## 5. Polinomi e matrici

Sia $V$ uno spazio vettoriale sul campo $K$, e sia $f$ un suo endomorfismo. Sia poi $P(X) = \sum_i a_i X^i$ un polinomio a coefficienti in $K$. Poniamo

$$P(f) = \sum_i a_i f^i \,,$$

dove $f^i$ sta per la composizione di $f$ con se stesso $i$ volte. Allo stesso modo, se $A$ è una matrice $n \times n$ a coefficienti in $K$, porremo

$$P(A) = \sum_i a_i A^i \,.$$

Se $V$ ha dimensione finita e $A$ è la matrice di $f$ rispetto a una sua base, allora $P(A)$ è la matrice di $P(f)$ rispetto a questa stessa base. Ciò segue dal fatto che la composizione di applicazioni lineari corrisponde al prodotto di matrici. Sia ora $Q(X)$ un altro polinomio. Valgono le seguenti proprietà

$$\begin{aligned} P(f) + Q(f) &= (P+Q)(f) \,, \\ P(f)Q(f) &= PQ(f) \,. \end{aligned}$$

Solo la seconda merita un cenno di dimostrazione. Scriviamo $Q(X) = \sum_i b_i X^i$, cosicché

$$PQ(X) = \sum_i \left( \sum_{h+k=i} a_h b_k \right) X^i \,.$$

D'altra parte

$$\begin{aligned} P(f)Q(f) &= (\sum_h a_h f^h)(\sum_k b_k f^k) \\ &= \sum_i \left( \sum_{h+k=i} a_h f^h b_k f^k \right) \\ &= \sum_i \left( \sum_{h+k=i} a_h b_k \right) f^i \,, \end{aligned}$$

come si voleva. Naturalmente proprietà analoghe a quelle appena dimostrate valgono quando al posto di $f$ vi è una matrice quadrata.

TEOREMA (5.1) (CAYLEY-HAMILTON). *Sia $V$ uno spazio vettoriale di dimensione finita sul campo $K$, sia $f : V \to V$ una applicazione lineare, e sia $P(X)$ il suo polinomio caratteristico. Allora $P(f) = 0$.*

Scegliamo una base di $V$, e sia $A$ la matrice di $f$ rispetto a questa base. Dunque $P(X) = \det(XI - A)$, e dobbiamo dimostrare che $P(A) = 0$. Siamo naturalmente liberi di cambiare la base o, che è lo stesso, di rimpiazzare la matrice $A$ con una sua coniugata, se necessario. La dimostrazione è per induzione su $n$. Il teorema è sicuramente vero per $n = 1$, in quanto in questo caso $A$ è uno scalare $a$ e $P(X) = X - a$. Per $n > 1$ ci baseremo sul seguente semplice risultato algebrico, che enunciamo senza dimostrazione.

LEMMA (5.2). *Esiste un campo $F$ di cui $K$ è sottocampo (esiste cioè una estensione $F$ di $K$) tale che $P$ abbia una radice in $F$.*

Possiamo considerare la matrice $A$ come matrice a elementi in $F$, senza che questo cambi il suo polinomio caratteristico. L'annullarsi o meno di $P(A)$ è inoltre indipendente dal fatto che si lavori sul campo $K$ o sul campo $F$. Possiamo dunque supporre che $P$ abbia una radice $\lambda$ già in $K$. Scegliamo la base di $V$ in modo che il suo primo elemento sia un autovettore per $\lambda$. Allora

$$A = \begin{pmatrix} \lambda & C \\ 0 & B \end{pmatrix},$$

dove $B$ è un blocco $(n-1) \times (n-1)$ e $C$ un blocco $1 \times (n-1)$. Applicando la regola di Laplace e sviluppando il determinante rispetto alla prima colonna si ottiene

$$P(X) = \det(XI - A) = (X - \lambda) \det(XI - B)\,.$$

In altre parole, il polinomio caratteristico di $A$ è il prodotto di $X - \lambda$ e del polinomio caratteristico di $B$, che indichiamo con $Q(X)$. Per ipotesi induttiva sappiamo che $Q(B) = 0$. È immediato mostrare che

