Polynomials over Noncommutative Rings and the Cayley-Hamilton Theorem

Mathematics 482/526, Spring 2005

In this note we give a proof of the Cayley-Hamilton theorem by making use of the theory of polynomials over noncommutative rings. The argument we give appears in an article by Greenberg in the American Mathematical Monthly [1]. The motivation for this approach is the following suggestive but incorrect proof. Let $\Psi(x) = \det(xI - A)$ be the characteristic polynomial of a square matrix A. One would like to say $\Psi(A) = \det(AI - A) = \det(0) = 0$, but there is a mistake in this argument, as we will see in Example 2. By being careful about what is going on, we will be able to recover the idea of this proof.

Let R be a possibly noncommutative ring. In the ring of polynomials $R[x]$ we will always write coefficients on the left; this distinction is important, as we will see below. Let $a \in R$. If $f(x) = r_0 + r_1 x + \cdots + r_n x^n$, then we define the *evaluation* $f(a)$ at a by $f(a) = r_0 + r_1 a + \cdots + r_n a^n$. Since $ra \neq ar$ in general in R, it is in defining $f(a)$ that we have to be careful to write coefficients on one side or another of x. The main problem with evaluation for noncommutative rings is that the map that sends $f(x)$ to $f(a)$ does not preserve multiplication.

Example 1. Let $R = M_2(\mathbb{R})$, and let

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Note that $ab \neq ba$ since $ab \neq 0$ but $ba = 0$. If $f(x) = x$ and $g(x) = b$, then $f(a) = a$ and $g(a) = b$. Moreover, $fg(x) = bx$. So, $fg(a) = ba$, but $f(a)g(a) = ab$, so $fg(a) \neq f(a)g(a)$.

Example 2. Let $a \in M_n(F)$ and set $\Psi(x) = \det(xI - a)$. If $b \in M_n(F)$, then $\Psi(b) \neq \det(bI - a)$ in general, since $\Psi(b)$ is a matrix but $\det(bI - a)$ is a scalar. Even more, it is not true that $\Psi(b) = \det(bI - a)I$ in general; to see this, let a, b be as in the previous example. Then $\psi(x) = x(x - 1) = x^2 - x$. Thus, $\psi(b) = b^2 - b$. However, $b^2 = 0$, so $\psi(b) = -b$. On the other hand, $\det(bI - a) = \det(b - a) = 0$. Therefore, the "proof" above of the Cayley-Hamilton theorem certainly is false.

Example 3. The significance of the fact that evaluation does not preserve multiplication is in the consideration of roots. If $a \in R$ with $f(a) = 0$, then we cannot conclude that a is also a root of $f(x)g(x)$ for any $g(x)$. To see this, with a, b as in the previous example, let $f(x) = bx$. Then $f(a) = ba = 0$. Let

$$c = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

If $g(x) = c$, then $(fg)(x) = bcx$. Then

$$(fg)(a) = bca = a,$$

while $f(a)g(a) = 0$ since $f(a) = 0$. Therefore, a is not a root of $fg(x)$ even though it is a root of $f(x)$.

The main properties of the evaluation function are given in the following proposition. The third statement below is the key fact we need to prove the Cayley-Hamilton theorem.

Proposition 4. *Let $f(x), g(x) \in R[x]$ and let $a \in R$.*

(1) $(f + g)(a) = f(a) + g(a)$.

(2) *If $f(x) = rx^n$ for some n and some $r \in R$, then $(fg)(a) = rg(a)a^n$.*

(3) *If $g(a) = 0$, then $(fg)(a) = 0$.*

(4) *If $c = g(a)$ is a unit in R, then $(fg)(a) = f(cac^{-1})g(a)$.*

Proof. (1): Let $f(x) = r_0 + \cdots + r_n x^n$ and $g(x) = s_0 + \cdots + s_n x^n$; by adding zero coefficients if necessary we write f and g in this way. Then $(f + g)(x) = (r_0 + s_0) + \cdots + (r_n + s_n)x^n$, so

$$\begin{aligned}(f + g)(a) &= (r_0 + s_0) + \cdots + (r_n + s_n)a^n \\ &= r_0 + r_1 a + \cdots + r_n a^n + s_0 + \cdots + s_n a^n \\ &= f(a) + g(a).\end{aligned}$$

