$$

sicché per cominciare abbiamo

$$\|b_1\| = \left\| \frac{v_1}{\|v_1\|} \right\| = \frac{\|v_1\|}{\|v_1\|} = 1.$$

Ora cerchiamo un vettore $v_2 - xb_1$ che sia ortogonale a b_1. Deve essere

$$0 = (v_2 - xb_1) \cdot b_1 = v_2 \cdot b_1 - x,$$

dunque si sceglie $x = v_2 \cdot b_1$, e poi si prende

$$b_2 = \frac{v_2 - xb_1}{\|v_2 - xb_1\|}.$$

In generale, se ho già definito $b_1, \dots, b_{k-1}$ in modo che sia $b_i \cdot b_j = 0, 1$ a seconda se $i \neq j$ o $i = j$, allora cerco $x_1, \dots, x_{k-1}$ tali che per ogni $i = 1, \dots, k-1$ valga

$$0 = b_i \cdot (v_k - x_1 b_1 - \dots - x_{k-1} b_{k-1}) = b_i \cdot v_k - x_i,$$

da cui $x_i = b_i \cdot v_k$, e poi come sopra si divide il vettore $v_k - x_1 b_1 - \dots - x_{k-1} b_{k-1}$ per la sua lunghezza per trovare b_k.

6.11. **Ancora sugli spazi vettoriali.**

6.11.1. **Proposizione.** *Sia V uno spazio vettoriale, e sia*

$$v_1, \dots, v_k$$

un sistema di generatori.

Allora una parte dei v_i forma una base di V.

Proof. Se i v_i sono linearmente indipendenti, è fatta, per il Teorema 5.5.3.

Altrimenti i v_i *non* sono linearmente indipendenti, dunque esistono $a_1, \dots, a_k$ *non tutti nulli* tali che

$$a_1 v_1 + \dots + a_k v_k = 0.$$

Cambiando eventualmente l'ordine, possiamo supporre che sia $a_k \neq 0$. Dunque

$$v_k = -a_k^{-1} a_1 v_1 - \dots - -a_k^{-1} a_{k-1} v_{k-1}. \tag{6.7}$$

Ma i v_i erano un sistema di generatori, cioè ogni $v \in V$ si scrive nella forma

$$v = c_1 v_1 + \dots + c_{k-1} v_{k-1} + c_k v_k.$$

Sostituendo (6.7) si ottiene

$$\begin{aligned} v &= c_1 v_1 + \dots + c_{k-1} v_{k-1} + \\ &\quad c_k(-a_k^{-1} a_1 v_1 - \dots - -a_k^{-1} a_{k-1} v_{k-1}) \\ &= (c_1 - a_k^{-1} a_1) v_1 + \dots + (c_{k-1} - a_k^{-1} a_{k-1}) v_{k-1}. \end{aligned}$$

Dunque $v_1, \dots, v_{k-1}$ sono un sistema di generatori. Ripetendo l'argomento (in realtà, argomentando per induzione), si ottiene il risultato. □

6.11.2. **Proposizione.** *Sia V uno spazio vettoriale,*

$$v_1, \dots, v_k$$

linearmente indipendenti, e

$$b_1, \dots, b_n$$

una base.

Allora esistono alcuni b_i che aggiunti a tutti i v_i formano una base.

Proof. Consideriamo i vettori

$$v_1, \dots, v_k, b_1, \dots, b_n.$$

Sono senz'altro un sistema di generatori, perché già i b_i lo sono. Se sono linearmente indipendenti, è fatta, sempre per il Teorema 5.5.3.

Altrimenti esistono x_i, y_i *non tutti nulli* tali che

$$x_1 v_1 + \dots + x_k v_k + y_1 b_1 + \dots + y_n b_n = 0.$$

Ora non tutti gli y_i possono essere zero, perché per ipotesi i v_i sono linearmente indipendenti, e dunque anche gli x_i dovrebbero essere tutti zero.