$$\begin{pmatrix} \lambda & C \\ 0 & B \end{pmatrix}^k = \begin{pmatrix} \lambda^k & D \\ 0 & B^k \end{pmatrix},$$

dove $D$ è una matrice $1 \times (n-1)$. Ne segue che

$$A - \lambda I = \begin{pmatrix} 0 & E \\ 0 & G \end{pmatrix}, \qquad Q(A) = \begin{pmatrix} Q(\lambda) & M \\ 0 & Q(B) \end{pmatrix} = \begin{pmatrix} Q(\lambda) & M \\ 0 & 0 \end{pmatrix},$$

dove $G$ è una matrice $(n-1) \times (n-1)$ e $E$, $M$ sono matrici $1 \times (n-1)$. Dunque

$$P(A) = (A - \lambda I)Q(A) = \begin{pmatrix} 0 & E \\ 0 & G \end{pmatrix} \begin{pmatrix} Q(\lambda) & M \\ 0 & 0 \end{pmatrix} = 0\,,$$

come si doveva dimostrare.

Nel resto di questa sezione supporremo sempre che $V$ abbia dimensione finita. Il teorema di Cayley-Hamilton asserisce tra l'altro l'esistenza di un polinomio non nullo a coefficienti in $K$ che si annulla se valutato su $f$. Tra tutti i polinomi non nulli con questa proprietà consideriamo quelli di grado minimo. Questi polinomi sono tutti proporzionali fra loro, poiché in caso contrario si potrebbe costruire per combinazione lineare un polinomio

non nullo di grado più basso che si annulla se valutato su $f$; vi è dunque tra loro un solo polinomio monico, che si chiama *polinomio minimo* di $f$. Allo stesso modo si definisce il polinomio minimo di una matrice quadrata. È chiaro che, se $A$ è la matrice di $f$ rispetto a una qualche base, i polinomi minimi di $f$ e di $A$ coincidono.

Sia ora $Q$ il polinomio minimo di $f$, e sia $G$ un polinomio a coefficienti in $K$ tale che $G(f) = 0$. Vogliamo mostrare che esiste un altro polinomio $H$ a coefficienti in $K$ tale che $G = HQ$. Infatti se dividiamo $G$ per $Q$ possiamo scrivere

$$G = HQ + R\,,$$

dove $R$ è nullo o ha grado più basso di quello di $Q$. È chiaro che $R(f) = 0$; quindi, se $R$ non fosse nullo, $Q$ non potrebbe essere il polinomio minimo di $f$.

PROPOSIZIONE (5.3). *Sia $V$ uno spazio vettoriale di dimensione finita su $K$, sia $f$ un suo endomorfismo, e indichiamo con $Q$ il polinomio minimo di $f$. Allora:*

*i) ogni autovalore di $f$ è radice di $Q$;*

*ii) $f$ è diagonalizzabile (su $K$) se e solo se $Q$ ha $\deg(Q)$ radici distinte in $K$.*

Sia $\lambda$ un autovalore di $f$; dunque c'è un vettore non nullo $v \in V$ tale che $f(v) = \lambda v$. Ne segue che $f^i(v) = \lambda^i v$ per ogni $i$, e quindi che $H(f)(v) = H(\lambda)v$ per ogni polinomio $H$. In particolare $0 = Q(f)(v) = Q(\lambda)v$, e come conseguenza $Q(\lambda) = 0$. Supponiamo ora che $f$ sia diagonalizzabile, cioè che $V$ abbia una base $v_1, \dots, v_n$ costituita da autovettori. Riordinando questa base possiamo supporre che

$$f(v_i) = \begin{cases} \lambda_1 v_i & \text{se } 0 = n_0 < i \le n_1 \\ \lambda_2 v_i & \text{se } n_1 < i \le n_2 \\ \lambda_3 v_i & \text{se } n_2 < i \le n_3 \\ \dots & \\ \lambda_h v_i & \text{se } n_{h-1} < i \le n_h = n \end{cases}$$

dove $\lambda_1, \dots, \lambda_h$ sono distinti. Dico che $Q(X) = \prod_i (X - \lambda_i)$. Dato che, come si è visto, ognuno dei $\lambda_i$ è radice di $Q$, basta mostrare che $H(f) = 0$, dove $H(X) = \prod_i (X - \lambda_i)$. In effetti, se $j$ è un intero compreso tra 1 e $n$, e $i$ è scelto in modo che $n_{i-1} < j \le n_i$, allora $f(v_j) = \lambda_i v_j$ e quindi

$$H(f)(v_j) = \prod_{k \neq i} (f - \lambda_k \mathbf{1})(f(v_j) - \lambda_i v_j) = 0\,.$$

Poiché $v_j$ è un elemento arbitrario della base scelta per $V$, $H(f)$ deve essere nullo.