To prove (2), writing $g(x) = s_0 + \cdots + s_m x^m$, we have

$$fg(x) = rs_0 x^n + rs_1 x^{n+1} + \cdots + rs_m x^{n+m},$$

so

$$\begin{aligned}(fg)(a) &= rs_0 a^n + \cdots + rs_m a^{n+m} \\ &= r\left(s_0 a^n + \cdots + s_m a^{n+m}\right) \\ &= r\left(s_0 + \cdots + s_m a^m\right) a^n \\ &= rg(a)a^n.\end{aligned}$$

For (3) and (4), write $f(x) = r_0 + \cdots + r_n x^n$. Then $fg(x) = \sum_{i=0}^{n} r_i x^i g(x)$. Then $(fg)(a) = \sum_{i=0}^{n} r_i g(a) a^i = 0$ by (1) and (2) since $g(a) = 0$. If, on the other hand, $c = g(a)$ is a unit in R, then

$$\begin{aligned}(fg)(a) &= \sum_{i=0}^{n} r_i g(a) a^i = \sum_{i=0}^{n} r_i c a^i = \sum_{i=0}^{n} r_i c a^i c^{-1} c \\ &= \sum_{i=0}^{n} r_i \left(cac^{-1}\right)^i c = \left(\sum_{i=0}^{n} r_i \left(cac^{-1}\right)^i\right) c \\ &= f(cac^{-1}) g(a).\end{aligned}$$

□

Corollary 5. *If $a \in R$ is a root of $g(x)$, then a is a root of $f(x)g(x)$ for any $f(x) \in R[x]$.*

The lack of symmetry of the corollary comes from our choice to write coefficients on the left. If we wrote them on the left, then roots of $f(x)$ would be roots of $f(x)g(x)$ for any $g(x) \in R[x]$, as a simple modification of the results above will show. Part (4) of the proposition is most applicable for polynomials over a division ring.

Example 6. Another difference in considering polynomials over noncommutative rings is in the number of roots a polynomial can have. Let $R = \mathbb{H}$, the ring of Hamilton's quaternions. This is the ring

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\},$$

where multiplication is determined by the basic relations

$$\begin{aligned} i^2 = j^2 = k^2 &= -1, \\ ji = -ij &= -k. \end{aligned}$$

Consider $f(x) = x^2 + 1$. Then $f(i) = f(j) = f(k) = 0$. Thus, f has more roots than its degree. In fact, the situation is even more strange than it appears. If $\alpha = bi + cj + dk$, then $\alpha^2 = -(b^2 + c^2 + d^2)$. Therefore, if we choose b, c, d so that $b^2 + c^2 + d^2 = 1$, then α is a root of $f(x)$. Thus, f has infinitely many roots in $\mathbb{H}$.

To further illustrate Proposition 4, consider the polynomial $h(x) = (x - i)(x - j) = x^2 - (i + j)x + k$. An easy calculation shows $h(i) = i^2 - (i + j)i + k = 2k \neq 0$. Thus, i is not a root of $h(x)$ even though $x - i$ is a factor of $h(x)$. Setting $f(x) = x - i$ and $g(x) = x - j$, we have $g(i) = i - j$. Using part (4) of the Proposition, we should have $h(i) = f(cic^{-1})g(i)$, where $c = g(i)$. We have, $g(i)ig(i)^{-1} = (i - j)i(i - j)^{-1} = -j$, and so $h(i) = f(-j)g(i) = (-j - i)(i - j) = -(i + j)(i - j) = 2k$, as we obtained earlier.

We now consider the polynomial ring $M_n(F)[x]$. Recall that if S is a commutative ring and $A \in M_n(S)$, then, by identifying α with αI in $M_n(F)$, the scalar $\det(A)$ is given by the formula $\det(A) = \text{adj}(A) \cdot A$, where $\text{adj}(A)$ is the adjoint of A. For example, if $A \in M_n(F)$,

then we can view $xI - A \in M_n(F[x])$, and by doing so, we have $\Psi(x) = \det(xI - A) = \text{adj}(xI - A) \cdot (xI - A)$. However, in the proof of the Cayley-Hamilton theorem below, it is more advantageous to view $xI - A \in M_n(F)[x]$.

Theorem 7 (Cayley-Hamilton). *Let $A \in M_n(F)$. If $\Psi(x)$ is the characteristic polynomial of A, then $\Psi(A) = 0$.*

Proof. Let $f(x) = \text{adj}(xI - A)$ and $g(x) = xI - A$. Then $\Psi(x) = f(x)g(x)$, and it is obvious that $g(A) = 0$. Thus, by Part 3 of Proposition 4, $\Psi(A) = 0$. □

References

[1] M. J. Greenberg, *Note on the Cayley-Hamilton theorem*, Amer. Math. Monthly **91** (1984), no. 3, 193–195.