Allora sarà per esempio $y_n \neq 0$. (In caso, cambio l'ordine ai b_i.) Ma allora, come visto nella dimostrazione della Proposizione 6.11.1, sarà un sistema di generatori anche

$$v_1, \dots, v_k, b_1, \dots, b_{n-1}.$$

Ripetendo l'argomento (cioè procedendo per induzione) si arriva al risultato. □

E ora il Teorema con la dimostrazione forse più difficile di tutto il corso. Premettiamo alcune osservazioni ricorrenti, alcune delle quali già usate.

6.11.3. **Lemma.**

(1) *Se i vettori $v_1, \dots, v_n$ sono linearmente indipendenti, allora sono tutti diversi da O.*
(2) *Se il vettore v è combinazione lineare dei vettori $w_1, \dots, w_n$, allora i vettori $v, w_1, \dots, w_n$ sono linearmente dipendenti.*
(3) *Se i vettori $v_1, \dots, v_n$ sono linearmente indipendenti, allora qualsiasi parte di essi è linearmente indipendente.*
(4) *Se i vettori $v_1, \dots, v_n$ sono un sistema di generatori per lo spazio vettoriale V, allora aggiungendo altri vettori ho sempre un sistema di generatori.*
(5) *Se v_k è combinazione lineare di $v_1, \dots, v_{k-1}$, allora il sottospazio generato da $v_1, \dots, v_k$ è eguale al sottospazio generato da $v_1, \dots, v_{k-1}$.*

Proof.

Punto (1). Se fosse ad esempio $v_1 = O$, allora

$$1 \cdot v_1 + 0 \cdot v_2 + \dots + 0 \cdot v_n = O,$$

e non tutti i coefficienti sono zero (guardate il primo).

Punto (2). Se

$$v = a_1 w_1 + \dots + a_n w_n,$$

allora

$$1 \cdot v - a_1 w_1 - \dots - a_n w_n = O,$$

e non tutti i coefficienti sono zero (guardate il primo). (In effetti questo punto ingloberebbe il primo.)

Punto (3). Basta far vedere che $v_1, \dots, v_{n-1}$ sono linearmente indipendenti. Se

$$0 = a_1 v_1 + \dots + a_{n-1} v_{n-1} = a_1 v_1 + \dots + a_{n-1} v_{n-1} + 0 v_n,$$

allora $a_1 = \dots = a_{n-1} = 0$.

Punto (4). Sia $v \in V$, allora esistono a_i tali che

$$v = a_1 v_1 + \dots + a_n v_n = a_1 v_1 + \dots + a_n v_n + 0 w_1 + \dots + 0 w_m,$$

qualunque siano i w_i.

Punto (5). Lo abbiamo già visto all'interno della Proposizione 6.11.1. □

6.11.4. **Teorema.** *Sia V uno spazio vettoriale che abbia una base con m elementi e una base con n elementi.*

Allora $m = n$.

Ricordate che il numero di elementi di una base si dice *dimensione* dello spazio vettoriale.

Proof. Siano

$$v_1, \dots, v_m$$

e

$$w_1, \dots, w_n$$

le due basi, e supponiamo che sia $m < n$. Vogliamo vedere che questo porta a una contraddizione. (Se fosse $n < m$, poco male, scambiamo i ruoli dei v_i e dei w_i. E se fosse $m = n$, beh, è proprio quello che vogliamo dimostrare.)

Dato che i v_i sono un sistema di generatori, sarà

$$w_1 = a_1 v_1 + \dots + a_m v_m,$$

e non tutti gli a_i sono zero. Diciamo che sia $a_1 \neq 0$, allora

$$v_1 = a_1^{-1} w_1 - a_1^{-1} a_2 v_2 - \dots - a_1^{-1} a_m v_m, \tag{6.8}$$

dunque

$$v_2, \dots, v_m, w_1 \tag{6.9}$$

sono un sistema di generatori. Questo perché i v_i sono una base, dunque ogni vettore u si può scrivere come combinazione lineare dei v_i,

$$u = c_1 v_1 + \dots + c_m v_m,$$

ma (6.8) ci dice

$$\begin{aligned} u &= c_1(a_1^{-1} w_1 - a_1^{-1} a_2 v_2 - \dots - a_1^{-1} a_m v_m) + \\ &\quad c_2 v_2 + \dots + c_m v_m \\ &= c_1 a_1^{-1} w_1 + (c_2 - c_1 a_1^{-1} a_2) v_2 + \dots + (c_m - c_1 a_1^{-1} a_m) v_m, \end{aligned}$$

dunque u si scrive anche come combinazione lineare dei (6.9).