Per completare la dimostrazione di (5.3) resta da mostrare che, se $Q(X) = \prod_i (X - \lambda_i)$, dove i $\lambda_i$ sono distinti, allora $f$ è diagonalizzabile. Ragioneremo per induzione sul grado di $Q$. Se questo è 1, $f$ è una omotetia, e quindi diagonalizzabile. Descriviamo ora il passo induttivo. Scriviamo $Q(X) = (X - \lambda_1)R(X)$, dove $R(X) = \prod_{k>1} (X - \lambda_k)$. Osserviamo che, poiché $\lambda_1$ non è una radice di $R$, dividendo $R$ per $(X - \lambda_1)$ si può scrivere

$$R(X) = C(X)(X - \lambda_1) + c\,,$$

dove $c$ è una costante non nulla. Dividendo per $c$ se ne ricava che ci sono polinomi $A$ e $B$ tali che

$$(X-\lambda_1)A(X)+R(X)B(X)=1\,. \tag{5.4}$$

Sia ora $v$ un elemento di $V$. Da (5.4) si deduce che

$$v=(f-\lambda_1\mathbf{1})A(f)(v)+R(f)B(f)(v)\,.$$

Si noti che $(f-\lambda_1\mathbf{1})R(f)B(f)(v)=Q(f)B(f)(v)=0$ e che $R(f)(f-\lambda_1\mathbf{1})A(f)(v)=Q(f)A(f)(v)=0$. Ciò mostra che $v$ è somma di un elemento di $V_1=\ker(R(f))$ e di uno di $V_2=\ker(f-\lambda_1\mathbf{1})$; in altre parole, $V$ è somma di $V_1$ e $V_2$. Questa somma è diretta. In effetti, se $v\in V_1\cap V_2$, allora da (5.4) si ricava che

$$v=A(f)(f(v)-\lambda_1 v)+B(f)R(f)(v)=0\,.$$

Osserviamo ora che $fR(f)(w)=R(f)(f(w))$ per ogni $w\in V$, e quindi in particolare $f(w)\in\ker R(f)$ ogni volta che $w\in\ker R(f)$. In altri termini, se conveniamo di chiamare *invariante* per $f$ un sottospazio $W$ di $V$ nel caso in cui $f(W)\subset W$, il sottospazio $V_1=\ker R(f)$ è invariante. Ne segue che $f$ induce per restrizione un endomorfismo $g$ di $V_1$. Il polinomio minimo di $g$ è $R(X)$. Per ipotesi induttiva $g$ è diagonalizzabile, cioè vi è una base di $V_1$ costituita da autovettori di $g$, e quindi di $f$. Una base di $V$ costituita da autovettori si ottiene aggiungendo a questa una base di $V_2$, che è l'autospazio di $f$ relativo a $\lambda_1$. La dimostrazione di (5.3) è completa.

Ricordiamo che due polinomi si dicono primi fra loro se non vi sono polinomi, oltre alle costanti, che li dividano entrambi.

Proposizione (5.5). *Sia $V$ uno spazio vettoriale di dimensione finita su $K$ e sia $f$ un suo endomorfismo. Siano $P_1,\dots,P_h$ polinomi a due a due primi fra loro tali che $\prod P_i(f)=0$. Allora $V$ è somma diretta dei sottospazi $V_i=\ker(P_i(f))$, $i=1,\dots,h$.*

Questa proposizione e la sua dimostrazione generalizzano la parte finale della dimostrazione di (5.3). Iniziamo dal caso $h=2$. Faremo uso del seguente semplice lemma algebrico.

