# The Cayley–Hamilton Theorem*

Attila Máté
Brooklyn College of the City University of New York

March 23, 2016

## Contents



# 1 Introduction

Given a square matrix $A$ and writing $I$ for the identity matrix of the same size, the characteristic polynomial of $A$ is the determinant of the matrix $\lambda I - A$, which is a polynomial of $\lambda$. The Cayley–Hamilton Theorem asserts that if one substitutes $A$ for $\lambda$ in this polynomial, then one obtains the zero matrix. This result is true for any square matrix with entries in a commutative ring.

*Written for the course Mathematics 4101 at Brooklyn College of CUNY.

For a matrix of a given size, this theorem can be restated as a number of polynomial identities in terms of the entries of the matrix $A$ – namely, one identity for each entry being 0 of the matrix resulting by substituting $A$ for $\lambda$; if such an identity is satisfied over the ring of integers then it is satisfied over any commutative ring (see Subsection 1.1). Therefore, in proving the Cayley–Hamilton Theorem it is permissible to consider only matrices with entries in a field, since if the identities are true in the field of reals then they are also true in the ring of integers.

There are two basic approaches to proving such a result. In one approach, one considers $A$ as representing a linear transformation on a vector space, and obtains the result as a consequence of studying the structure of linear transformations on vector spaces. In such an approach it is important to make sure that $A$ is a matrix over a field, since structures similar to vector spaces over rings (called modules) lack many of the basic properties of vector spaces. Another approach establishes the result directly as a consequence of properties of matrices and determinants. This type of approach may work directly for matrices over commutative rings.

Below we describe a proof using the second approach. When one wants to substitute the matrix $A$ for $\lambda$ in the determinant of $\lambda I - A$, one cannot do this directly, since a determinant must have scalar entries; so first one needs to rewrite the determinant as a polynomial. In the approach we use, the determinant $\lambda I - A$ will be written as a product of two polynomials with matrix coefficients, and the result of substituting $A$ for $\lambda$ will clearly give the zero matrix. The argument completely avoids calculations, but to follow it there are subtle points of algebra that need to be clearly understood. To this end we need to make a detour into a formal discussion as to what a polynomial is, and what kind of an algebraic structure they form.

## 1.1 A multivariate polynomial zero on all integers is identically zero

Let $P(x_1, x_2, \ldots, x_k)$ be a polynomial with integer coefficients, i.e., a sum of integer multiples of products formed with the variables $x_1$, $x_2$, …, $x_k$, and assume that $P(u_1, u_2, \ldots, u_k) = 0$ for any choice of the integers $u_1$, $u_2$, …, $u_k$. We claim that then $P(x_1, x_2, \ldots, x_k) = 0$ identically, i.e., after all cancelations everything will cancel, that is, $P(x_1, x_2, \ldots, x_k)$ will be a sum of a number zero of products.

Indeed, we can write

$$P(x_1, x_2, \ldots, x_k) = \sum_{i=0}^{n} P_i(x_1, x_2, \ldots, x_{k-1}) x_k^i$$

Choosing $x_1 = u_1$, $x_2 = u_2$, …, $x_{k-1} = u_{k-1}$ for some integers $u_1$, $u_2$, …, $u_{k-1}$, this is a polynomial in the single variable $x_k$. Since a polynomial of degree $n$ can have at most $n$ zeros, this polynomial being zero for all $x_k$ means that all the coefficients $P_i(u_1, u_2, \ldots, u_{k-1})$ are zero. Since this is true for any choice of the integers $u_1$, $u_2$, …, $u_{k-1}$, this implies by induction on the number $k$ of variables in $P(x_1, x_2, \ldots, x_k)$ that all coefficients $P_i(x_1, x_2, \ldots, x_{k-1})$ are identically zero.