Ora i vettori di (6.9) sono anche linearmente indipendenti, perché se fossero linearmente dipendenti, allora esisterebbero b_i non tutti nulli tali che

$$b_2 v_2 + \cdots + b_m v_m + b_1 w_1 = 0.$$

Qui deve essere $b_1 \neq 0$, dato che i v_i sono linearmente indipendenti. Dunque

$$w_1 = b_1^{-1} b_2 v_2 + \cdots + b_1^{-1} b_m v_m,$$

e mettendo insieme come appena fatto con (6.8), si vede che v_1 è combinazione lineare di $v_2, \dots, v_m$, il che per il punto (2) del Lemma 6.11.3 vuol dire che i v_i sono linearmente dipendenti, contro l'ipotesi che siano una base.

Dunque

$$v_2, \dots, v_m, w_1$$

è una base. Continuando con questo sistema (e riservandoci ogni volta di scambiare l'ordine dei v_i), ogni

$$v_{i+1}, \dots, v_m, w_1, \dots, w_i$$

è una base. Dato che abbiamo assunto $m < n$, i v_i finiscono prima di aver aggiunto tutti i w_i, dunque è una base

$$w_1, \dots, w_m.$$

ma allora w_n (che, dato che $m < n$, non fa parte dei vettori appena scritti) si scrive come combinazione lineare di $w_1, \dots, w_m$, e dunque per il punto (2) del Lemma 6.11.3 i w_i sono linearmente dipendenti, contro l'ipotesi che siano una base. □

6.11.5. **Corollario.**

(1) *In uno spazio vettoriale di dimensione n, se i vettori*

$$v_1, v_2, \dots, v_k$$

sono linearmente indipendenti, allora $k \leq n$. (Utile anche la riformulazione: ogni sistema di più di n vettori è linearmente dipendente.)

(2) *In uno spazio vettoriale di dimensione n, ogni sistema di generatori ha almeno n elementi.*

(3) *Sia V uno spazio vettoriale di dimensione n, e U un suo sottospazio. Allora $\dim(U) \leq n$.*

(4) *Sia V uno spazio vettoriale di dimensione n, e siano U, W due sottospazi tali che $U \subseteq W$. Allora sono equivalenti*
   - *$U = W$, e*
   - *$\dim(U) = \dim(W)$.*

Proof.

Punto (1). Per la Proposizione 6.11.2, posso aggiungere dei vettori a v_i per ottenere una base. Siccome per il Teorema 6.11.4 ogni base ha n elementi, deve essere $k \leq n$.

Punto (2) Stavolta uso la Proposizione 6.11.1 e il Teorema 6.11.4.

Punto (3). Prendo una base di U, che avrà k elementi, che sono dunque linearmente indipendenti. Uso il punto (1).

Punto (4). Abbiamo appena visto che $s = \dim(U) \leq \dim(W) = t$. Inoltre una base di U è fatta di s elementi di W che sono linearmente indipendenti, dunque posso aggiungerci dei vettori di W in modo da fare una base di W, fatta di t elementi. Me le due dimensioni sono eguali, cioè $s = t$, non c'è niente da aggiungere, dunque una base di U è già una base anche di W, e dunque $U = W$. $\square$

Dipartimento di Matematica, Università degli Studi di Trento, via Sommarive 14, 38123 Trento

Email address: andrea.caranti@unitn.it

URL: http://www.science.unitn.it/~caranti/