Lemma (5.6). *Siano $P_1$ e $P_2$ polinomi a coefficienti in $K$. Se $P_1$ e $P_2$ sono primi fra loro esistono polinomi $A$ e $B$ a coefficienti in $K$ tali che*

$$A(X)P_1(X)+B(X)P_2(X)=1\,.$$

Osserviamo, per inciso, che (5.4) è un caso particolare di questo lemma. Per dimostrare (5.6) indichiamo con $H$ un polinomio non nullo di grado minimo tra quelli della forma $MP_1+NP_2$, dove $M$ e $N$ sono polinomi. Possiamo scrivere $P_1(X)=C(X)H(X)+R(X)$, dove $R$ è nullo o ha grado minore di quello di $H$. Quindi $R=(1-CM)P_1-CNP_2$. Per la minimalità del grado di $H$, $R$ deve essere nullo. Dunque $H$ divide $P_1$, e lo stesso

ragionamento mostra che divide $P_2$. Poiché $P_1$ e $P_2$ sono primi fra loro, ne segue che $H$ è costante. Dividendo per questa costante si ottiene la relazione cercata.

Torniamo alla dimostrazione di (5.5). Se $v \in V$ segue da (5.6) che

$$v = \mathbf{1}(v) = P_2(f)B(f)(v) + P_1(f)A(f)(v)\,.$$

Dato che $P_1(f)P_2(f)B(f)(v) = 0$ e $P_2(f)P_1(f)A(f)(v) = 0$, ciò esprime $v$ come somma di un elemento di $V_1$ e di uno di $V_2$. Se poi $v$ appartiene a $V_1 \cap V_2$ possiamo scrivere, sempre usando (5.6):

$$v = \mathbf{1}(v) = B(f)P_2(f)(v) + A(f)P_1(f)(v) = 0\,.$$

Ciò mostra che $V = V_1 \oplus V_2$, completando la dimostrazione nel caso $h = 2$.

Per $h > 2$ si procede per induzione su $h$. Faremo uso di un altro lemma algebrico.

LEMMA (5.7). *Siano $P_1, \dots, P_h$ polinomi a coefficienti in $K$. Se $P_1$ è primo con $P_2, \dots, P_h$, allora $P_1$ è primo con $Q = P_2 \cdots P_h$.*

Ragioniamo per assurdo. Se quanto affermato dal lemma non fosse vero potremmo trovare polinomi di grado positivo che sono divisori sia di $P_1$ che di $Q$. Sia $H$ un polinomio di grado minimo tra questi; in particolare $H$ non ha divisori di grado minore di quello di $H$ a parte le costanti. Dato che $P_1$ e $P_2$ sono primi fra loro, ciò implica che $H$ deve essere primo con $P_2$, e quindi, per (5.6), che si può scrivere

$$1 = AH + BP_2\,,$$

da cui

$$\prod_{i\geq 3} P_i = \prod_{i\geq 3} P_i AH + BQ\,.$$

Dato che $H$ divide $Q$ se ne deduce che divide anche $P_3 \cdots P_h$. Iterando questo ragionamento si giunge alla conclusione che $H$ divide $P_h$, il che è assurdo visto che $P_1$ e $P_h$ sono primi fra loro.

Torniamo alla dimostrazione di (5.5). Il lemma che abbiamo appena dimostrato dice che $P_1$ e $Q = P_2 \cdots P_h$ sono primi fra loro. Usando il caso $h = 2$ della proposizione ne deduciamo che $V$ è somma diretta di $V_1$ e del nucleo di $Q(f)$. Osserviamo che $fQ(f) = Q(f)f$, e che quindi $f(\ker Q(f)) \subset \ker Q(f)$. Ne segue che $f$ induce per restrizione un endomorfismo $g$ di $\ker Q(f)$. Per la definizione di $\ker Q(f)$ si ha che $Q(g) = 0$, e quindi, per ipotesi induttiva, se ne ricava che

$$\ker Q(f) = \bigoplus_{i\geq 2} \ker P_i(g)\,,$$

o anche che

$$V = V_1 \oplus \ker Q(f) = V_1 \oplus \bigoplus_{i\geq 2} \ker P_i(g)\,.$$

Per concludere basta osservare che, per $i \geq 2$,

$$\ker P_i(g) = \ker Q(f) \cap \ker P_i(f) = \ker P_i(f)\,,$$

dato che $P_i$ divide $Q$ e quindi $\ker P_i(f) \subset \ker Q(f)$. La dimostrazione di (5.5) è ora completa.

