Math 154. Generalized Cayley-Hamilton and integrality

1. From endomorphisms to matrices

Let $A \to B$ be an injective map of commutative rings and assume B is finitely generated as an A-module. We seek to prove that any $b \in B$ satisfies a *monic* polynomial $f_b \in A[x]$ (so b is integral over A). The idea for how to construct f_b is to compute certain A-linear maps $B \to B$ in terms of "$n \times n$ matrices" over A, and to exploit a version of the Cayley-Hamilton theorem over rings.

To bring matrices into the picture, pick a spanning set $\{b_1, \dots, b_n\}$ for B as an A-module, so $B = \sum Ab_i$. There is no linear independence to grab onto, but nonetheless the idea of computing with matrices will be useful. For $b \in B$, we have the A-linear multiplication map $m_b : B \to B$ defined by $x \mapsto bx$.

Letting $\mathrm{End}_A(M)$ denote the set of A-linear maps $T : M \to M$ ("endomorphisms") for an A-module M, the element $m_b \in \mathrm{End}_A(B)$ determines b because $m_b(1) = b$. Also, it is clear from the commutativity of B that $m_b \circ m_{b'} = m_{bb'}$ and $m_{ab+a'b'} = am_b + a'm_{b'}$ in $\mathrm{End}_A(B)$. In other words, $b \mapsto m_b$ defines an *injective A-algebra homomorphism*

$$B \hookrightarrow \mathrm{End}_A(B),$$

where the target is an associative (generally non-commutative) A-algebra using composition as "multiplication" in $\mathrm{End}_A(B)$. (In the non-commutative setting, the phrase "A-algebra" encode the condition that A commutes with everything, which is the case in $\mathrm{End}_A(B)$ by the very definition of the A-module structure on B and the meaning of an "A-linear" endomorphism.)

It follows that if $f \in A[x]$ is a polynomial then $m_{f(b)} = f(m_b)$ in $\mathrm{End}_A(B)$, so $f(b) = 0$ if and only if $f(m_b) = 0$. Hence, to find a monic polynomial over A that kills b in B it is the same to find one that kills m_b in $\mathrm{End}_A(B)$. Thus, it suffices to prove rather generally that for *any* finitely generated A-module M and *any* $T \in \mathrm{End}_A(M)$ there is a monic polynomial $f \in A[x]$ such that $f(T) = 0$ in $\mathrm{End}_A(M)$. To find such an f, we will express T in terms of a "matrix" (even though M usually does not have an A-basis!) and then show that the characteristic polynomial of that matrix does the job.

Pick a finite spanning set $\{v_1, \dots, v_n\}$ for M as an A-module, so for every j we can express $T(v_j)$ as an A-linear combination of the v_i's. That is, there exist elements $a_{ij} \in A$

$$T(v_j) = \sum_i a_{ij} v_i.$$

Of course, in the absence of linear independence there are probably zillions of different ways to choose these a_{ij}'s, but that won't matter. The key point is that if $T' \in \mathrm{End}_A(M)$ is another endomorphism and we make a choice of elements $a'_{ij} \in A$ such that $T'(v_j) = \sum_i a'_{ij} v_i$ then for any $a, a' \in A$ we have

$$(aT + a'T')(v_j) = \sum_i (aa_{ij} + a'a'_{ij}) v_i, \quad (T \circ T')(v_j) = \sum_i (\sum_h a_{ih} a'_{hj}) v_i$$

by the exact same formal computations as used in linear algebra to show that "matrix algebra" computes linear combinations and composition among linear endomorphisms of a vector space. (The point is that those computations never use the linear independence aspect of bases, only their spanning property, so they carry over to the present setting without change.) Thus, if we let $\mu = (a_{ij}) \in \mathrm{Mat}_n(A)$ ($n \times n$ matrices over A) then for any integer $e \geq 0$ the matrix power μ^e computes the effect of $T^e \in \mathrm{End}_A(M)$ in the sense that $T^e(v_j)$ is the A-linear combination of the v_i's given by the jth column of the matrix μ^e. Hence, likewise, for any $f \in A[x]$, the endomorphism

$f(T) \in \mathrm{End}_A(M)$ is computed by $f(\mu)$ in the sense that $f(T)(v_j)$ is the A-linear combination of the v_i's given by the jth column of the matrix $f(\mu) \in \mathrm{Mat}_n(A)$. But the v_i's *span* M over A, so although they may not be linearly independent (and hence we cannot generally *define* an element of $\mathrm{End}_A(M)$ by arbitrarily specifying its effect on the v_i's separately) we see that an element of $\mathrm{End}_A(M)$ is *uniquely determined* by its effect on the v_i's. In particular, an element of $\mathrm{End}_A(M)$ that kills the v_i's must be 0.

