# Corso di Algebra A. Anno accademico 2004-2005

## Note integrative

- Il piccolo teorema di Fermat.

*Teorema. Se $p$ è un numero primo e $a$ è un numero intero positivo che non è un multiplo di $p$, allora vale*

$$a^{p-1} \equiv 1 \quad (p)$$

*Dimostrazione.* Per prima cosa dimostriamo che $p$ divide $\begin{pmatrix} p \\ i \end{pmatrix}$ quando $0 < i < p$. Infatti sappiamo che

$$\begin{pmatrix} p \\ i \end{pmatrix} i!\,(p-i)! = p!$$

e, per il teorema di decomposizione unica in prodotto di fattori primi, $p$, che divide il membro di destra, deve dividere il membro di sinistra. Poiché $p$ non può dividere $i!\,(p-i)!$ (che sono prodotti di numeri positivi strettamente minori di $p$) possiamo dedurre, per la proprietà caratterizzante dei numeri primi (ossia quella che dice che se $p$ è primo e divide un prodotto $ab$ allora $p$ deve dividere $a$ o $p$ deve dividere $b$), che $p$ deve dividere $\begin{pmatrix} p \\ i \end{pmatrix}$.

A questo punto possiamo osservare che, dati due numeri interi $a$ e $b$, lo sviluppo del binomio $(a+b)^p$ ha, modulo $p$, una scrittura molto semplificata. Infatti vale

$$(a+b)^p = \sum_{i=0}^{p} \begin{pmatrix} p \\ i \end{pmatrix} a^i b^{p-i} \equiv a^p + b^p \quad (p)$$

dato che, come sappiamo, $p$ divide tutti i coefficienti $\begin{pmatrix} p \\ i \end{pmatrix}$ $(1 < i < p)$.

In particolare, nel caso $b = 1$, possiamo concludere che

$$(a+1)^p \equiv a^p + 1 \quad (p)$$

Ora proviamo che, per ogni $a \in \mathbb{Z}$ vale

$$a^p \equiv a \quad (p)$$

Questa relazione, nel caso in cui $a$ non è multiplo di $p$, ci dà (dividendo per $a$) l'enunciato del teorema.

Ci basta dimostrare che, per ogni $a \in \mathbb{N}$,

$$a^p \equiv a \quad (p)$$

(il caso dei numeri negativi si ricava poi immediatamente).

Lo dimostriamo per induzione su $a$.

Il caso base, per $a = 0$,

$$0^p \equiv 0 \quad (p)$$

è banale.

Supponiamo ora che questa relazione sia vera fino ad $a = n$ e proviamo a dimostrare che

$$(n+1)^p \equiv n+1 \quad (p)$$

(se ci riusciamo la nostra dimostrazione è terminata).

Ora, per quanto visto sopra possiamo scrivere che

$$(n+1)^p \equiv n^p + 1 \quad (p)$$

Ma, per ipotesi induttiva, $n^p \equiv n \quad (p)$ per cui

$$(n+1)^p \equiv n+1 \quad (p)$$

■

Il piccolo teorema di Fermat si può anche esprimere con il linguaggio degli anelli $\mathbb{Z}_p$. Ecco come:

*Teorema. Se $p$ è un numero primo e $[a] \neq [0]$ in $\mathbb{Z}_p$ allora in $\mathbb{Z}_p$ vale*

$$[a]^{p-1} = [1]$$

Questo è (ovviamente) un enunciato equivalente:

*Teorema. Se $p$ è un numero primo allora per ogni $[a]$ in $\mathbb{Z}_p$ vale*

$$[a]^p = [a]$$

Visto che è chiaro che stiamo lavorando in $\mathbb{Z}_p$ si possono, per brevità, omettere le parentesi quadre:

*Teorema. Se $p$ è un numero primo allora per ogni $a$ in $\mathbb{Z}_p$ vale*

$$a^p = a$$

- La definizione formale di anello e di campo.

*Definizione.* Un anello (non banale, con identità) $R$ è un insieme che possiede due elementi speciali diversi fra loro (0 e 1) e dove sono definite due operazioni, che chiamiamo addizione ($+$) e moltiplicazione ($\cdot$), che soddisfano le seguenti proprietà:

- $\forall\ a, b, c \in R$ vale $(a+b)+c = a+(b+c)$ (proprietà associativa dell'addizione).
- $\forall\ a, b \in R$ vale $a+b = b+a$ (proprietà commutativa dell'addizione).
- $\forall\ a \in R$ vale $a+0 = a$ (0 è l'elemento identico per l'addizione).
- $\forall\ a \in R$ esiste un elemento "$-a$" in $R$ tale che $a+(-a) = 0$ (esistenza dell'opposto per l'addizione).
- $\forall\ a, b, c \in R$ vale $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (proprietà associativa della moltiplicazione).
- $\forall\ a \in R$ vale $a \cdot 1 = a$ (1 è l'elemento identico per la moltiplicazione).
- $\forall\ a, b, c \in R$ vale $(a+b) \cdot c = a \cdot c + b \cdot c$ e anche $a \cdot (b+c) = a \cdot b + a \cdot c$ (proprietà distributive).

*Osservazione importante*: la definizione data sopra, per "colpa" della richiesta $0 \neq 1$, non include l' anello "banale" $A = \{0\}$, in cui tutte le operazioni sono banali e dunque lo 0 funziona da elemento identico sia per la somma che per la moltiplicazione. Noi vogliamo considerare $A$ come un anello con identità e dunque lo aggiungiamo a quelli (non banali) che risultano dalla definizione.

*Definizione.* Un anello (con identità) $R$ che soddisfa anche la seguente proprietà si dice "commutativo":
– $\forall\ a, b \in R$ vale $a \cdot b = b \cdot a$ (proprietà commutativa della moltiplicazione).

*Definizione.* Un anello (con identità) $R$ che soddisfa anche la seguente proprietà si dice "privo di divisori di 0":
– $\forall\ a, b \in R$ vale che $a \cdot b = 0$ implica $a = 0$ o $b = 0$.

*Definizione.* Un elemento $u$ di un anello con identità $R$ si dice una "unità" o "invertibile" se esiste un $v \in R$ tale che $u \cdot v = v \cdot u = 1$ (cioè se esiste un inverso sinistro e destro di $u$ rispetto alla moltiplicazione).

*Definizione.* Un anello non banale, con identità, commutativo, che soddisfa anche la seguente proprietà è detto "campo":
– ogni $a \in R - \{0\}$ è invertibile (esistenza dell'inverso rispetto alla moltiplicazione per tutti gli elementi diversi da 0).

*Osservazione.* Notiamo che un campo è anche automaticamente "privo di divisori dello 0".

- Il teorema di fattorizzazione unica per polinomi.

Visto che in classe lo abbiamo enunciato velocemente, riportiamo qui il testo preciso del teorema di fattorizzazione unica per polinomi (la dimostrazione è un esercizio caldamente consigliato, ed è del tutto simile a quella per la fattorizzazione unica in $\mathbb{Z}$; la trovate comunque sul Childs).

*Teorema. Ogni polinomio di grado $\geq 1$ in $K[x]$ (dove $K$ è un campo) è irriducibile o si fattorizza come prodotto di polinomi irriducibili. Inoltre, se*

$$f = p_1 p_2 \cdots p_s = q_1 q_2 \cdots q_t$$

*sono due fattorizzazioni del polinomio $f$ come prodotto di irriducibili, allora vale che $s = t$ e che i polinomi $p_i$ e i polinomi $q_j$ sono a due a due associati.*

Notiamo che nel teorema precedente, quando si scrive

$$f = p_1 p_2 \cdots p_s$$

si intende che i $p_i$ sono scritti anche con ripetizioni. Un altro modo di scrivere la fattorizzazione di un polinomio è quello di evidenziare gli irriducibili scrivendo ciascuno con l'esponente con cui compare. Per esempio si scriverà

$$h = q_1^{r_1} q_2^{r_2} \cdots q_t^{r_t}$$

dove i $q_j$ sono polinomi irriducibili e gli $r_i$ sono numeri naturali positivi.

In questo modo è facile individuare, proprio come avveniva in $\mathbb{Z}$, un $MCD$ di due polinomi. Se infatti prendiamo una fattorizzazione di un altro polinomio:

$$g = p_1^{s_1} p_2^{s_2} \cdots p_j^{r_j}$$

allora un $MCD$ $(h, g)$ si otterrà facendo il prodotto degli irriducibili che compaiono sia fra i $p_m$ che fra i $q_n$, ciascuno preso col minimo esponente fra i due esponenti che troviamo nelle due fattorizzazzioni.

Per esempio, in $\mathbb{Q}[x]$, se

$$h(x) = (x-1)^2(x^2-5)^3(x^4-7x+7)$$

$$g(x) = (x-1)^7(x^2-5)(x^5+11x^2+11)^2$$

allora un $MCD$ $(h, g)$ è

$$(x-1)^2(x^2-5)$$

Gli altri $MCD$ $(h, g)$, come sappiamo, sono tutti i polinomi associati a $(x-1)^2(x^2-5)$.

*Domanda.* Perché i polinomi che compaiono in queste fattorizzazioni sono irriducibili in $\mathbb{Q}[x]$ ? Per alcuni è facile provarlo, per altri ci vogliono tanti calcoli. Come affrontereste il problema ? Vedremo più avanti in questo corso il criterio di irriducibilità di Eisenstein che ci darà un aiuto.

*Osservazione.* Questo teorema di fattorizzazione unica vale in $K[x]$ quando $K$ è un CAMPO. Guardate cosa può succedere in $\mathbb{Z}_{30}[x]$:

$$x^2 - 1 = (x-1)(x-29) = (x-19)(x-11)$$

Fate i conti e verificate che è vero. Queste sono due distinte fattorizzazioni in irriducibili (giacché si tratta di polinomi di grado 1) !

- Il Lemma di Bezout per polinomi.