COROLLARIO (5.8). *Sia $V$ uno spazio vettoriale di dimensione finita su $K$ e sia $f$ un suo endomorfismo. Supponiamo che il polinomio caratteristico $P$ di $f$ si decomponga completamente in fattori lineari, e scriviamo*

$$P(X) = \prod_{i=1}^{h} (X - \lambda_i)^{\mu_i} ,$$

*dove i $\lambda_i$ sono distinti. Allora $V$ è somma diretta dei sottospazi $V_i = \ker(f - \lambda_i \mathbf{1})^{\mu_i}$, $i = 1, \dots, h$. Inoltre la dimensione di $V_i$ è $\mu_i$.*

Dato che $P(f) = 0$ per il teorema di Cayley-Hamilton, e che $(X - \lambda_i)^{\mu_i}$ e $(X - \lambda_j)^{\mu_j}$ sono primi fra loro se $i \neq j$, è necessario dimostrare solo l'asserzione sulle dimensioni dei $V_1$. Scegliamo una base $v_1, \dots, v_{n_1}, v_{n_1+1}, \dots, v_{n_2}, \dots$ per $V$ in modo che $v_1, \dots, v_{n_1}$ sia una base di $V_1$, $v_{n_1+1}, \dots, v_{n_2}$ una base di $V_2$, e così via. Dato che $fP_i(f) = P_i(f)f$, e quindi $f(V_i) \subset V_i$, per ogni $i$, la matrice di $f$ rispetto a questa base è diagonale a blocchi ed ha sulla diagonale blocchi di dimensioni $n_1 = \dim V_1$, $n_2 - n_1 = \dim V_2$, $n_3 - n_2 = \dim V_3$, e così via. Inoltre $f$ induce per restrizione un endomorfismo $f_i$ di $V_i$. Le matrici di questi endomorfismi non sono altro che i blocchi diagonali della matrice di $f$. Ne segue in particolare che $P(X) = \prod_i P_i(X)$, dove $P_i$ è il polinomio caratteristico di $f_i$. Si ha che $P_i(X) = \det(X\mathbf{1} - f_i) = \det((X - \lambda_i)\mathbf{1} - (f_i - \lambda_i \mathbf{1})) = Q(X - \lambda_i)$, dove $Q$ è il polinomio caratteristico di $N_i = f_i - \lambda_i \mathbf{1}$. Poiché il grado di $P_i$ è la dimensione di $V_i$, per concludere basta mostrare che $Q$ è una potenza di $X$ per ogni $i$. Ricordiamo che un endomorfismo si dice *nilpotente* se una sua potenza è nulla. Gli endomorfismi $N_i$ sono nilpotenti. Infatti, per definizione, $V_i$ è il nucleo di $(f - \lambda_i \mathbf{1})^{\mu_i}$, e quindi $N_i^{\mu_i} = 0$. Quanto dobbiamo dimostrare segue dunque dal seguente risultato.

LEMMA (5.9). *Sia $V$ uno spazio vettoriale di dimensione finita $n$, e sia $N$ un suo endomorfismo. Allora $N$ è nilpotente se e solo se il suo polinomio caratteristico è $X^n$.*

La dimostrazione è semplice. Da un lato, se il polinomio caratteristico è $X^n$, il teorema di Cayley-Hamilton dice che $N^n = 0$. Per dimostrare il viceversa poniamo $V_i = \ker N^i$, e osserviamo che $N(V_i) \subset V_{i-1} \subset V_i$. Se $N$ è nilpotente, $V_i = V$ per $i$ abbastanza grande. Vi è dunque una filtrazione crescente di $V$

$$\{0\} = V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1} = V .$$

Possiamo costruire una base $v_1, \dots, v_{n_1}, v_{n_1+1}, \dots, v_{n_2}, \dots$ in modo che $v_1, \dots, v_{n_1}$ sia una base di $V_1$, $v_1, \dots, v_{n_2}$ una base di $V_2$, e così via. Rispetto a questa base la matrice $A$ di $N$ è triangolare superiore e ha zeri sulla diagonale. Quindi $XI - A$ è triangolare superiore e i suoi elementi diagonali sono tutti uguali a $X$. Di conseguenza $\det(XI - A) = X^n$, come si voleva.