To summarize, we have shown that if $f \in A[x]$ is a polynomial and $f(\mu) = 0$ in $\mathrm{Mat}_n(A)$ then $f(T) = 0$ in $\mathrm{End}_A(M)$. Hence, to find a monic f that kills T it suffices to find such an f that kills the $n \times n$ matrix μ over A. Rather generally, for any $\mu \in \mathrm{Mat}_n(A)$ we can consider its characteristic polynomial $\chi_\mu \in A[x]$ defined as the determinant

$$\chi_\mu = \det(x \cdot \mathrm{id}_n - \mu) \in A[x]$$

(determinant of a matrix with entries in the commutative ring $A[x]$, with id_n the $n \times n$ identity matrix), and by inspection this is a *monic* polynomial of degree n. Thus, we will be done if we can show that $\chi_\mu(\mu) = 0$ in $\mathrm{Mat}_n(A)$. This is what we shall now do.

2. A universal identity

We continue to let A be a commutative ring. Our aim is to prove the following generalization of the Cayley-Hamilton theorem.

Theorem 2.1. *For any $\mu = (a_{ij}) \in \mathrm{Mat}_n(A)$ with $n \geq 1$ and its characteristic polynomial $\chi = x^n + \cdots \in A[x]$, the matrix $\chi(\mu) \in \mathrm{Mat}_n(A)$ vanishes.*

To prove this result, we first note that when A is a field then this is the usual Cayley-Hamilton theorem. We will reduce the general case to the special case of fields, or even just algebraically closed fields of characteristic 0 (or even just $\mathbf{C}$!) by means of a powerful trick: to prove "universal" identities, we reduce the problem to a "universal" case which occurs over a ring with special features (e.g., a domain) and then we exploit those special features to prove the result in the universal case by a method which is often not directly applicable in other cases.

In our setting, the method goes as follows. Consider the "universal $n \times n$ matrix", by which we just mean the matrix $\mu^{\mathrm{univ}} = (X_{ij}) \in \mathrm{Mat}_n(R^{\mathrm{univ}})$ where $R^{\mathrm{univ}} = \mathbf{Z}[X_{ij}]$ is the polynomial ring in n^2 variables over $\mathbf{Z}$ (indexed by pairs of integers $1 \leq i, j \leq n$). This example is universal in the sense that for *any* A and *any* $\mu = (a_{ij})$ there is a unique ring map $\phi : R^{\mathrm{univ}} \to A$ under which $\mathrm{Mat}_n(R^{\mathrm{unv}}) \to \mathrm{Mat}_n(A)$ carries μ^{univ} to μ. Indeed, the unique such ϕ is given by $X_{ij} \mapsto a_{ij}$. Actually, the uniqueness of ϕ is not what matters; rather, we care about just the existence. That is, we have made a specific $n \times n$ matrix μ^{univ} over the ring R^{univ} which has some extra properties (it is a domain!!) so that our original μ is obtained from μ^{univ} by some ring homomorphism $\phi : R^{\mathrm{univ}} \to A$.

The utility of this is due to the observation that since ϕ is a ring homomorphism, the induced map $R^{\mathrm{univ}}[x] \to A[x]$ carries $\chi_{\mu^{\mathrm{univ}}}$ to χ_μ and so likewise the map of matrix rings

$$\mathrm{Mat}_n(R^{\mathrm{univ}}) \to \mathrm{Mat}_n(A)$$

carries $\chi_{\mu^{\mathrm{univ}}}(\mu^{\mathrm{univ}})$ to $\chi_\mu(\mu)$. (This just expresses the fact that the formation of $\chi_\mu(\mu)$ is a "universal formula" which has nothing to do with the specifics of A or μ.) Hence, if we can prove the vanishing result for the pair $(\mu^{\mathrm{univ}}, R^{\mathrm{univ}})$ then it follows for the given pair (μ, A)!

This formalism reduces the general problem to the special case of μ^{univ} over R^{univ}. But what extra properties does this case have which are not available in the general case? The main point is that R^{univ} is a domain, so we have reduced the general case to the special case when the coefficient ring for the matrices is a *domain*.

Remark 2.2. We have also gained other properties; e.g., R^{univ} has fraction field of characteristic 0 that even embeds into $\mathbf{C}$, and one can show that the characteristic polynomial of μ^{univ} is diagonalizable over an algebraic closure of the fraction field. This won't be needed for our purposes, but it is a mechanism for reducing various matrix identities to the special case of diagonalizable matrices, thereby vindicating the belief of most physicists that "all" matrices are diagonalizable.

Now we may and do assume A is a domain, say with fraction field K. Then the vanishing of $\chi_\mu(\mu)$ in $\mathrm{Mat}_n(A)$ is the same as the analogue in $\mathrm{Mat}_n(K)$. But this version over K is the Cayley-Hamilton theorem!