3. *Cayley-Hamilton and McCoy Theorems*

A little excursion into linear algebra.

For R a commutative ring, let $M_n(R)$ be the R-algebra of $n \times n$ matrices with entries in R. We denote the identity matrix by I.

3.1 Lemma. Let R be a commutative ring. For $A = (a_{ij}) \in M_n(R)$, define the **adjoint** of A as $\operatorname{adj}(A) = B = (b_{ij}) \in M_n(R)$ with $b_{ij} = (-1)^{i+j} \det(A^i_j)$, where $A^i_j \in M_{n-1}(R)$ results from A by removing the i-th column and the j-th row.

Then

$$AB = BA = (\det A)I.$$

Proof. Routine verification. $\square$

The adjoint of a matrix can be used to give a short proof of Cayley-Hamilton's theorem. It uses the natural R-algebra homomorphism between $M_n(R[x])$ and $M_n(R)[x]$, which maps the matrix whose (i, j)-th entry is $\sum_k a_k^{(i,j)} x^k \in R[x]$, to the polynomial $\sum_k A_k x^k$, where $A_k \in M_n(R)$ is the matrix whose (i, j)-th entry is $a_k^{(i,j)}$.

If S is a non-commutative ring, such as $M_n(R)$, and we want to substitute a ring element s for the variable in $f(x) = \sum a_k x^k \in S[x]$, we must specify whether substitution happens on the right or on the left of the coefficients, i.e., if $f(s)$ means $\sum a_k s^k$ (right substitution) or $\sum s^k a_k$ (left substitution).

In any case, it is important to remember that substitution (whether right or left) is in general not a homomorphism: if $f, g \in S[x]$ and $s \in S$, it is in general not true that $(f \cdot g)(s) = f(s) \cdot g(s)$ (unless s commutes with the coefficients of f and g). One thing that does work similarly to polynomials over commutative rings is the correspondence between zeros and linear factors:

3.2 Lemma. Let S be a (possibly non-commutative) ring and $f = \sum a_k x^k \in S[x]$. For $s \in S$ define $f(s) = \sum a_k s^k$. Then $f(s) = 0$ if and only if $f(x) = g(x)(x - s)$ for some $g \in S[x]$.

Proof. If $f(s) = 0$ then $f(x) = f(x) - f(s) = \sum_k a_k x^k - \sum_k a_k s^k = \sum_k a_k(x^k - s^k) = \sum_k a_k(\sum_{j=0}^{k-1} s^j x^{k-j-1})(x - s) = g(x)(x - s)$.

If, on the other hand, $f(x) = g(x)(x - s)$ with $g(x) = \sum_{k=0}^{n} b_k x^k$ then $f(x) = \sum_{k=0}^{n} b_k x^{k+1} - \sum_{k=0}^{n} b_k s x^k = \sum_{k=0}^{n+1} (b_{k-1} - b_k s) x^k$ and right substitution yields $f(s) = 0$. $\square$

3.3 Theorem (Cayley-Hamilton). Let C be a $n \times n$ matrix with entries in a commutative ring R, and $\chi(x) = \det(xI - C)$ its characteristic polynomial. Then $\chi(C) = 0$.

Proof. Let $B \in M_n(R[x])$ be the adjoint of $xI - C$ then

$$B \cdot (xI - C) = \chi(x)I.$$

Applying the R-algebra isomorphism $M_n(R[x]) \simeq M_n(R)[x]$, we get

$$B(x)(x - C) = \chi(x),$$

where $B(x) \in M_n(R)[x]$ is the polynomial corresponding to the matrix $B \in M_n(R[x])$. Now $\chi(x)$ has a factor $x - C$ in $M_n(R)[x]$, and therefore $\chi(C) = 0$. $\square$

With the argument we used to prove Cayley-Hamilton (3.3), one can actually show a stronger result, characterizing the ideal of those polynomials in $R[x]$, which have a given matrix $C \in M_n(R)$ as a zero:

3.4 Theorem (McCoy). Let C be a $n \times n$ matrix with entries in a commutative ring R. For $1 \leq k \leq n$ let $J_k(xI - C)$ be the ideal of $R[x]$ generated by the $k \times k$ minors of $xI - C$. Then the ideal of all polynomials $f \in R[x]$ with $f(C) = 0$ is equal to the ideal quotient $(J_n(xI - C) : J_{n-1}(xI - C)) = (\chi_C(x)R[x] :_{R[x]} J_{n-1}(xI - C))$.