Completiamo questa sezione dando un'altra dimostrazione del teorema di Cayley-Hamilton (5.1), che però non fa uso del lemma (5.2). Anche questa dimostrazione è per induzione sulla dimensione $n$ dello spazio vettoriale $V$. Come si è già osservato nel corso della prima dimostrazione, il teorema è banalmente vero per $n = 1$. Se $n > 1$ distinguiamo

due casi. Il primo è quello in cui esiste un sottospazio vettoriale $W \subset V$, diverso da $\{0\}$ e da $V$, tale che $f(W) \subset W$, il secondo quello in cui ciò non accade. Nel primo caso possiamo scegliere una base $v_1, \dots, v_n$ per $V$ in modo che $v_1, \dots, v_h$ sia una base di $W$, dove $h < n$. Rispetto a questa base la matrice di $f$ è della forma

$$A = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix},$$

dove $A_1$ è un blocco $h \times h$, $A_2$ un blocco $(n-h) \times (n-h)$ e $B$ un blocco $h \times (n-h)$.

LEMMA (5.10). $\det(A) = \det(A_1)\det(A_2)$ *e* $P(X) = P_1(X)P_2(X)$, *dove* $P_1(X)$ *è il polinomio caratteristico di* $A_1$ *e* $P_2(X)$ *quello di* $A_2$.

Dato che

$$P(X) = \det(XI - A) = \det \begin{pmatrix} XI - A_1 & -B \\ 0 & XI - A_2 \end{pmatrix},$$

è sufficiente dimostrare la prima affermazione. Sappiamo che questa è vera se $B = 0$, e vogliamo ridurci a questo caso. Se le righe di $A_2$ sono tra loro dipendenti, anche le righe di $A$ lo sono, e quindi $\det(A) = \det(A_2) = 0$; in questo caso, dunque, il lemma è vero. Se invece le righe di $A_2$ sono indipendenti, costituiscono una base di $K^{n-h}$. Ne segue che ogni riga di $B$ è combinazione lineare di righe di $A_2$ e quindi, per eliminazione Gaussiana, possiamo ridurre la nostra matrice alla forma

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

senza cambiarne il determinante, che vale perciò $\det(A_1)\det(A_2)$. Questo conclude la dimostrazione del lemma.

Torniamo alla dimostrazione del teorema di Cayley-Hamilton. Induttivamente, possiamo supporre che esso valga per $A_1$ e $A_2$; possiamo cioè supporre che $P_1(A_1) = 0$, $P_2(A_2) = 0$. È immediato mostrare che

$$\begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}^k = \begin{pmatrix} A_1^k & C \\ 0 & A_2^k \end{pmatrix},$$

dove $C$ è una matrice $h \times (n-h)$. Ne segue che

$$P_1(A) = \begin{pmatrix} P_1(A_1) & D \\ 0 & E \end{pmatrix} = \begin{pmatrix} 0 & D \\ 0 & E \end{pmatrix}, \qquad P_2(A) = \begin{pmatrix} F & G \\ 0 & P_2(A_2) \end{pmatrix} = \begin{pmatrix} F & G \\ 0 & 0 \end{pmatrix},$$

dove $F$ è una matrice $h \times h$, $E$ una matrice $(n-h) \times (n-h)$, $D$ e $G$ matrici $h \times (n-h)$. Dunque

$$P(A) = P_1(A)P_2(A) = \begin{pmatrix} 0 & D \\ 0 & E \end{pmatrix} \begin{pmatrix} F & G \\ 0 & 0 \end{pmatrix} = 0,$$

come si doveva dimostrare.