*Teorema (il "Lemma di Bezout per polinomi"). Dimostrare che un massimo comun divisore $d$ di $f$ e $g$ si può scrivere nella forma $d = af + bg$ per opportuni polinomi $a$, $b$ di $\mathbb{K}[x]$.*

*Dimostrazione.*

Consideriamo l'insieme $Z = \{z = xf + yg : \ (x, y \in \mathbb{K}[x]) \wedge (z \neq 0)\}$ e scegliamo un elemento $d = af + yg \in Z$ tale che $\forall z \in Z$ vale $deg(d) \leq deg(z)$. Adesso dividiamo $f$ per $d$ ottenendo $f = qd + r$ con $r = 0$ oppure $deg(r) < deg(d)$ e osserviamo come $r$ sia un polinomio della forma $xf + yg$:

$$r = f - qd = f - q(af + yg) = f - qaf - qyg = (1 - qa)f + (-qy)g$$

Se $r$ fosse diverso da zero apparterrebbe a $Z$, ma avendo supposto $d$ come elemento di grado minimo di $Z$ e sapendo che $deg(r) < deg(d)$ dobbiamo scartare questa ipotesi, per cui $r$ è il polinomio zero. Dunque:

$$f(x) = c(x)d(x)$$

cioè $d \mid f$. Ripetendo lo stesso ragionamento per $g$ troviamo che $d \mid g$.

Resta da mostrare che se un polinomio $c$ è tale che $c \mid f$ e $c \mid g$ allora $c \mid d$. Per ipotesi abbiamo che $f = ck$ e $g = cl$ per opportuni polinomi $k$ e $l$; inoltre, considerando quanto visto al punto precedente, possiamo scrivere:

$$d = af + bg = a(ck) + b(cl) = c(ak + bl)$$

dunque $c \mid d$.

*Ringraziamo Giacomo Del Rio per avere scritto questa dimostrazione, che era stata lasciata come compito per casa.*

- I numeri complessi.

*Definizione* $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$, e inoltre si hanno due operazioni di somma e di prodotto, $\cdot, +$ definite come:
$(a + bi) + (c + di) = a + c + (b + d)i$,
$(a + bi)(c + di) = ac - bd + (ad + bc)i$.

*Teorema* Con gli elementi $1 = 1 + 0i$ e $0 = 0 + 0i$ si ha che $(\mathbb{C}, +, \cdot, 0, 1)$ è un campo
*Dimostrazione.* La dimostrazione è lasciata come esercizio. □
I numeri complessi si possono rappresentare come punti del piano $\mathbb{R}^2$, mandando il numero $a + bi$ nel punto $(a, b)$.
*Definizione* Dato $a + bi \in \mathbb{C}$, definiamo la sua *norma* come $|a + bi| = \sqrt{a^2 + b^2}$, e il suo *coniugato* come $\bar{a + bi} = a - bi$.
*Proposizione* Per ogni numero complesso $z$ si ha che $|z|^2 = z\bar{z}$, e inoltre $\left|\frac{z}{|z|}\right| = 1$.
*Dimostrazione.* La dimostrazione è lasciata come esercizio. □
*Definizione* Dato un numero complesso $z$ di norma uguale a 1, l'angolo corrispondente al punto sulla circonferenza unitaria associato a $z$ in $\mathbb{R}^2$ si chiama *argomento* di $z$, ed è definito a meno di aggiungere multipli di $2\pi$. L'argomento è definito anche come quell'angolo per il quale vale (per $z$ di norma unitaria)

$$z = cos(arg(z)) + isin(arg(z))$$

Se $z$ è un numero complesso qualsiasi diverso da 0, definiamo $arg(z)$ come $arg(z) = arg\left(\frac{z}{|z|}\right)$.
*Proposizione* Dati due numeri complessi $z, w$ diversi da 0, si ha che

$$|zw| = |z||w|,\ arg(zw) = arg(z) + arg(w)$$

*Dimostrazione.* Se $s = arg(z)$ e $t = arg(w)$, possiamo scrivere

$$z = |z|(cos(s) + isin(s)),\quad w = |w|(cos(t) + isin(t))$$

Si ha che

$$zw = |z||w|(cos(s) + isin(s))(cos(t) + isin(t)) =$$

$$|z||w|\left((cos(s)cos(t) - sin(s)sin(t)) + (cos(s)sin(t) + sin(s)cos(t))i\right)$$

Non dimostriamo il fatto (standard) che per ogni $s, t$ vale

$$cos(s)cos(t) - sin(s)sin(t) = cos(s+t), \quad cos(s)sin(t) + sin(s)cos(t) = sin(s+t)$$

da cui seguono immediatamente le uguaglianze cercate (esercizio). □

La proposizione precedente permette di rappresentare graficamente il prodotto di numeri complessi: dati due numeri, il loro prodotto si ottene ruotando di un angolo pari alla somma degli angoli relativi ai due numeri iniziali, e allontanandosi dall'origine di una lunghezza pari al prodotto delle lunghezze relative ai due numeri iniziali. Questo permette facilmente di dimostrare che ogni numero complesso diverso da 0 ha un inverso rispetto alla moltiplicazione. Il seguente esempio dimostra lo stesso fatto in un altro modo, ossia risolvendo un sistema lineare.

*Esempio* Dato un numero complesso $z = a + bi$ diverso da zero, il numero complesso $z^{-1} = x + iy$ può essere trovato risolvendo il problema

$$(a + bi)(x + iy) = 1$$

Questo implica le condizioni $ax - by = 1$ e $ay + bx = 0$. Queste due condizioni determinano un sistema

$$\begin{cases} ax - by = 1 \\ bx + ay = 0 \end{cases}$$

Se $a \neq 0$, dalla prima equazione ricaviamo $x = \frac{1}{a}(1 + by)$, e andando a sostituire nella seconda equazione $\frac{b}{a}(1 + by) + ay = 0$, da cui $y = -(a + \frac{b^2}{a})^{-1}\frac{b}{a}$ e sostituendo nella equazione per $x$ si ottiene anche il valore di $x$. Se invece $a = 0$, allora deve essere $x = 0$ e $y = -\frac{1}{b}$.

- Come costruire nuovi anelli e campi.

Ripassiamo le congruenze fra numeri interi:

$$a \equiv b \quad (m)$$

significa che $m \mid a - b$ oppure, equivalentemente, che il resto della divisione euclidea di $a$ per $m$ è uguale al resto della divisione euclidea di $b$ per $m$.

Questa definizione si basa dunque sulla divisione euclidea. Ma, come abbiamo visto, abbiamo una divisione euclidea anche in $K[x]$ (dove $K$ è un campo).

Allora, dato in $K[x]$ un polinomio $m(x)$ diverso da 0, possiamo definire cosa vuol dire che due polinomi $a(x)$ e $b(x)$ sono congrui fra di loro modulo $m(x)$:

$$a(x) \equiv b(x) \quad (m(x))$$

significa che $m(x) \mid a(x) - b(x)$ oppure, equivalentemente, che il resto della divisione euclidea di $a(x)$ per $m(x)$ è uguale al resto della divisione euclidea di $b(x)$ per $m(x)$.

Continuiamo a osservare le analogie con il caso dei numeri interi. La congruenza lineare

$$ax \equiv b \quad (m)$$

ha soluzione se e solo se $MCD\ (a, m) \mid b$. Questo si otteneva utilizzando il lemma di Bezout.

Ma noi conosciamo una versione del lemma di Bezout per polinomi. Applicandolo, troviamo che, dati i polinomi $a(x)$, $b(x)$, $m(x)$, l'equazione

$$a(x)f(x) \equiv b(x) \quad (m(x))$$

ha soluzione se e solo se $MCD\ (a(x), m(x)) \mid b(x)$.

*Osservazione. Qui la notazione è un po' imprecisa, perché ci sono più di un* $MCD\ (a(x), m(x))$ *- differiscono fra di loro per uno scalare; ovviamente intendiamo dire che ogni* $MCD\ (a(x), m(x))$ *divide* $b(x)$*.*

Utilizzando le congruenze fra interi abbiamo introdotto gli anelli $Z_m$. Gli elementi di $Z_m$ sono le classi $[0], [1], \ldots, [m-1]$ dove per esempio

$$[1] = \{n \in \mathbb{Z} \mid n \equiv 1 \quad (m)\}$$

ossia è l'insieme di tutti i numeri interi che sono congrui a 1 modulo $m$.

Estendendo un po' la notazione abbiamo deciso di indicare la classe [1] anche per esempio come $[1+m]$ o come $[1+17m]$. Insomma abbiamo convenuto che per indicare una classe non conta quale fra i suoi elementi mettiamo fra parentesi quadre (del resto i numeri interi congrui a 1 modulo $m$ sono la stessa cosa dei numeri interi congrui a $1+m$ - o a $1+17m$ - modulo $m$).

Con questa notazione, la somma e la moltiplicazione in $\mathbb{Z}_m$ sono definite così:

$$[a] + [b] = [a+b]$$

$$[a][b] = [ab]$$

Si ottiene così un anello commutativo (con unità - che è la classe [1]).

Se poi $m = p$ è un numero primo, allora $\mathbb{Z}_p$ è un campo: infatti se $[a] \neq [0]$ l'equazione

$$[a][x] = [1]$$

ha sempre soluzione (perchè' ? Il motivo lo abbiamo detto poche righe sopra...) e dunque ogni elemento diverso da zero è invertibile.

Vorremmo adesso, tenendo presente la costruzione di $\mathbb{Z}_m$, utilizzare le congruenze fra polinomi per costruire dei nuovi anelli e dei nuovi campi.

Prendiamo in $K[x]$ un polinomio $m(x)$, che per semplicità supponiamo di grado $\geq 1$ (per il caso degenere in cui $m(x)$ è una costante diversa da zero si veda un esercizio più avanti), e costruiamo un anello che chiameremo $\dfrac{K[x]}{(m(x))}$.