## 2 Polynomials over a ring

**Definition 1.** A *ring* is a set equipped with two binary operation, called addition (symbol $+$) and multiplication (symbol $\cdot$, usually omitted) with the following properties. For all $a, b, c \in R$ we have

$$\begin{aligned}(a+b)+c &= a+(b+c),\\ a+b &= b+a,\\ (ab)c &= a(bc),\\ a(b+c) &= ab+ac,\\ (a+b)c &= ac+bc,\end{aligned}$$

i.e., the addition is associative and commutative, and the multiplication is associative and distributive over addition. Further, there are elements $0, 1 \in R$ such that

$$\begin{aligned}a+0 &= a,\\ a\cdot 1 = 1\cdot a &= a\end{aligned}$$

for all $a \in R$ (additive and multiplicative identities, respectively), and, finally, for every $a \in R$ there is an element $-a \in R$ such that

$$a+(-a) = 0$$

($-a$ is called an additive inverse of $a$).

Not every author requires the existence of a multiplicative identity in a ring, but recently it has been common to require the existence of a multiplicative identity. A ring without a multiplicative identity might be called a *pseudo-ring*. There are simple proofs that the additive and multiplicative identities and the additive inverse of an element are unique. A *commutative ring* is a ring in which the multiplication is commutative.

A *formal power series* over a ring $R$ is intuitively described as a sum

$$\sum_{i=0}^{\infty} a_i\lambda^i \qquad (a_i \in R),$$

where $\lambda$ is a formal variable; the $a_i$'s are called coefficients. If all but a finite number of the coefficients are 0, then a formal power series is called a *polynomial*. The addition addition and multiplication of formal power series is defined as

$$\begin{aligned}\sum_{i=0}^{\infty} a_i\lambda^i + \sum_{i=0}^{\infty} b_i\lambda^i &= \sum_{i=0}^{\infty}(a_i+b_i)\lambda^i,\\ \sum_{i=0}^{\infty} a_i\lambda^i \cdot \sum_{i=0}^{\infty} b_i\lambda^i &= \sum_{i=0}^{\infty}\Big(\sum_{k=0}^{i} a_k b_{i-k}\Big)\lambda^i.\end{aligned}$$

The sum notation here is merely formal, no actual addition is meant. A more formal definition can be given as follows.

## 2.1 A formal definition of polynomials

Write $\mathbb{N}$ for the set $\{0, 1, 2, \ldots\}$ of natural numbers (nonnegative integers).[1] Given a function $f$, write $f\text{`}x$ for the value of the function at $x$.[2]

**Definition 2.** A *formal power* series over a ring $R$ is defined as a function $f : \mathbb{N} \to R$, with the operations $+$ and $\cdot$ defined as follows. Given formal power series $f$ and $g$ over $R$, we define $f + g$ and $fg$ as formal power series such that for all $n \in N$

$$(f + g)\text{`}n = f\text{`}n + g\text{`}n,$$
$$(fg)\text{`}n = \sum_{k=0}^{n} (f\text{`}k)\big(g\text{`}(n - k)\big).$$

A polynomial over a ring $R$ is a formal power series $p$ such that $p\text{`}n = 0$ for all but finitely many $n \in \mathbb{N}$.

Writing $\lambda$ for the formal power series over $R$ such that $\lambda\text{`}1 = 1$ and $\lambda\text{`}n = 0$ for $n \neq 1$ ($n \in \mathbb{N}$), the intuitive description above of a formal power series can be given a precise meaning. $\lambda$ is called a *formal variable*. We will mostly use the more suggestive notation given in this intuitive description rather than the formal description given in the definition above, except when we want to be meticulously precise. The polynomials over a ring $R$ with operations given in the above definition form a ring. If $\lambda$ is the name of the formal variable, this ring is denoted as $R[\lambda]$. If $p$ is a polynomial, $p\text{`}n$ for $n \in \mathbb{N}$ will be called the $n$th *coefficient* of $p$.

## 2.2 The evaluation operator

An *operator* assigns objects to certain given objects. While a distinction can be made between operators and functions, such a distinction will not be necessary for our purposes.