Proof. Since the generators of $J_{n-1}(xI - C)$, the $(n-1) \times (n-1)$-minors of $(xI - C)$, are (up to sign) the entries of $adj(xI - C)$),

$$g(x) \in (\chi(x)R[x] :_{R[x]} J_{n-1}(xI - C))$$

is equivalent to

$$g(x)adj(xI - C) \in \chi(x)M_n(R[x]).$$

Multiplying by $(xI - C)$, which, being monic, is certainly not a zero-divisor in $M_n(R)[x]$, we again get an equivalent statement:

$$g(x)\chi(x)I \in \chi(x)M_n(R[x])(xI - C).$$

We may cancel the scalar $\chi(x) \in R[x]$ (a monic polynomial - not a zero-divisor in R[x]) to arrive at the equivalent statement

$$g(x)I \in M_n(R[x])(xI - C),$$

and apply the isomorphism between $M_n(R[x])$ and $M_n(R)[x]$ to get

$$g(x) \in M_n(R)[x](x - C),$$

which is (by 3.2) equivalent to C being a zero of $g(x)$. $\square$

From Cayley-Hamilton theorem for matrices it is possible to derive a version of Cayley-Hamilton theorem for linear operators, by associating matrices to linear operators:

Suppose M is an R-module generated by $m_1, \dots, m_n$. Then every $b \in M$ is representable as $b = b_1 m_1 + \dots + b_n m_n$ with (in general non-unique) $b_1, \dots, b_n$ in R. Let $\varphi \in \mathrm{End}_R(M)$. We say that a matrix $C \in M_n(R)$ represents φ with respect to the system of generators $m_1, \dots, m_n$, if for every $b \in M$, $b = b_1 m_1 + \dots + b_n m_n$ implies $\varphi(b) = b'_1 m_1 + \dots + b'_n m_n$ with $C[b_1, \dots, b_n] = [b'_1, \dots, b'_n]$. (We use square brackets to denote column vectors: $[b_1, \dots, b_n]$ means the transpose of $(b_1, \dots, b_n)$. It is easy to see that C represents φ if and only if the k-th column of C is a coordinate vector of $\varphi(m_k)$, i.e., if $\varphi(m_k) = c_{1k} m_1 + c_{2k} m_2 + \dots + c_{nk} m_n$.

If the generators $m_1, \dots, m_n$ are not R-linearly independent, then the matrix associated to a linear operator $\varphi \in \mathrm{End}_R(M)$ is non-unique, and not every matrix in $M_n(R)$ represents a linear operator. If C represents φ and D represents ψ, however, then $C + D$ and CD represent $\varphi + \psi$ and $\varphi \circ \psi$, respectively, and, for $r \in R$, rC represents $r\varphi$, so the set $E \subseteq M_n(R)$ of matrices representing R-endomorphisms of M is a R-subalgebra of $M_n(R)$, and there is a surjective R-algebra homomorphism $\pi\colon E \to \mathrm{End}_R(M)$, mapping every matrix in E to the endomorphism it represents. In particular, for every matrix C represening some $\varphi \in \mathrm{End}_R(M)$, restriction of π to $R[C]$ gives a surjective R-algebra homomorphism $\pi\colon R[C] \to R[\varphi]$, mapping $f(C) \in R[C]$ to $f(\varphi) \in R[\varphi]$ for every $f \in R[x]$.

3.5 Lemma. Let M be an R-module generated by n elements and $\varphi \in \mathrm{End}_R(M)$. If $C \in M_n(R)$ is a matrix associated to φ (with respect to a system of generators $m_1, \dots, m_n \in M$) and $\chi(x)$ the characteristic polynomial of C then $\chi(\varphi) = 0$.

Proof. By Cayley-Hamilton (3.3), $\chi(C) = 0$, where $\chi = \det(xI - C)$ is the characteristic polynomial of C. Applying the R-algebra homomorphism $\pi\colon R[C] \to R[\varphi]$ mentioned above, we see that $0 = \pi(\chi(C)) = \chi(\varphi)$. □

3.6 Corollary. Let M be an R-module generated by n elements and $\varphi \in \mathrm{End}_R(M)$, such that $\varphi(M) \subseteq JM$ for some ideal $J \trianglelefteq R$. Then φ is a zero of a monic polynomial in $R[x]$ of degree n, $f(x) = x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ with $a_k \in J^{n-k}$ for $0 \le k < n$.

Proof. If $m_1, \dots, m_n$ generate M as a R-module and $\varphi(M) \subseteq JM$ then there exist $c_{ij} \in J$ with $\varphi(m_k) = c_{1k} m_1 + c_{2k} m_2 + \dots + c_{nk} m_n$. Now $\chi(x) = \det(xI - C)$, the characteristic polynomial of $C = (c_{ij})$ is a monic polynomial of degree n with $\chi(\varphi) = 0$ and the property that the coefficient of x^k is in J^{n-k}. □