Gli elementi di $\dfrac{K[x]}{(m(x))}$ sono in corrispondenza con i polinomi di $K[x]$ che sono i possibili resti della divisione euclidea per $m(x)$: si tratta dunque delle classi $[g(x)]$, dove $g(x)$ varia fra tutti i polinomi di grado strettamente minore di $m(x)$ e

$$[g(x)] = \{f(x) \in K[x] \mid f(x) \equiv g(x) \quad (m(x))\}$$

ossia è l'insieme di tutti i polinomi che sono congrui a $g(x)$ modulo $m(x)$.

Come nel caso dei numeri interi, posso rappresentare una classe mettendo fra parentesi quadre uno qualunque dei suoi elementi: per esempio se $m(x) = x^3 + 1$, la classe $[x^2]$ la posso rappresentare anche come $[x^2 + 3x(x^3 + 1)]$.

Analogamente al caso $Z_m$, la somma e la moltiplicazione in $\frac{K[x]}{(m(x))}$ sono definite così:

$$[a(x)] + [b(x)] = [a(x) + b(x)]$$

$$[a(x)][b(x)] = [a(x)b(x)]$$

Si verifica facilmente che $\frac{K[x]}{(m(x))}$ è un anello commutativo (con unità - che è [1]).

Se poi $m(x) = p(x)$ è un polinomio irriducibile allora $\frac{K[x]}{(p(x))}$ è un campo: infatti se $[a(x)] \neq [0]$ l'equazione

$$[a(x)][f(x)] = [1]$$

ha sempre soluzione (perchè ? Vedi sopra...) e dunque ogni elemento diverso da zero è invertibile.

*Esempio 1.* Da questo punto di vista, $\mathbb{C} = \frac{\mathbb{R}[x]}{(x^2+1)}$. Verificate ! Nel costruire i numeri complessi abbiamo "aggiunto" ai numeri reali una radice (che abbiamo chiamato $i$) di $x^2 + 1$. Questo suggerisce l'idea che quando si costruisce il campo $\frac{K[x]}{(p(x))}$ quello che stiamo facendo è aggiungere a $K$ una radice del polinomio irriducibile $p(x)$.....come vedremo nel secondo esempio.

*Esempio 2.* Consideriamo $\mathbb{Q}[x]$ e il polinomio $x^5 - 7x^2 + 14$ (tale polinomio è irriducibile - potete verificarlo con Eisenstein per esempio). Costruiamo ora il campo $\frac{\mathbb{Q}[x]}{(x^5-7x^2+14)}$ e chiamiamolo $K$. Consideriamo il polinomio $y^5 - 7y^2 + 14$ in $K[y]$. Tale polinomio ha radici in $K$ ? Come sappiamo non ha radici in $\mathbb{Q}$, anzi, è irriducibile in $\mathbb{Q}[x]$.

Ma in $K$ abbiamo l'elemento $[x]$, ossia la classe $[x]$. Proviamo a vedere se è una radice del polinomio $y^5 - 7y^2 + 14$, ossia proviamo a calcolare $[x]^5 - 7[x]^2 + 14$.

Come sappiamo, viste le leggi di addizione e moltiplicazione, questo equivale a $[x^5 - 7x^2 + 14)]$. Ma la classe $[x^5 - 7x^2 + 14]$ è la stessa cosa di $[0]$ in $\frac{\mathbb{Q}[x]}{(x^5-7x^2+14)}$. Infatti il resto della divisione euclidea di $x^5 - 7x^2 + 14$ per $x^5 - 7x^2 + 14$ è 0 !

Dunque la classe $[x]$ è una radice del polinomio $y^5 - 7y^2 + 14$. Possiamo insomma pensare che abbiamo costruito il nuovo campo $\frac{\mathbb{Q}[x]}{(x^5-7x^2+14)}$ "aggiungendo" a $\mathbb{Q}$ una radice del polinomio $x^5 - 7x^2 + 14$.

*Esempio 3.* Consideriamo il caso in cui $K = \mathbb{Z}_p$, con $p$ numero primo.

Per esempio prendiamo $\mathbb{Z}_7[x]$ e il polinomio $x^3+2x+1$. Tale polinomio è irriducibile in $\mathbb{Z}_7[x]$ (perchè ? Ha grado 3 e si verifica che non ha radici in $\mathbb{Z}_7$..).

Allora $\dfrac{\mathbb{Z}_7[x]}{(x^3+2x+1)}$ è un campo. Quanti elementi ha ? Come abbiamo osservato sopra, gli elementi di $\dfrac{\mathbb{Z}_7[x]}{(x^3+2x+1)}$ sono in corrispondenza biunivoca con i polinomi in $\mathbb{Z}_7[x]$ di grado strettamente minore a 3. Dunque sono tanti quanti i polinomi di grado $\leq 2$ a coefficienti in $\mathbb{Z}_7$. Scriviamo un tale polinomio: $ax^2+bx+c$. Abbiamo 7 scelte per $a$, 7 per $b$ e 7 per $c$. Dunque in $\dfrac{\mathbb{Z}_7[x]}{(x^3+2x+1)}$ ci sono $7^3 = 343$ elementi.

Il nuovo campo che abbiamo costruito è ancora finito. L'importanza di questo esempio nasce dal fatto che tutti i campi finiti si possono ottenere così:

*Teorema. Ogni campo finito si può costruire come* $\dfrac{Z_p[x]}{(q(x))}$ *dove* $p$ *è un numero primo e* $q(x)$ *è un polinomio irriducibile in* $Z_p[x]$.

Questo teorema non lo dimostriamo. Sappiamo però subito ricavare dal suo enunciato:

*Corollario. Ogni campo finito ha cardinalità uguale alla potenza di un numero primo, ossia del tipo* $p^n$*, dove* $p$ *è un numero primo.*

Per completare le nostre informazioni su questo argomento, manca un ultimo teorema che enunciamo in maniera un po' informale:

*Teorema (enunciato informalmente). Viceversa, per ogni potenza di un numero primo* $p^n$*, esiste un campo finito di cardinalità* $p^n$*, e in un certo senso si può dire che ne esiste uno solo.*

Osserviamo che questo teorema implica un fatto notevole che nel corso non abbiamo dimostrato, ossia che, per ogni numero primo $p$, in $\mathbb{Z}_p[x]$ ci sono polinomi irriducibili per ogni grado $n \geq 1$ (solo grazie alla loro esistenza, infatti, potrò costruire un campo con $p^n$ elementi). Sotto questo aspetto i campi $\mathbb{Z}_p$ si comportano come $\mathbb{Q}$ e non come $\mathbb{R}$ o $\mathbb{C}$.

*Esercizio.* Cosa è $\dfrac{K[x]}{(m(x))}$ nel caso in cui facciamo la stessa costruzione con $m(x)$ uguale ad una costante (diversa da zero) ?

*Esercizio.* Cosa accade in $\dfrac{K[x]}{(m(x))}$ nel caso in cui $m(x)$ non sia irriducibile ? Per esempio: è vero che ogni elemento di $\dfrac{K[x]}{(m(x))}$ è invertibile ? È vero o no che $\dfrac{K[x]}{(m(x))}$ è un anello privo di divisori dello zero ?

- La definizione formale di spazio vettoriale.

  Uno spazio vettoriale su un campo $K$ è un insieme $V$ su cui sono definite due operazioni: la somma fra due elementi di $V$, il cui risultato è ancora un elemento di $V$, e il prodotto di un elemento del campo $K$ (ossia di uno "scalare") per un elemento di $V$, il cui risultato è un elemento di $V$. Tali operazioni soddisfano le seguenti proprietà:

  - $\forall\, u, v, w \in V$ vale $(u+v)+w = u+(v+w)$ (proprietà associativa dell'addizione).

  - $\forall\, v, w \in V$ vale $v + w = w + v$ (proprietà commutativa dell'addizione).

  - esiste $O \in V$ tale che $\forall\, v \in V$ vale $v + O = v$ ($O$ è l'elemento identico per l'addizione. ATTENTI ! L'elemento $O$ appartiene a $V$, non va confuso con lo scalare 0 che appartiene al campo $K$).

  - $\forall\, v \in V$ esiste un elemento "$-v$" in $V$ tale che $v + (-v) = 0$ (esistenza dell'opposto per l'addizione).

  - $\forall \lambda, \mu \in K,\ \forall\, v, w \in V$ vale $\lambda(v + w) = \lambda v + \lambda w$ e anche $(\lambda + \mu)v = \lambda v + \mu v$ (proprietà distributive della moltiplicazione per scalare).

  - $\forall \lambda, \mu \in K,\ \forall\, v \in V$ vale $(\lambda\mu)v = \lambda(\mu v)$ (proprietà associativa della moltiplicazione per scalare).

  - $\forall\, v \in V$ vale $1v = v$ e $0v = O$ (proprietà di 0 e 1 rispetto alla moltiplicazione).

  Di solito chiameremo "vettori" gli elementi di uno spazio vettoriale $V$, mentre, come si è visto, gli elementi del campo $K$ saranno gli "scalari".

  L'esempio che studieremo di più è quello di $\mathbb{R}^n$. Si tratta dello spazio vettoriale sul campo $\mathbb{R}$ dei numeri reali così definito:

  (1) Gli elementi, ossia i vettori, sono le liste ordinate formate da $n$ numeri reali:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ \dots \\ a_{n-1} \\ a_n \end{pmatrix}$$

(2) La somma fra vettori è definita da:

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ ... \\ ... \\ a_{n-1} \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ ... \\ ... \\ b_{n-1} \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \\ ... \\ ... \\ a_{n-1} + b_{n-1} \\ a_n + b_n \end{pmatrix}$$

(3) Il prodotto per scalari è definito da:

$$\lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ ... \\ ... \\ a_{n-1} \\ a_n \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \lambda a_3 \\ ... \\ ... \\ \lambda a_{n-1} \\ \lambda a_n \end{pmatrix}$$

- Qualche osservazione sulle basi di spazi vettoriali.

Abbiamo dato in classe la definizione di base di uno spazio vettoriale. In realtà la abbiamo data nel caso di $\mathbb{R}^n$, ma la definizione generale è del tutto simile: una base, lo ricordiamo, è data da un insieme di vettori che generano lo spazio e sono linearmente indipendenti.