**Definition 3.** Given a ring $R$, the evaluation operator is a function $\mathrm{ev} : R[\lambda] \times R \to R$ such that, for $p \in R[\lambda]$ $a \in R$ we have

$$\mathrm{ev}\,\text{`}(p, a) = \sum_{n=0}^{\infty} (p\text{`}n)a^n.$$

Since we did not assume that $R$ is commutative, one needs to be a little careful, since if $p$ and $q$ are polynomials over $R$, one does not in general have $\mathrm{ev}\,\text{`}(pq, a) = \big(\mathrm{ev}\,\text{`}(p, a)\big)\big(\mathrm{ev}\,\text{`}(q, a)\big)$. In fact, we could have called the operator ev the right-evaluation operator (since the formal variable is substituted on the right), and we could have similarly defined a left-evaluation operator. The reason we need to deal with noncommutative rings here is that we will consider polynomials with matrix coefficients. An important exception where the equality $\mathrm{ev}\,\text{`}(pq, a) = \big(\mathrm{ev}\,\text{`}(p, a)\big)\big(\mathrm{ev}\,\text{`}(q, a)\big)$ does indeed hold is described by the following

**Lemma 1** (Evaluation Lemma)**.** *Let $R$ be a ring, and let $a \in R$ and $p, q \in R[\lambda]$. Assume that the element $a$ commutes with the coefficients of $q$, that is*

$$a \cdot (q\text{`}n) = (q\text{`}n) \cdot a \qquad \textit{for all } n \in \mathbb{N}.$$

[1] The number 0 is sometimes considered a natural number, sometimes it is not. In these notes we will always regard 0 as a natural number.

[2] This is the notation used by Kurt Gödel [1]. We will use the notation $f(x)$ as the result of the application *evaluation operator* (see below), to be distinguished from the value of a function.

*Then*

$$\mathrm{ev}\,`(pq,a) = \big(\mathrm{ev}\,`(p,a)\big)\big(\mathrm{ev}\,`(q,a)\big).$$

*Proof.* The proof is a fairly simple calculation. We will give the details, using informal notation. Writing $b_n = p`n$ and $c_n = q`n$ for the coefficients $p$ and $q$, we have

$$pq = \sum_{n=0}^{\infty}\Big(\sum_{k=0}^{n} b_k c_{n-k}\Big)\lambda^n;$$

the outside sum is of course finite, since only finitely many among the $b_n$ and $c_n$ are not zero. Noting that $ac_n = c_n a$ for all $n$ by our assumptions, we have

$$\begin{aligned}\mathrm{ev}\,`(pq,a) &= \sum_{n=0}^{\infty}\Big(\sum_{k=0}^{n} b_k c_{n-k}\Big)a^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n}(b_k a^k)(c_{n-k}a^{n-k})\\ &= \Big(\sum_{k=0}^{\infty} b_k a^k\Big)\Big(\sum_{l=0}^{\infty} c_l a^l\Big) = \big(\mathrm{ev}\,`(p,a)\big)\big(\mathrm{ev}\,`(q,a)\big),\end{aligned}$$

establishing the desired equality. □

One customarily writes $p(a) = \mathrm{ev}\,`(p,a)$. The use of this notation necessitated for us to change the notation for the value of a function, since the result of the evaluation operation needs to be distinguished from the value of a function.

## 2.3 $R$ as a subring of $R[\lambda]$

The way we defined polynomials, $R$ is not a subset of $R[\lambda]$. It is, however, natural to identify a constant polynomial $a\lambda^0$ with $a$; with this identification, $R$ becomes a subring of $R[\lambda]$. One could modify the definition of polynomials in such a way that one would indeed have the inclusion $R \subset R[\lambda]$ at the price of some additional complications. In informal or semi-formal discussions we will indeed assume that $R \subset R[\lambda]$.

If $R$ is a ring and $R[\lambda]$ is the ring of polynomials over $R$, one can introduce a new formal variable $\mu$ and consider the ring of polynomials $R[\lambda][\mu]$ over $R[\lambda]$. If, using intuitive notation, we have

$$p = \sum_{n=0}^{\infty} a_n \lambda^n,$$

then, identifying an element $a$ of $R$ with the polynomial $a\lambda^0$,

$$q = \sum_{n=0}^{\infty} a_n \mu^n,$$

is polynomial with coefficients in $R[\lambda]$, i.e., it is an element of $R[\lambda][\mu]$. Writing $\mathrm{ev}_{R[\lambda][\mu]}$ for the evaluation operator $\mathrm{ev}_{R[\lambda][\mu]} : R[\lambda][\mu] \to R[\lambda]$, we have

$$\mathrm{ev}_{R[\lambda][\mu]}\,`(q,\lambda) = p.$$

The reader is encouraged to work out the formal details. On account of this observation, attention is called to the notational confusion in the uses of $p(\lambda)$ and $p(a)$: sometimes, $p(\lambda)$ indicates the polynomial $p$ itself, at other times one uses the same notation for the polynomials $p$ and $q$ above, and writes $p(\mu)$ and $p(\lambda)$ instead of writing $q$ and $\mathrm{ev}_{R[\lambda][\mu]}\,`(q,\lambda)$, and at yet other times one loosely considers $p$ as a function $p : R \to R$, and $p(a)$ denotes the value of this function for some $a \in R$.