Resta da esaminare il caso in cui non vi sono sottospazi non banali $W \subset V$ tali che $f(W) \subset W$. Scegliamo un elemento non nullo $v$ di $V$, e sia $h$ il minimo tra gli interi $m$ per cui $v, f(v), f^2(v), \dots, f^m(v)$ sono linearmente dipendenti. In altre parole, $v, f(v), f^2(v), \dots, f^{h-1}(v)$ sono linearmente indipendenti ma si ha una relazione di dipendenza lineare

$$f^h(v) + \sum_{i=0}^{h-1} a_i f^i(v) = 0\,.$$

Questa relazione mostra che, se $W$ è il sottospazio di $V$ generato da $v, f(v), \dots, f^{h-1}(v)$, allora $f(W) \subset W$. Dato che $W$ è diverso da $\{0\}$, deve essere necessariamente uguale a $V$. In altre parole, $v, f(v), f^2(v), \dots, f^{n-1}(v)$ costituiscono una base di $V$; per brevità porremo $v_i = f^{i-1}(v)$, per $i = 1, \dots, n$. Dato che $f^n(v)$ è combinazione lineare dei $v_i$ si ha una relazione di dipendenza lineare

$$f^n(v) + \sum_{i=0}^{n-1} a_i f^i(v) = 0\,. \tag{5.11}$$

Poniamo

$$Q(X) = X^n + \sum_{i=0}^{n-1} a_i X^i\,.$$

La relazione (5.11) dice che $Q(f)(v) = 0$. Applicando $f^{j-1}$ alla (5.11) si ottiene che

$$0 = f^{n+j-1}(v) + \sum_{i=0}^{n-1} a_i f^{i+j-1}(v) = f^n(f^{j-1}(v)) + \sum_{i=0}^{n-1} a_i f^i(f^{j-1}(v))\,,$$

cioè che $Q(f)(v_j) = 0$ per ogni $j$. Dato che i $v_j$ generano $V$ se ne deduce che $Q(f) = 0$. Per concludere basterà mostrare che $Q(X)$ non è altro che il polinomio caratteristico $P(X)$. La matrice di $f$ rispetto alla base $v_1, \dots, v_n$, che indicheremo con $A$, è particolarmente semplice. In effetti segue dalla definizione dei $v_i$ e da (5.11) che

$$\begin{aligned} f(v_i) &= v_{i+1} && \text{se } i < n\,,\\ f(v_n) &= -\sum_{i=1}^{n} a_{i-1} v_i\,. \end{aligned}$$

Dunque

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix},$$

e quindi

$$P(X) = \det \begin{pmatrix} X & 0 & 0 & \cdots & 0 & a_0 \\ -1 & X & 0 & \cdots & 0 & a_1 \\ 0 & -1 & X & \cdots & 0 & a_2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & -1 & X + a_{n-1} \end{pmatrix} .$$

Per calcolare questo determinante usiamo la regola di Laplace, sviluppando rispetto all'ultima colonna. Si ottiene

$$\begin{aligned} P(X) &= (X + a_{n-1}) \det(M_{n-1}) + \sum_{i=0}^{n-2} (-1)^{n-i-1} a_i \det \begin{pmatrix} M_i & 0 \\ 0 & L_{n-i-1} \end{pmatrix} \\ &= (X + a_{n-1}) \det(M_{n-1}) + \sum_{i=0}^{n-2} (-1)^{n-i-1} a_i \det(M_i) \det(L_{n-i-1}) , \end{aligned}$$

dove $M_i$ è la matrice $i \times i$

$$\begin{pmatrix} X & 0 & \cdot & \cdots & 0 & 0 \\ -1 & X & 0 & \cdots & 0 & 0 \\ 0 & -1 & X & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdot & \cdots & -1 & X \end{pmatrix}$$

e $L_i$ è la matrice $i \times i$

$$\begin{pmatrix} -1 & X & \cdot & \cdots & 0 & 0 \\ 0 & -1 & X & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & -1 & X \\ 0 & 0 & \cdot & \cdots & 0 & -1 \end{pmatrix} .$$

Dato che

$$\det(M_i) = X^i , \qquad \det(L_i) = (-1)^i ,$$

si conclude che

$$\begin{aligned} P(X) &= (X + a_{n-1}) X^{n-1} + \sum_{i=0}^{n-2} (-1)^{n-i-1} a_i X^i (-1)^{n-i-1} \\ &= X^n + \sum_{i=0}^{n-1} a_i X^i \\ &= Q(X) , \end{aligned}$$

come si voleva. Questo completa la dimostrazione del teorema di Cayley-Hamilton.