Fissiamo uno spazio vettoriale $V$ (sul campo $K$) e supponiamo che ammetta una base finita $\{v_1, v_2, \ldots, v_n\}$.

*Proposizione. Ogni vettore $v \in V$ si scrive IN MODO UNICO come combinazione lineare degli elementi della base.*

*Dimostrazione.* Il vettore $v$ si può scrivere come combinazione lineare degli elementi della base perché gli elementi della base generano $V$. L'unicità di una tale combinazione lineare è conseguenza della lineare indipendenza degli elementi della base. Infatti, supponiamo che si possa scrivere:

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n = b_1v_1 + b_2v_2 + \cdots + b_nv_n$$

dove gli $a_i$ e i $b_j$ sono elementi del campo $K$. Sottraendo abbiamo:

$$O = (a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_n - b_n)v_n$$

Ma sappiamo che i vettori $v_1, v_2, \ldots, v_n$ sono linearmente indipendenti. Dunque la combinazione lineare che abbiamo scritto sopra, e che ha come risultato $O$, deve avere tutti i coefficienti nulli. Così possiamo concludere che $a_i = b_i$ per ogni $i$, ossia che esiste un solo modo di scrivere $v$ come combinazione lineare degli elementi della base data.

■

Come annunciato, in questo corso considereremo quasi sempre spazi vettoriali che ammettono una base finita. Abbiamo detto in classe che risulterà che tutte le basi di uno spazio che ammette una base finita hanno la stessa cardinalità e che tale cardinalità si chiama la "dimensione" dello spazio vettoriale; discuteremo più avanti nel corso questa affermazione. Per il momento ci limitiamo ad osservare, mediante il seguente teorema, che se uno spazio vettoriale ammette un insieme finito di generatori, allora ammette anche una base finita. In altre parole i generatori dati potrebbero essere molti, e sovrabbondanti, ma è sempre possibile estrarre dall'insieme dei generatori un sottoinsieme che è una base.

*Teorema.* Sia $V$ uno spazio vettoriale (sul campo $K$) e supponiamo che $V$ sia generato da un insieme finito di vettori $\{w_1, w_2, \ldots, w_s\}$. Allora è possibile estrarre da $\{w_1, w_2, \ldots, w_s\}$ un sottoinsieme $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ che è una base di $V$.

*Dimostrazione.* Consideriamo l'insieme:

$$\mathcal{M} = \{A \subset \{w_1, w_2, \ldots, w_s\} \mid A \text{ è un insieme di vettori lin. indip.}\}$$

e notiamo che $\mathcal{M}$ non è vuoto, in quanto contiene certamente i sottoinsiemi di $\{w_1, w_2, \ldots, w_s\}$ di cardinalità 1, tipo $\{w_1\}$ o $\{w_2\}$. Fra tutti gli elementi di $\mathcal{M}$, ossia fra tutti i sottoinsiemi che appartengono a $\mathcal{M}$, ne prendo uno di cardinalità massima: $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$; per la osservazione precedente sarà $n \geq 1$ ossia tale insieme sarà non vuoto.

Questo $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ è proprio il nostro candidato ad essere una base di $V$.

Parte bene perché per come lo abbiamo costruito è un insieme di vettori linearmente indipendenti. Resta da dimostrare che genera $V$. Per questo basterà mostrare che con combinazioni lineari dei vettori di $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ posso ottenere uno qualunque dei vettori di $\{w_1, w_2, \ldots, w_s\}$, visto che sappiamo che questi generano $V$ (verificare di aver capito bene questo passaggio !).

Se $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\} = \{w_1, w_2, \ldots, w_s\}$ abbiamo già finito. Se invece

$$\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\} \subsetneq \{w_1, w_2, \ldots, w_s\}$$

allora prendiamo un vettore, diciamo $w_r$, che non appartiene a $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$. Dobbiamo dimostrare che $w_r$ si può scrivere come combinazione lineare dei vettori $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$.

Se consideriamo l'insieme $\{w_r, w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ notiamo che certamente questo non è un insieme di vettori linearmente indipendenti, se no apparterrebbe a $\mathcal{M}$ e non sarebbe più vero che $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$ aveva cardinalità massima fra gli elementi di $\mathcal{M}$.

Dunque esiste una combinazione lineare:

$$a_r w_r + a_{i_1} w_{i_1} + a_{i_2} w_{i_2} + \cdots + a_{i_n} w_{i_n} = 0$$

che è non banale, ossia i coefficienti non sono tutti zero. In particolare risulta che non può essere $a_r = 0$, altrimenti resterebbe una combinazione lineare non banale:

$$a_{i_1} w_{i_1} + a_{i_2} w_{i_2} + \cdots + a_{i_n} w_{i_n} = 0$$

che contraddirrebbe la lineare indipendenza di $\{w_{i_1}, w_{i_2}, \ldots, w_{i_n}\}$.

Visto dunque che $a_r \neq 0$, si può dividere tutto per $a_r$ ottenendo:

$$w_r = -\frac{a_{i_1}}{a_r} w_{i_1} - \frac{a_{i_2}}{a_r} w_{i_2} - \cdots - \frac{a_{i_n}}{a_r} w_{i_n}$$

che è la combinazione lineare cercata.

■

- Applicazioni lineari e matrici.

Consideriamo una applicazione lineare:

$$L \,:\, V \to W$$

dove $V$ e $W$ sono due spazi vettoriali sul campo $K$.
Prendiamo in $V$ una base $\{e_1, e_2, \ldots, e_n\}$ e in $W$ una base $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$.
Dato un elemento $v \in V$, proviamo a scrivere la sua immagine $L(v)$. Sappiamo che $v$ si può scrivere in modo unico come combinazione lineare degli elementi della base scelta:

$$v = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$$

Per la linearità di $L$ allora:

$$L(v) = a_1 L(e_1) + a_2 L(e_2) + \cdots + a_n L(e_n)$$

Dunque, per conoscere $L$, ossia per saper dire qual è l'immagine di un qualsiasi elemento $v \in V$, basta conoscere $L(e_1), L(e_2), \ldots, L(e_n)$.

Per poter descrivere $L(e_1), L(e_2), \ldots, L(e_n)$, che sono vettori di $W$, possiamo servirci della base $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$: per ogni $i$, $L(e_i)$ si può scrivere in modo unico come

$$L(e_i) = a_{1i}\epsilon_1 + a_{2i}\epsilon_2 + \cdots + a_{mi}\epsilon_m$$

Ecco che entrano in scena le matrici.

*Definizione La matrice associata all'applicazione lineare L nelle basi* $\{e_1, e_2, \ldots, e_n\}$ *e* $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$ *è data dalla seguente griglia di m righe per n colonne:*

$$[L]_{\substack{e_1, e_2, \ldots, e_n \\ \epsilon_1, \epsilon_2, \ldots, \epsilon_m}} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}$$

*che si può anche scrivere sinteticamente come:*

$$[L]_{\substack{e_1, e_2, \ldots, e_n \\ \epsilon_1, \epsilon_2, \ldots, \epsilon_m}} = (a_{ij}) \quad \begin{matrix} i = 1, 2, 3, \ldots, n \\ j = 1, 2, 3, \ldots, m \end{matrix}$$

Notiamo che la matrice $[L]$ (da ora in poi, per semplificare la notazione, ometteremo il riferimento alle basi tutte le volte che questo non creerà ambiguità) si ottiene ponendo uno accanto all'altro i vettori $L(e_1), L(e_2), \ldots, L(e_n)$, scritti come vettori colonna nella base scelta di $W$. Dunque guardando la matrice possiamo sapere tutto quello che ci serve sull'applicazione lineare $L$.

Applichiamo ciò che abbiamo detto fin qui al calcolo di $L(v)$ dove, come prima, $v \in V$ e

$$v = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$$

Allora

$$L(v) = a_1 L(e_1) + a_2 L(e_2) + \cdots + a_n L(e_n)$$

ossia

$$L(v) = a_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \cdots \\ \cdots \\ a_{m1} \end{pmatrix} + a_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \cdots \\ \cdots \\ a_{m2} \end{pmatrix} + \cdots + a_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \cdots \\ \cdots \\ a_{mn} \end{pmatrix}$$

Ribadiamo che i vettori che compaiono nella espressione scritta qui sopra sono i vettori colonna della matrice $[L]$, e che stiamo considerando lo spazio $W$ munito della sua base $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$. Dunque per esempio il vettore colonna

$$\begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \cdots \\ \cdots \\ a_{mn} \end{pmatrix}$$

rappresenta l'elemento

$$L(e_n) = a_{1n}\epsilon_1 + a_{2n}\epsilon_2 + \cdots + a_{mn}\epsilon_m$$

Possiamo allo stesso modo rappresentare i vettori di $V$ come vettori colonna relativamente alla base $\{e_1, e_2, \ldots, e_n\}$; in tal modo allora per esempio il nostro $v$ si scrive:

$$v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \cdots \\ \cdots \\ a_n \end{pmatrix}$$

Ricordando il prodotto "righe per colonne" di cui abbiamo parlato in classe (e che qui non ridefiniamo), possiamo anche scrivere così:

$$L(v) = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \cdots \\ \cdots \\ a_n \end{pmatrix}$$

Dunque la matrice $[L]$, mediante il prodotto righe per colonne, ci permette di calcolare come agisce la applicazione $L$ sui vettori di $V$.