# 3 Determinants and the adjugate matrix

## 3.1 Matrices

Given positive integers $m$ and $n$, an $m \times n$ matrix over $R$ is a rectangular arrangement of elements of $R$ with the well-known definition of matrix operations. To give a more formal definition, such a matrix $A$ is defined as a function

$$A : \{i : 1 \le i \le n\} \times \{j : 1 \le j \le n\} \to R,$$

and when one writes $A = (a_{ij})$ one means $a_{ij} = A`(i, j)$.

## 3.2 Determinants

The theory of determinants can in a natural way developed over commutative rings [2, Chapter 5, pp. 140–156] or [3, §17, pp. 152–159]. Much of discussion in [4, Chapter 4, pp. 161–201] can also be developed in rings, but the proof of the product theorem for determinants[3] given in [4, (4.1.5) Theorem, p. 171] relies on the row-echelon form of a matrix, and so it works only for matrices over a field.

Let $n$ be a positive integer, and $A = (a_{ij})$ be an $n \times n$ matrix over a ring $R$ ($a_{ij} \in R$). The determinant of $A$ is defined as

$$\det` A = \sum_{\sigma} (\operatorname{sgn}` \sigma) \prod_{i=1}^{n} a_{i\sigma`i}, \tag{1}$$

where $\sigma$ runs over all permutations (one-to-one mappings of a set onto itself) of the set $\{1, 2, \ldots, n\}$, and $\operatorname{sgn}` \sigma$ is the sign of the permutation $\sigma$ (i.e., 1 for an even permutation and $-1$ for an odd permutation); this is called the Leibniz formula for determinants.

For the matrix $A$, the cofactor $A_{ij}$ can be obtained as follows. Let $A(i|j)$ denote the matrix obtained by deleting the $i$th row and the $j$th column of $A$, and put

$$A_{ij} = (-1)^{i+j} \det` A(i|j).$$

For any $i$ with $1 \le i \le n$ we have

$$\det` A = \sum_{k=1}^{n} a_{ik} A_{ik} \quad \text{and} \quad \det` A = \sum_{k=1}^{n} a_{ki} A_{ki}.$$

This is called the Laplace expansion of determinants, and it can be used instead of Leibniz's formula in a recursive definition of determinants if one starts by saying that for $a \in A$ and the $1 \times 1$ matrix $(a)$ we put $\det`(a) = a$. Write

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \ne j; \end{cases}$$

$\delta_{ij}$ is called Kronecker's delta. It is well known that a determinant with two equal rows or two equal columns is 0. Therefore, for any $i$, $j$ with $1 \le i, j \le$ we have

$$\sum_{k=1}^{n} a_{ik} A_{jk} = \delta_{ij} \det` A \quad \text{and} \quad \sum_{k=1}^{n} a_{ki} A_{ki} = \delta_{ij} \det` A. \tag{2}$$

[3] Saying that if $A$ and $B$ are two square matrices of the same size, then, $\det`(AB) = (\det` A)(\det` B)$, where $\det`\cdot$ denotes the determinant of a square matrix.

Indeed, if $i = j$, these formulas give Laplace expansion of the determinant of $A$, and if $i \neq j$, they give Laplace expansions of a determinant with two equal rows or two equal columns. The matrix

$$\operatorname{adj}`A \stackrel{def}{=} (A_{ij})_{j,i} = (A_{ij})^T$$

is called the *adjugate* of the matrix $A$.[4] Equations (2) can also be expressed in matrix form as

$$(\operatorname{adj}`A)A = A(\operatorname{adj}`A) = (\det`A)I \tag{3}$$

for any $n \times n$ matrix over a ring $R$, where $I$ denotes the $n \times n$ identity matrix. If $\det`A$ has a multiplicative inverse, then the inverse matrix of $A$ can be given as $(\det`A)^{-1} \operatorname{adj}`A$; we will not use this latter formula, since, dealing with rings, we wil not be interested in multiplicative inverses.