*Esempio.* Facciamo ora un esempio che mostra come la matrice associata ad una applicazione lineare dipenda dalle basi scelte. Consideriamo gli spazi vettoriali $\mathbb{R}^4$ e $\mathbb{R}^3$ con le loro basi standard, rispettivamente

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

e

$$\epsilon_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \epsilon_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \epsilon_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Consideriamo poi la applicazione lineare

$$L : \mathbb{R}^4 \to \mathbb{R}^3$$

così definita (sappiamo, per quanto osservato sopra, che per definire una applicazione lineare basta dare il suo valore sugli elementi di una base):

$$L(e_1) = 2\epsilon_1 + \sqrt{3}\epsilon_2$$
$$L(e_2) = 3\epsilon_1 + \epsilon_2 + \epsilon_3$$
$$L(e_3) = \epsilon_1 + 7\epsilon_2 + 8\epsilon_3$$
$$L(e_4) = 2\epsilon_2 + 4\epsilon_3$$

Come sappiamo, a questa applicazione corrisponde la seguente matrice relativamente alle basi standard:

$$[L]_{\substack{e_1, e_2, e_3, e_4 \\ \epsilon_1, \epsilon_2, \epsilon_3}} = \begin{pmatrix} 2 & 3 & 1 & 0 \\ \sqrt{3} & 1 & 7 & 2 \\ 0 & 1 & 8 & 4 \end{pmatrix}$$

Dunque, preso per esempio il vettore

$$v = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

(scritto rispetto alla base standard) per calcolare $L(v)$ basta fare il prodotto:

$$\begin{pmatrix} 2 & 3 & 1 & 0 \\ \sqrt{3} & 1 & 7 & 2 \\ 0 & 1 & 8 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

che dà come risultato

$$L(v) = \begin{pmatrix} 11 \\ \sqrt{3} + 31 \\ 42 \end{pmatrix}$$

che è un vettore di $\mathbb{R}^3$ scritto nella base standard.

Supponiamo ora di voler cambiare le basi. Prendiamo allora in $\mathbb{R}^4$ la nuova base (verificare che si tratta davvero di una base !):

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

e in $\mathbb{R}^3$ la nuova base (anche qui verificare !):

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, w_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, w_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

Proviamo a scrivere la matrice

$$[L]_{\substack{v_1, v_2, v_3, v_4 \\ w_1, w_2, w_3}}$$

che rappresenterà la solita applicazione lineare $L$ (ma sarà diversa dalla matrice trovata prima, relativa alle basi standard).

Nella prima colonna della matrice che stiamo per costruire, dovremo mettere il vettore $L(v_1)$ scritto in termini della base $\{w_1, w_2, w_3\}$. Calcoliamolo, facendo in un primo tempo riferimento alle basi standard (d'altra parte la nostra $L$ la abbiamo definita tramite le basi standard, dunque non possiamo far altro che ripartire da quella definizione).

$$L(v_1) = L\left(\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right) = L\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right) + L\left(\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 5 \\ \sqrt{3}+1 \\ 1 \end{pmatrix}$$

Fin qui questo vettore è scritto ancora in termini della base standard di $\mathbb{R}^3$. Ora lo esprimiamo in termini della base $\{w_1, w_2, w_3\}$. Si verifica che risulta

$$\begin{pmatrix} 5 \\ \sqrt{3}+1 \\ 1 \end{pmatrix} = (4-\sqrt{3})\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + (\sqrt{3}+1)\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 2\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

e dunque

$$L(v_1) = (4-\sqrt{3})w_1 + (\sqrt{3}+1)w_2 - 2w_3$$

Allora il vettore da inserire come prima colonna della matrice

$$[L]_{\substack{v_1, v_2, v_3, v_4 \\ w_1, w_2, w_3}}$$

è

$$\begin{pmatrix} 4-\sqrt{3} \\ \sqrt{3}+1 \\ -2 \end{pmatrix}$$

Procedendo allo stesso modo per le altre colonne si ottiene (verificare!):

$$[L]_{\substack{v_1,v_2,v_3,v_4\\ w_1,w_2,w_3}} = \begin{pmatrix} 4-\sqrt{3} & -4 & -8 & -2 \\ \sqrt{3}+1 & 8 & 9 & 2 \\ -2 & \frac{5}{2} & \frac{11}{2} & 2 \end{pmatrix}$$

- Rissunto della sesta lezione: le operazioni elementari di colonna, le matrici a scalini, il rango di una matrice.

In classe abbiamo descritto un gioco: si prende una matrice $m \times n$, a coefficienti in un campo $K$, e abbiamo a disposizione tre tipi di mosse sulle colonne. Ecco le mosse:
– si somma alla colonna $i$ la colonna $j$ moltiplicata per uno scalare $\lambda$;
– si moltiplica la colonna $s$ per uno scalare $k \neq 0$;
– si permutano fra di loro due colonne, diciamo la $i$ e la $j$.
Si è visto che è sempre possibile, usando le mosse descritte sopra, ridurre la matrice in una forma detta "a scalini (per colonne)". Per intenderci, ecco alcuni esempi di matrici in forma a scalini (sempre per colonne, vedremo presto le matrici in forma a scalini per righe):

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{3}+1 & 1 & 0 & 0 \\ -2 & \frac{5}{2} & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{3}+1 & 1 & 0 & 0 \\ -2 & 7 & 0 & 0 \\ \sqrt{3} & 4 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & \frac{5}{2} & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -8 & 4 & 1 & 0 \\ -5 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ -8 & 4 & 1 & 0 & 0 \\ -2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 & 1 \end{pmatrix}$$

Operando poi sempre con le solite operazioni di colonna (in particolare con la operazione del primo tipo, partendo dalle righe basse e risalendo) si può trasformare ancora la matrice ponendola in forma "a scalini ridotta". Ecco le forme a scalini

ridotte negli esempi appena visti:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{3}+1 & 1 & 0 & 0 \\ -2 & \frac{5}{2} & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ \sqrt{3}+1 & 1 & 0 & 0 \\ -2 & 7 & 0 & 0 \\ \sqrt{3} & 4 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2-7-7\sqrt{3} & 7 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & \frac{5}{2} & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & \frac{5}{2} & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -8 & 4 & 1 & 0 \\ -5 & 1 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 1 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ -8 & 4 & 1 & 0 & 0 \\ -2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

A questo punto in classe abbiamo dimostrato (riguardare !!), utilizzando le mosse di colonna, il seguente importante teorema, che era stato annunciato fin dalla quarta lezione:
*Teorema. Se uno spazio vettoriale $V$ ha una base di cardinalità $n \in \mathbb{Z}^+$, allora tutte le altre basi di $V$ hanno la stessa cardinalità.*

Come conseguenza, abbiamo potuto "ufficialmente" dare la seguente definizione:

*Definizione. Se uno spazio vettoriale $V$ ha una base di cardinalità $n \in \mathbb{Z}^+$ si dice che $V$ ha dimensione $n$ e si scrive anche dim $V = n$.*

Abbiamo anche osservato che ogni singola mossa sulle colonne corrisponde a moltiplicare la nostra matrice $m \times n$, a destra, per una matrice $n \times n$ invertibile. Per esempio la mossa:
– si somma alla colonna $i$ la colonna $j$ moltiplicata per uno scalare $\lambda$;
corrisponde a moltiplicare la nostra matrice per la matrice $n \times n$ che ha tutti 1 sulla diagonale, e 0 in tutte le altre caselle eccetto che nella casella identificata da "riga $j$, colonna $i$", dove troviamo $\lambda$.
Consideriamo allora una applicazione lineare:

$$L : V \to W$$

dove $V$ e $W$ sono due spazi vettoriali sul campo $K$, di dimensione $n$ e $m$ rispettivamente.
Prendiamo in $V$ una base $\{e_1, e_2, \ldots, e_n\}$ e in $W$ una base $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_m\}$.
A questo punto alla $L$ si può associare una matrice $[L]$, di forma $m \times n$, nelle basi scelte. Ora agiamo sulle colonne di $[L]$ fino a ridurla in forma a scalini ridotta: questo equivale a dire che moltiplichiamo $[L]$ a destra per tante matrici invertibili $[E_1], [E_2], \ldots, [E_k]$ fino a che $[L][E_1][E_2]\cdots[E_k]$ è in forma a scalini ridotta.

Per semplificare la notazione, chiamiamo $[E] = [E_1][E_2]\cdots, [E_k]$: sappiamo che $[E]$ è invertibile, visto che è il prodotto di matrici invertibili (la sua inversa è $[E]^{-1} = [E_k]^{-1}[E_{k-1}]^{-1}\cdots, [E_1]^{-1}$).

Ma la matrice $[E]$ è associata ad una applicazione lineare - che chiameremo $E$, come c'era da aspettarsi - da $V$ in $V$: infatti, fissata la base, in questo caso $\{e_1, e_2, \ldots, e_n\}$, sappiamo che la corrispondenza fra applicazioni lineari e matrici è bigettiva, ossia data una applicazione lineare costruiamo la matrice, data una matrice troviamo la applicazione lineare da essa rappresentata.

E chi è la applicazione lineare associata a $[L][E]$ ? È proprio la

$$L \circ E \ : \ V \to W$$

dato che il prodotto fra matrici è stato definito in modo da rispettare la composizione fra applicazioni.

A questo punto in classe abbiamo dimostrato (riguardare !!) che:

*Proposizione. Vale che $Imm\ L \circ E = Imm\ L$, ossia, scritto con un'altra notazione, $L \circ E\ (V) = L(V)$.*

Sappiamo che l'immagine di una applicazione lineare è un sottospazio vettoriale: più esattamente sappiamo che è il sottospazio vettoriale generato dalle immagini degli elementi di una base dello spazio di partenza, ossia, una volta fissate le basi, dai vettori colonna della matrice che rappresenta la applicazione.

Tradotto nel nostro caso: $Imm\ L$ è il sottospazio vettoriale di $W$ generato da $L(e_1)$, $L(e_2)$,..., $L(e_n)$, ossia dai vettori colonna di $[L]$. Ma la proposizione ci ha appena garantito che $Imm\ L = Imm\ L \circ E$, dunque $L(V)$ è anche generato dai vettori colonna di $[L][E]$.

Qui siamo arrivati al punto "vincente" del ragionamento, perché $[L][E]$ ha delle colonne "semplici" con cui lavorare, visto che è in forma a scalini ridotta. Si vede subito che le colonne non nulle di $[L][E]$ formano un insieme di vettori linearmente indipendenti, e dunque che tali colonne, siccome appunto sono linearmente indipendenti e inoltre generano $Imm\ L$, sono UNA BASE di $Imm\ L$.
Se anche avessimo ridotto la matrice $[L]$ in forma a scalini con altre mosse, e dunque magari avessimo trovato una forma a scalini diversa, rappresentata dalla matrice $[L][E']$, di nuovo lo stesso ragionamento ci porterebbe a dire che le colonne non nulle di $[L][E']$ danno UNA BASE di $Imm\ L$.
Poiché tutte le basi hanno la stessa cardinalità, con le osservazioni appena fatte siamo riusciti a dimostrare:

*Teorema. Data una applicazione lineare $L$ come sopra e fissate le basi, vale che $dim\ Imm\ L$ è uguale al numero di colonne non nulle che si trovano quando si trasforma $[L]$ in forma a scalini ridotta (o anche solo a scalini, visto che il numero di colonne non nulle è lo stesso).*

*Osservazione.* Se avessimo fissato altre basi avremmo avuto una matrice $[L]$ diversa, ma, trasformandola in forma a scalini ridotta, avremmo ancora ovviamente trovato lo stesso numero di colonne non nulle, giacché tale numero è $dim\ Imm\ L$, ossia è invariante, dipende dalla applicazione (è la dimensione della sua immagine) e non dalle basi scelte.

Vogliamo insistere sul teorema appena dimostrato, perché è importantissimo. Enunciamolo di nuovo in un altro modo, per coglierne tutto il significato.
Fissiamo le basi come sempre; sappiamo che $Imm\ L$ è il sottospazio vettoriale di $W$ generato dai vettori colonna di $[L]$. Da questi vettori, come abbiamo visto in una nota precedente ("Qualche osservazione sulle basi di spazi vettoriali"), è possibile estrarre una base di $Imm\ L$.
Dunque, facendo il conto da questo punto di vista, quanto è $dim\ Imm\ L$ ? È uguale al numero di elementi di una base di $Imm\ L$, ossia (rivedete la costruzione nella nota) è uguale al MASSIMO NUMERO DI COLONNE LINEARMENTE INDIPENDENTI di $[L]$.
La proposizione enunciata poco fa ci garantisce che $dim\ Imm\ L = dim\ Imm\ L \circ E$. Lo stesso ragionamento ci dice in realtà che:

*Proposizione. Quando consideriamo una applicazione lineare invertibile*

$$B : V \to V$$

*vale $dim\ Imm\ L = dim\ Imm\ L \circ B$. Dunque $dim\ Imm\ L$, ossia il massimo numero di colonne linearmente indipendenti di $[L]$, è uguale a $dim\ Imm\ L \circ B$, ossia al massimo numero di colonne linearmente indipendenti di $[L][B]$.*

Allora quando facciamo una mossa sulle colonne di una matrice che ha $m$ come numero massimo di colonne linearmente indipendenti, la nuova matrice avrà ancora $m$ come numero massimo di colonne linearmente indipendenti. In altre parole, fare una mossa di colonne non cambia tale numero.

Enunciamo allora di nuovo il teorema in una forma "più lunga":

*Teorema. Data una applicazione lineare $L$ come sopra e fissate le basi, vale che $dim\ Imm\ L$ è uguale al massimo numero di colonne linearmente indipendenti di $[L]$. Visto che fare mosse di colonna non cambia tale numero, vale che $dim\ Imm\ L$ è uguale anche al massimo numero di colonne linearmente indipendenti che si trovano quando si trasforma $[L]$ in forma a scalini ridotta (o anche solo a scalini). Data la particolare forma delle matrici a scalini (le cui colonne non nulle costituiscono ovviamente un insieme di vettori linearmente indipendenti), $dim\ Imm\ L$ non è altro che il numero di colonne non nulle che si trovano quando si trasforma $[L]$ in forma a scalini.*

- Quando un insieme di vettori è una base ?

Le considerazioni riassunte nella nota precedente ci permettono di discutere una importante applicazione del metodo di riduzione di una matrice in forma a scalini per colonne.

*Applicazione.* Domanda: dato uno spazio vettoriale $V$ di dimensione $n$ con una base nota $e_1, e_2, \ldots, e_n$, come si fa a stabilire se $n$ vettori dati $v_1, v_2, \ldots, v_n$ danno anch'essi una base di $V$ ?

Risposta: Mettiamo i vettori $v_1, v_2, \ldots, v_n$, che avremo espresso in termini della base $e_1, e_2, \ldots, e_n$, in colonna uno accanto all'altro. Così facendo otteniamo una matrice $[M]$ che è $n \times n$. Ora possiamo ridurre $[M]$ in forma a scalini ridotta $[M']$: se $[M']$ ha tutte le colonne non nulle (ossia è la matrice identità) allora i vettori $v_1, v_2, \ldots, v_n$ sono una base, altrimenti no.

Perché questo è vero ? Sappiamo per quanto detto sopra che il massimo numero di colonne linearmente indipendenti di $[M']$ è uguale al massimo numero di colonne linearmente indipendenti di $[M]$. Ma se $[M']$ ha degli scalini lunghi, ha per forza almeno una colonna tutta nulla: dunque il massimo numero di colonne linearmente indipendenti di $[M']$ (cioè di $[M]$ ) è $< n$. In tal caso i vettori $v_1, v_2, \ldots, v_n$ non sono linearmente indipendenti (sono "troppi", perché sono $n$....) e dunque non danno una base.
Perché diano una base, è allora necessario che la forma a scalini ridotta $[M']$ abbia scalini tutti corti ossia che $[M']$ sia la matrice identità: infatti ribadiamo che questo implica che $[M]$ ha $n$ colonne linearmente indipendenti ossia che $v_1, v_2, \ldots, v_n$ sono linearmente indipendenti.
Basta per dire che sono una base ? Resta da vedere che generano $V$. Pensiamo alla applicazione $M'$ che è associata a $[M']$: visto che $[M']$ è la matrice identità si osserva subito che $Imm\ M' = V$. Ma per le proposizioni discusse nella nota precedente $Imm\ M' = Imm\ M$, dunque $Imm\ M = V$, cioè le colonne di $[M]$ generano $V$, cioè $v_1, v_2, \ldots, v_n$ generano $V$, come volevamo.

Nelle ultime righe qui sopra abbiamo dimostrato "sottobanco" un teorema che vale la pena enunciare per ricordarlo bene:

*Teorema. In uno spazio vettoriale $V$ di dimensione $n$, se ho $n$ vettori linearmente indipendenti questi sono anche automaticamente una base di $V$.*

Facciamo ora un semplice esempio concreto di "riconoscimento" di una base. Consideriamo $\mathbb{R}^4$ con la sua base standard e poi i vettori

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, v_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Si noti che si tratta dei vettori utilizzati in un esempio alla fine della nota "Applicazioni lineari e matrici"; avrete dunque già verificato che si tratta di una base. Ma ora lo possiamo fare col nuovo metodo.

Scriviamo dunque la matrice

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

e cerchiamo di portarla in forma a scalini ridotta. Sottraendo la quarta colonna alla terza otteniamo

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Sottraendo la terza colonna alla seconda otteniamo

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e infine sottraendo la seconda colonna alla prima troviamo la matrice identità come volevamo:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Dunque $\{v_1, v_2, v_3, v_4\}$ è una base di $\mathbb{R}^4$. In questo esempio i calcoli erano particolarmente semplici, ma è già possibile notare la "convenienza" di questo metodo.

- Teorema del completamento.

Nel paragrafo precedente abbiamo dimostrato questo teorema:
*Teorema. In uno spazio vettoriale $V$ di dimensione $n$, se ho $n$ vettori linearmente indipendenti questi sono anche automaticamente una base di $V$.*
Utilizziamolo per dimostrare il teorema di completamento ad una base. La dimostrazione qui di seguito è una riscrittura di quella fatta in classe e mette bene in risalto l'uso del principio di induzione. Ci viene offerta da Giacomo Del Rio, che ringraziamo.

*Teorema. Dato uno spazio vettoriale $V$ di dimensione $n$, ogni sottoinsieme $B = \{v_1, \ldots, v_k\} \subset V$ di vettori linearmente indipendenti di cardinalità $k \leqslant n$, può essere completato ad una base di $V$ aggiungendo a $B$ $n - k$ vettori di $V \setminus Span(B)$.*

*Dimostrazione.*

Procediamo per induzione su $i = n - k$.

Passo base. Per $i = 1$, ossia per $k = n - 1$, aggiungiamo a $B$ un vettore $s \in (V \setminus Span(B))$, cioè costruiamo $B' = B \cup \{s\} = \{v_1, \ldots, v_{n-1}, s\}$.
I vettori di $B'$ sono linearmente indipendenti. Infatti

$$a_1 v_1 + \ldots + a_{n-1} v_{n-1} + a_n s = 0$$

è una combinazione lineare di coefficienti tutti nulli, dal momento che se fosse $a_n \neq 0$ si avrebbe $s = \frac{1}{a_n}(-a_1 v_1 - \ldots - a_{n-1} v_{n-1})$ che è impossibile per la scelta di $s$, viceversa con $a_n = 0$ si avrebbe $a_1 v_1 + \ldots + a_{n-1} v_{n-1} = 0$ che per l'indipendenza dei vettori di $B$ implica $a_1 = \ldots = a_{n-1} = 0$. Per il teorema ricordato sopra sappiamo che ogni sottoinsieme di vettori linearmente indipendenti di cardinalità $n$ è una base se lo spazio che li contiene ha dimensione $n$, dunque abbiamo concluso che $B'$ è una base.