# 4 The Cayley–Hamilton Theorem

**Definition 4.** Given a positive integer $n$ and an $n \times n$ matrix over a commutative ring $R$, the *characteristic polynomial* of $A$ is the polynomial $\det`(\lambda I - A)$, where $I$ is the $n \times n$ identity matrix.

Assuming that $R$ is a commutative ring, the polynomial ring $R[\lambda]$ is commutative. The entries of the matrix $\lambda I - A$ are elements of the ring $R[\lambda]$. Since determinants can be defined for any square matrix over a commutative ring, the determinant of this matrix is well defined.[5]

**Theorem 1** (Cayley–Hamilton Theorem; informal version)**.** *Let $n$ be a positive integer, let $R$ be a commutative ring, and let $A$ be an $n \times n$ matrix over $R$. Then $A$ is a zero of its own characteristic polynomials; that is,*

$$\operatorname{ev}`\big(\det`(\lambda I - A), A\big) = 0 \tag{4}$$

There is a minor problem with the interpretation of the evaluation operator in the last formula: We have $\det`(\lambda I - A) \in R[\lambda]$, i.e. the coefficients of this polynomial are elements of $R$, yet we want to evaluate this polynomial for a matrix, that is, for an element of $R_{n\times n}$. While intuitively it is clear what is being meant, the formal interpretation is a little more delicate; this is why we called the theorem above the informal version of the Cayley–Hamilton theorem.

The Cayley–Hamilton Theorem asserts that the matrix $A$ itself is a zero of this polynomial. In other words, if $A = (a_{ij})$, then

$$\det`(\lambda I - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \dots & \dots & \dots & \dots \\ -a_{n1} & -a_{n2} & \dots & \lambda - a_{nn} \end{vmatrix} \tag{5}$$

Now, here one cannot directly substitute $A$ for $\lambda$, since the matrices we are dealing with are not supposed to have matrices for entries. However, when we expand this determinant, we obtain a polynomial of degree $n$, and if we substitute $A$ for $\lambda$ in that polynomial, then we obtain zero, according to the Cayley–Hamilton theorem.

[4] For a matrix $A$ we write $A^T$ for the transpose of the matrix $A$. If $A = (a_{ij})$ then we also indicate the transpose by writing $A^T = (a_{ij})_{j,i}$. The order of the subscripts $j$ and $i$ outside the parentheses indicate that $j$ is used to index rows and $i$, to index columns. The adjugate matrix was used to be called the adjoint of $A$, but later, in the theory of Hilbert spaces, the adjoint of an operator was defined with a different meaning. This different meaning is now also used to describe the adjoint of a matrix; to avoid conflicting terminology, the matrix $\operatorname{adj} A$ was renamed the *classical adjoint* of $A$, or, more recently, the adjugate of $A$.

[5] Some authors call the determinant of the matrix $A - \lambda I$ the characteristic polynomial.

*Informal proof.* Using equation (3), we have

$$\big(\det{}'(\lambda I - A)\big) I = \big(\operatorname{adj}'(\lambda I - A)\big)(\lambda I - A). \tag{6}$$

The matrix $\operatorname{adj}'(A - \lambda I)$ is a matrix whose entries are polynomials of $\lambda$, and this matrix can be written as a polynomial of $\lambda$ with matrix coefficients. This seems intuitively clear; as an illustration, here is an example $2 \times 2$ matrices showing how to write a matrix of polynomials as a polynomial of matrices.

$$\begin{pmatrix} a_{11}\lambda^2 + b_{11}\lambda + c_{11} & a_{12}\lambda^2 + b_{12}\lambda + c_{12} \\ a_{21}\lambda^2 + b_{21}\lambda + c_{21} & a_{22}\lambda^2 + b_{22}\lambda + c_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \lambda^2 + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \lambda + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}. \tag{7}$$

Now, after such rewriting as polynomials of matrices, if one substitutes $\lambda = A$ in equation (6), one indeed gets 0, since the second factor on the right-hand side is indeed 0. Noting that the coefficients of the of the polynomials in equation (6) are matrices, and matrix multiplication is not commutative, the evaluation on the right-hand side needs to invoke Lemma 1. □