Passo induttivo. Adesso supponiamo vera la tesi per $i = r$ e mostriamo che vale per $i = r + 1$. Sia $B = \{v_1, \ldots, v_k\} \subset V$ un insieme di vettori linearmente indipendenti con $k = n - (r + 1)$. In modo del tutto analogo a quanto visto per il passo base posso estendere $B$ con un vettore $s_1 \in (V \setminus Span(B))$ ed avere $B' = B \cup \{s_1\} = \{v_1, \ldots, v_k, s_1\}$ in modo tale che i vettori di $B'$ risultino linearmente indipendenti. Ora è possibile applicare l'ipotesi induttiva e completare $B'$ ad una base di $V$, provando il teorema per induzione. ∎

Abbiamo utilizzato il teorema di completamento in alcune dimostrazioni molto importanti svolte in classe, per esempio nelle dimostrazioni dei due teoremi seguenti:

(1) Formula di Grassmann per i sottospazi: dati due sottospazi $A, B$ di uno spazio vettoriale $V$ di dimensione finita, vale

$$dim\ A + dim\ B = dim\ A \cap B + dim\ (A + B)$$

(2) Considerata una applicazione lineare $L : V \to W$ dove $V$ e $W$ sono spazi vettoriali e $V$ ha dimensione finita, vale

$$dim\ Ker\ L + dim\ Imm\ L = dim\ V$$

- Autovalori e autovettori di un endomorfismo lineare.

Sia $T : V \to V$ una applicazione lineare da uno spazio vettoriale $V$ (sul campo $K$) in sé. Una tale $T$ è chiamata anche endomorfismo lineare.

*Definizione. Un vettore $v \in V - \{O\}$ si dice un autovettore di $T$ se*

$$T(v) = \lambda v$$

*per un certo $\lambda \in K$.*

In altre parole un autovettore di $T$ è un vettore NON NULLO dello spazio $V$ che ha la seguente proprietà: la T lo manda in un multiplo di sé stesso.

*Definizione. Se $v \in V - \{O\}$ è un autovettore di $T$ tale che*

$$T(v) = \lambda v$$

*allora lo scalare $\lambda \in K$ si dice autovalore di $T$ relativo a $v$ (e viceversa si dirà che $v$ è un autovettore relativo all'autovalore $\lambda$).*

Si noti che è l'autovettore che non può essere $O \in V$, mentre l'autovalore può benissimo essere $0 \in K$: se per esempio $T$ non è iniettiva, ossia $Ker\ T \supsetneq \{0\}$, tutti gli elementi $w \in (Ker\ T) - \{O\}$ soddisfano

$$T(w) = O = 0\ w$$

ossia sono autovettori relativi all'autovalore 0.

*Definizione. Dato $\lambda \in K$ chiamiamo l'insieme*

$$V_\lambda = \{v \in V \mid T(v) = \lambda v\}$$

*"autospazio relativo a $\lambda$".*

Si verifica facilmente che un autospazio $V_\lambda$ è un sottospazio vettoriale di $V$.
Anche se abbiamo definito l' autospazio $V_\lambda$ per qualunque $\lambda \in K$, in realtà $V_\lambda$ è sempre uguale a $\{O\}$ a meno che $\lambda$ non sia un autovalore. Questo è dunque il caso interessante: se $\lambda$ è un autovalore di $T$ allora $V_\lambda$ è costituito da tutti gli autovettori relativi a $\lambda$ più lo $O$. In particolare abbiamo notato poco fa che $V_0 = Ker\ T$.

Perché per noi sono importanti autovettori e autovalori ?

Supponiamo che $V$ abbia dimensione $n$ e pensiamo a cosa succederebbe se riuscissimo a trovare una base di $V$, $\{v_1, v_2, \ldots, v_n\}$, composta solo da autovettori di $T$.

Avremmo, per ogni $i = 1, 2, \ldots, n$,

$$T(v_i) = \lambda_i v_i$$

per certi autovalori $\lambda_i$ (sui quali non sappiamo nulla, sono semplicemente elementi di $K$, addirittura potrebbero anche essere tutti uguali $\lambda_1 = \cdots = \lambda_n$).

Come sarebbe fatta la matrice

$$[T]\ {\scriptstyle \{v_1, v_2, \ldots, v_n\} \atop \scriptstyle \{v_1, v_2, \ldots, v_n\}}$$

associata a $T$ rispetto a questa base ?

Ricordandosi come si costruiscono le matrici osserviamo che la prima colonna conterrebbe il vettore $T(v_1)$ scritto in termini della base $\{v_1, v_2, \ldots, v_n\}$, ossia

$$T(v_1) = \lambda_1 v_1 + 0v_2 + 0v_3 + 0v_4 + \cdots + 0v_n$$

la seconda il vettore $T(v_2) = 0v_1 + \lambda_2 v_2 + 0v_3 + \cdots + 0v_n$ e così via. Quindi la matrice sarebbe diagonale:

$$[T]_{\substack{\{v_1, v_2, \ldots, v_n\} \\ \{v_1, v_2, \ldots, v_n\}}} = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_{n-1} & 0 \\ 0 & 0 & 0 & \ldots & 0 & \lambda_n \end{pmatrix}$$

Ora, una matrice diagonale è per noi "leggibilissima"; solo guardandola possiamo sapere tutto di $T$: il suo rango (dunque anche la dimensione del nucleo), quali sono esattamente i vettori di $Ker\ T$, quali sono i sottospazi in cui $T$ si comporta come l'identità, ossia i sottospazi costituiti dai vettori di $V$ che la $T$ lascia fissi...

Dunque studiamo gli autovalori e gli autovettori di $T$ nella speranza di trovare "basi buone" che ci permettano di conoscere bene il comportamento di $T$.

Ma esistono sempre queste "basi buone", ossia basi costituite solo da autovettori di $T$ ? NO, non sempre. Se per una certa $T$ esiste una base buona si dice che $T$ è "diagonalizzabile", altrimenti $T$ è "non diagonalizzabile".

Il nostro primo compito è allora quello di cercare di scoprire come possiamo trovare tutti gli autovalori e gli autovettori di $T$ e come possiamo poi decidere se esite per $T$ una base buona.

La nostra strategia sarà la seguente:

- PASSO 1. Data $T$, troviamo gli autovalori di $T$ utilizzando il polinomio caratteristico.

- PASSO 2. Supponiamo di aver trovato gli autovalori $\lambda_1, \lambda_2, \ldots, \lambda_k$: a questo punto scopriamo chi sono i relativi autospazi $V_{\lambda_1}, V_{\lambda_2}, \ldots, V_{\lambda_k}$.

- PASSO 3. Un teorema ci assicurerà che gli autospazi $V_{\lambda_1}, V_{\lambda_2}, \ldots, V_{\lambda_k}$ sono in somma diretta, ossia che vale

  $$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$$

  Quindi se

  $$dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots + dim\ V_{\lambda_k} = n = dim\ V$$

  allora è possibile trovare una base "buona", fatta da autovettori di $T$ e $T$ è diagonalizzabile. Per scrivere una base "buona" basta scegliere una base per ogni $V_{\lambda_i}$ e poi fare l'unione. Altrimenti, se

  $$dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots + dim\ V_{\lambda_k} < n = dim\ V$$

  $T$ non è diagonalizzabile.

- PASSO 4. Se $T$ è risultata diagonalizzabile, usando la base trovata si scrive la matrice diagonale $[T]$.

Vediamo i dettagli passo per passo.

PASSO 1.

Dato il nostro endomorfismo $T : V \to V$ e posto $n = dim\ V$ vogliamo trovare un criterio per decidere se uno scalare $\lambda \in K$ è o no un autovalore di $T$.

Perché sia una autovalore, secondo la definizione bisogna che esista un $v \in V - \{O\}$ tale che

$$T(v) = \lambda v$$

Questo si può riscrivere anche come

$$T(v) - \lambda I(v) = O$$

dove $I : V \to V$ è l'identità. Riscriviamo ancora:

$$(T - \lambda I)(v) = O$$

Abbiamo "scoperto" che la applicazione lineare $T - \lambda I$ non è iniettiva: infatti manda il vettore $v$ in $O$. Dunque, se scegliamo una base qualunque per $V$ e costruiamo la matrice $[T]$ associata a $T$, la matrice $[T] - \lambda[I]$ dovrà avere determinante uguale a 0: $det([T] - \lambda[I]) = 0 = det(\lambda[I] - [T])$. A dire la verità scrivere $[I]$ è un formalismo inutile: in qualunque base la si scriva, la matrice associata all'identità è comunque diagonale con tutti i coefficienti diagonali uguali a 1, quindi da ora in poi ometteremo le parentesi quadre e scriveremo semplicemente $I$.

Questa osservazione ci avvicina alla seguente definizione:

*Definizione. Dato un endomorfismo lineare $T : V \to V$ con $n = dim\ V$, scegliamo una base per $V$ e costruiamo la matrice $[T]$ associata a $T$ rispetto a tale base. Ora definiamo il polinomio caratteristico di $T$ $p_T(t) \in K[t]$ tramite la seguente equazione:*

$$p_T(t) \equiv det(t[I] - [T])$$

*Ossevazioni importanti.*

1) Perché la definizione data abbia senso innanzitutto bisogna verificare che $det(t[I] - [T])$ è veramente un polinomio. Questo si può dimostrare facilmente per induzione sulla dimensione $n$ di $V$. Sempre per induzione si può dimostrare un po' di più, ossia che $p_T(t)$ è un polinomio di grado $n$ con coefficiente direttivo 1: $p_T(t) = t^n + \ldots\ldots\ldots$ Queste dimostrazioni sono facoltative e consigliate !

2) Inoltre ci interessa molto che la definizione appena data non dipenda dalla base scelta di $V$: non sarebbe una definizione tanto buona se con la scelta di due basi diverse ottenessimo due polinomi caratteristici diversi !

Questo problema per fortuna non si verifica. In generale vale il fatto seguente: dato l'endomorfismo $A : V \to V$, scegliamo una base $b$ di $V$ e costruiamo la matrice

$[A]_{\substack{b\\b}}$. Poi scegliamo un'altra base $b'$ e costruiamo $[A]_{\substack{b'\\b'}}$. Le due matrici $[A]_{\substack{b\\b}}$ e $[A]_{\substack{b'\\b'}}$ sono legate dalla seguente relazione: esiste una matrice $[B]$ invertibile tale che

$$[B][A]_{\substack{b\\b}}[B]^{-1} = [A]_{\substack{b'\\b'}}$$

e quindi anche

$$[B]^{-1}[A]_{\substack{b'\\b'}}[B] = [A]_{\substack{b\\b}}$$

Lasciamo facoltativa la dimostrazione di quanto appena detto. Usando il teorema di Binet a questo punto sappiamo verificare che

$$det\ [A]_{\substack{b'\\b'}} = det[A]_{\substack{b\\b}}$$

Tornando al caso del nostro polinomio caratteristico, possiamo dunque stare sicuri che $p_T(t) = det\ (tI - [T])$ non dipende dalla scelta della base.

Assicuratici che la definizione è ben data, enunciamo il teorema principale che speiga l'utilità del polinomio caratteristico.

*Teorema. Considerata $T$ come sopra, vale che uno scalare $\lambda \in K$ è un autovalore di $T$ se e solo se $\lambda$ è una radice di $p_T(t)$, ossia se e solo se $p_T(\lambda) = 0$.*

*Dimostrazione.* Abbiamo già visto (l'osservazione prima della definizione del polinomio caratteristico) che se $\lambda$ è un autovalore di $T$ allora $det(\lambda I - [T]) = p_T(\lambda) = 0$. Resta dunque da dimostrare il viceversa. Supponiamo che $det(\lambda I - [T]) = p_T(\lambda) = 0$: allora l'applicazione lineare $\lambda I - T$ non è iniettiva. dunque esiste $v \in V - \{O\}$ tale che $(\lambda I - T)(v) = 0$, che si può riscrivere anche come

$$T(v) = \lambda v$$

dunque abbiamo trovato un autovettore che ha autovalore $\lambda$ e quindi abbiamo mostrato che, come volevamo, $\lambda$ è una autovalore di $T$. ∎

Questo conclude il passo 1: per sapere quali sono gli autovalori di un endomorfismo $T$ basta calcolare il polinomio caratteristico $p_T(t)$ e trovare le sue radici in $K$.

PASSO 2.

Supponiamo dunque di aver scoperto che $T$ ha i seguenti autovalori: $\lambda_1, \lambda_2, \ldots, \lambda_k$, tutti distinti fra loro. Vogliamo capire chi sono gli autospazi $V_{\lambda_1}, V_{\lambda_2}, \ldots V_{\lambda_k}$.

Per questo basterà risolvere dei sistemi lineari: per ogni $i = 1, 2, \ldots, k$, l'autospazio $V_{\lambda_i}$ è costituito per definizione dai vettori $v \in V$ tali che $T(v) = \lambda_i v$, ossia, scelta una base di $V$ e dunque trovata la matrice $[T]$, dalle soluzioni del sistema lineare

$$([T] - \lambda_i I) \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ \cdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ \cdots \\ 0 \\ 0 \end{pmatrix}$$

PASSO 3.

Qui cominciamo col dimostrare il seguente teorema.

*Teorema. Dato un endomorfismo lineare $T : V \to V$, siano $\lambda_1, \lambda_2, \ldots, \lambda_k$ degli autovalori di $T$ distinti fra loro. Consideriamo ora degli autovettori $v_1 \in V_{\lambda_1}, v_2 \in V_{\lambda_2}, \ldots, v_k \in V_{\lambda_k}$. Allora $\{v_1, v_2, \ldots, v_k\}$ è un insieme di vettori linearmente indipendenti.*

*Osservazione. Spesso ci si riferisce a questo teorema con la frase: "autovettori relativi ad autovalori distinti sono linearmente indipendenti".*

*Dimostrazione.* Procediamo per induzione su $k$. Per $k = 1$ l'enunciato è banale perché $\{v_1\}$ è banalmente un insieme di vettori linearmente indipendenti (c'è un vettore solo..). Supponiamo di aver dimostrato che l'enunciato è vero fino a $k - 1$ e cerchiamo di dimostrarlo per $k$. Supponiamo allora che valga:

$$a_1 v_1 + a_2 v_2 + \cdots + a_k v_k = O$$

Per mostrare che $\{v_1, v_2, \ldots, v_k\}$ è un insieme di vettori linearmente indipendenti dobbiamo mostrare che questo può accadere solo quando $a_1 = a_2 = \cdots = a_k = 0$.

Dalla equazione scritta ne ricaviamo due in due modi diversi. Prima applichiamo $T$ ad entrambi i membri e per linearità ricaviamo

$$a_1 T(v_1) + a_2 T(v_2) + \cdots + a_k T(v_k) = O$$

che svolgendo il calcolo diventa

$$a_1 \lambda_1 v_1 + a_2 \lambda_2 v_2 + \cdots + a_k \lambda_k v_k = O$$

Poi invece moltiplichiamo l'equazione per $\lambda_k$ ottenendo:

$$a_1 \lambda_k v_1 + a_2 \lambda_k v_2 + \cdots + a_k \lambda_k v_k = O$$

Per sottrazione da queste due equazioni ricaviamo:

$$a_1(\lambda_k - \lambda_1)v_1 + a_2(\lambda_k - \lambda_2)v_2 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1} = O$$

Ma questa è una combinazione lineare dei $k - 1$ vettori $\{v_1, v_2, \ldots, v_{k-1}\}$ uguale a $O$: per ipotesi induttiva tutti i coefficienti devono essere uguali a 0. Visto che gli scalari $\lambda_k - \lambda_i$ sono tutti diversi da zero (gli autovalori in questione sono distinti fra

loro per ipotesi) questo implica che $a_1 = a_2 = \cdots = a_{k-1} = 0$. Sostituendo nella equazione iniziale, notiamo che deve essere anche $a_k = 0$. ∎

Il seguente teorema è un rafforzamento del precedente.

*Teorema. Dato un endomorfismo lineare $T : V \to V$, siano $\lambda_1, \lambda_2, \ldots, \lambda_k$ degli autovalori di $T$ distinti fra loro. Gli autospazi $V_{\lambda_1}, V_{\lambda_2}, \ldots, V_{\lambda_k}$ sono in somma diretta, ossia che vale*

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$$

*Dimostrazione.* Dire che gli autospazi $V_{\lambda_1}, V_{\lambda_2}, \ldots, V_{\lambda_k}$ sono in somma diretta vuol dire che se ne prendo uno qualunque, diciamo $V_{\lambda_1}$ tanto per fissare la notazione, la sua intersezione con la somma di tutti gli altri è banale, ossia

$$V_{\lambda_1} \cap (V_{\lambda_2} + \cdots + V_{\lambda_k}) = \{O\}$$

Supponiamo per assurdo che non sia così, e che ci sia un vettore $w \neq O$ tale che

$$w \in V_{\lambda_1} \cap (V_{\lambda_2} + \cdots + V_{\lambda_k})$$

Allora possiamo scrivere $w$ in due modi:

$$w = v_1 \in V_{\lambda_1} - \{O\}$$

perché $w \in V_{\lambda_1}$, e

$$w = a_2 v_2 + a_3 v_3 + \cdots + a_k v_k$$

(dove gli $a_j$ sono scalari e i $v_j \in V_{\lambda_j}$ per ogni $j$) visto che $w \in V_{\lambda_2} + \cdots + V_{\lambda_k}$.

Dunque vale:

$$v_1 = w = a_2 v_2 + a_3 v_3 + \cdots + a_k v_k$$

ossia

$$v_1 - a_2 v_2 - a_3 v_3 - \cdots - a_k v_k = O$$

Questa è una combinazione lineare di autovettori relativi ad autovalori distinti, ma non è la combinazione lineare banale (infatti almeno il coefficiente di $v_1$ è 1 dunque non nullo). Dunque tali autovettori sarebbero linearmente dipendenti, assurdo perché contraddice il teorema precedente. ∎

Nelle ipotesi del teorema qui sopra, ci resta allora da osservare che, siccome

$$dim\ (V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}) = dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots + dim\ V_{\lambda_k}$$

(questo si può dimostrare per induzione applicando la formula di Grassmann: esercizio !!) allora abbiamo un criterio per dire se $T$ è diagonalizzabile o no. Infatti, se

$$dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots + dim\ V_{\lambda_k} = n = dim\ V$$

allora per questioni dimensionali deve essere

$$V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k} = V$$

e quindi è possibile trovare una base "buona", fatta da autovettori di $T$, insomma $T$ è diagonalizzabile.

Per scrivere una simile base "buona" basta scegliere una base per ogni $V_{\lambda_i}$ e poi fare

l'unione. Tale unione è una base di tutto $V$: infatti i vettori di tale base generano $V$ perché $V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$ e sono proprio $n$.

Altrimenti, se

$$dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots dim\ V_{\lambda_k} < n = dim\ V$$

$T$ non è diagonalizzabile. Infatti non è possibile trovare una base di autovettori: se la trovassimo contraddiremmo

$$dim\ V_{\lambda_1} + dim\ V_{\lambda_2} + \cdots dim\ V_{\lambda_k} < n = dim\ V$$

PASSO 4.

Se l'endomorfismo $T$ è diagonalizzabile, scegliamo dunque una base di autovettori nel modo descritto al passo 3, e avremo una matrice associata $[T]$ che risulterà diagonale. Manteniamo le notazioni introdotte al passo 3: allora sulla diagonale troveremo $dim\ V_{\lambda_1}$ coefficienti uguali a $\lambda_1$, $dim\ V_{\lambda_2}$ coefficienti uguali a $\lambda_2$, ...,$dim\ V_{\lambda_k}$ coefficienti uguali a $\lambda_k$. Il rango di $T$ sarà dunque uguale al numero dei coefficienti non zero che troviamo sulla diagonale, la dimensione del nucleo sarà uguale al numero dei coefficienti uguali a zero che troviamo sulla diagonale. In altre parole, se 0 non è un autovalore di $T$, allora $Ker\ T = \{O\}$, come avevamo già osservato in precedenza; se invece 0 è un autovalore di $T$ allora troveremo sulla diagonale $dim\ V_0$ coefficienti uguali a 0 - d'altronde avevamo già notato che $V_0 = Ker\ T$.