# 5 A formal restatement of the proof

The informal version of the proof of the Cayley–Hamilton Theorem is not up to formal standards: a matrix of polynomials is not a polynomial of matrices, and to say that it can be written as a polynomial of matrices is somewhat vague. Further, instead of substituting $A$ for $\lambda$, it is more rigorous to refer to the evaluation operator. This is especially so because one might object that one can only substitute scalars (i.e., elements of $R$) for $\lambda$ directly in the determinant on the left-hand side of (6); how come we are allowed to substitute a matrix on the right-hand side when rewritten as a polynomial?

## 5.1 Isomorphism between matrices of polynomials and polynomials of matrices; an informal description

Equation (7) gave an example as to how a matrix of polynomials can be written as a polynomial of matrices. More generally, an $n \times n$ matrix of polynomials in $R[\lambda]$ can informally be written as

$$\Big(\sum_{k=0}^{m} a_{ij,k}\lambda^k\Big)_{i,j},$$

where $a_{ij,k} \in R$, $m$ is the maximum degree of the polynomial entries of this matrix, and $1 \le i, j \le n$ are the row and column indices of the matrix. This matrix can be written as

$$\Big(\sum_{k=0}^{m} a_{ij,k}\lambda^k\Big)_{i,j} = \sum_{k=1}^{m} (a_{ij,k})_{i,j}\ \lambda^k = \sum_{k=1}^{m} (a_{ij,k})_{i,j}\ (\lambda I)^k,$$

where $I$ is the $n \times n$ identity matrix. On the right-hand side we replaced $\lambda$ with $\lambda I$ so that only the ring operations will be used for matrices (i.e., only matrix addition and matrix multiplication will be used, and no scalar multiplications of matrices except in forming $\lambda I$). The isomorphism $\phi$ from matrices of polynomials to polynomials of matrices will be obtained, informally, by replacing $\lambda I$ with the formal variable $\Lambda$; that is,

$$\phi'\Big(\sum_{k=0}^{m} a_{ij,k}\lambda^k\Big)_{i,j} = \phi'\Big(\sum_{k=1}^{m} (a_{ij,k})_{i,j}\ (\lambda I)^k\Big) = \sum_{k=1}^{m} (a_{ij,k})_{i,j}\ \Lambda^k.$$

The mapping $\phi$ is one-to-one, since from the polynomial on the right-hand side one can reconstruct the matrix of polynomials in the argument of $\phi$ on the left-hand side. It is also easy to show that for two matrices $P$ and $Q$ of polynomials we have $\phi`(P+Q) = (\phi`P)+(\phi`Q)$ and $\phi`(PQ) = (\phi`P)(\phi`Q)$,[6] showing that the mapping $\phi$ is indeed an isomorphism.

## 5.2 Isomorphism between the matrix of polynomials and polynomials of matrices; the formal description

We will give a formal description of the isomorphism $\phi$ from the ring $R[\lambda]_{n\times n}$ of $n \times n$ matrices over $R[\lambda]$, and the ring $R_{n\times n}[\Lambda]$ of polynomials over the ring $R_{n\times n}$ of matrices over $R$ described above informally. To this end, we need the formal descriptions of a matrices in Subsection 3.1 and of polynomials given in Definition 2. According to this, a matrix of polynomials is a function $A : \{i : 1 \le i \le n\} \times \{j : 1 \le j \le n\} \to R[\lambda]$, and given $i$ and $j$ with $1 \le i, j \le n$, $A`(i, j)$ is a function $A`(i, j) : \mathbf{N} \to R$. On the other hand, a polynomial of matrices as described is a function $p : \mathbb{N} \to R_{n\times n}$, and given $k \in \mathbb{N}$, $p`k$ is a function $p`k : \{i : 1 \le i \le n\} \times \{j : 1 \le j \le n\} \to R$. The isomorphism $\phi$ will be defined such that $\phi`A = p$ is and only if, for any $i$, $j$ with $1 \le i, j \le n$ and for any $k \in \mathbb{N}$ we have

$$(A`(i, j))`k = (p`k)`(i, j) \ \ (\in R).$$

Finally, to restrict this mapping to polynomials only rather than formal power series, we need to stipulate that $(A`(i, j)`k = 0$ except for finitely many $k$'s.

Writing a function $f$ with domain $\mathbb{N}$ as a sequence

$$(f`0, f`1, f`2, \ldots),$$

the formal variable of polynomials over $R$ will be the sequence

$$(0, 1, 0, 0, \ldots),$$

whereas the formal variable of polynomials over the ring $R_{n,n}$ of matrices will be the

$$(0I, I, 0I, 0I, \ldots),$$

where we wrote $0I$ for the zero matrix, to distinguish it from the element 0 of $R$. If we call the former formal variable $\lambda$, the latter formal variable needs a new name; we will call it $\Lambda$.[7]

## 5.3 The formal proof of the Cayley–Hamilton Theorem

Next we will restate Theorem 1 and we will complete its proof. The formal proof will be followed by an example in the next section, where the delicate steps of the proof are illustrated.

**Theorem 2** (Cayley–Hamilton Theorem; formal version)**.** *Let $n$ be a positive integer, let $R$ be a commutative ring, and let $A$ be an $n \times n$ matrix over $R$. Then $A$ is a zero of its own characteristic polynomials; that is,*

$$\mathrm{ev}`\Big(\phi`\big(\det`(\lambda I - A)I\big), A\Big) = 0 \tag{8}$$

[6] Informally, the second equation holds since the formal variable $\Lambda$ commutes with matrices.

[7] One might be tempted to write $\Lambda = \lambda I$. This, however, would create the false impression that, when using Lemma 1, we can evaluate a polynomial in $R_{n,n}[\Lambda]$ only for an $aI$ with $a \in R$. In actual fact, we can evaluate a polynomial in $R_{n,n}[\Lambda]$ for any $A \in R[n, n]$.

*The formal proof of the Cayley–Hamilton Theorem.* Under the isomorphism $\phi$ described above, equation (6) becomes

$$\phi\text{`}\Big(\big(\det\text{`}(\lambda I - A)\big)I\Big) = \phi\text{`}\Big(\big(\operatorname{adj}\text{`}(\lambda I - A)(\lambda I - A)\big)I\Big) = \phi\text{`}\big(\operatorname{adj}\text{`}(\lambda I - A)\big)\ \phi\text{`}(\lambda I - A) \tag{9}$$

The left-hand side here is the characteristic polynomial, a polynomial with scalar coefficients (i.e., the coefficients are of form $aI$ with $a \in R$). The first factor on the right-hand side is a polynomial with matrix coefficients, and the second factor is the polynomial $\phi\text{`}(\lambda I - A) = \Lambda - A$. Applying the evaluation operator with $\Lambda = A$ to this latter polynomial we clearly have

$$\operatorname{ev}\text{`}\big(\phi\text{`}(\lambda I - A), A\big) = \operatorname{ev}\text{`}(\Lambda I - A), A) = A - A = 0.$$

Therefore, noting that the coefficients of this polynomial commute with $A$ (since the only coefficients are $I$ and $-A$), by Lemma 1 we have

$$\operatorname{ev}\text{`}\bigg(\phi\text{`}\Big(\big(\det\text{`}(\lambda I - A)\big)I\Big), A\bigg) = \operatorname{ev}'\Big(\phi\text{`}\big(\operatorname{adj}\text{`}(\lambda I - A)\big), A\Big)\ \operatorname{ev}\text{`}\big(\phi\text{`}(\lambda I - A), A\big) = 0,$$

which is the assertion of the Cayley–Hamilton theorem. □

# 6 An example

As an illustration, we will follow through the steps of the proof above with an example. We believe that it is not necessary for the reader to check the calculations in detail to benefit from this. The example should be helpful simply by allowing the reader to visualize the concrete forms of the abstract formulas above. Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & -1 \\ -2 & 4 & 2 \end{pmatrix}$$

be a matrix over the ring of integers. Writing $I$ for the $3 \times 3$ identity matrix, we have

$$\lambda I - A = \begin{pmatrix} \lambda - 1 & -2 & -1 \\ -3 & \lambda - 2 & 1 \\ 2 & -4 & \lambda - 2 \end{pmatrix}.$$

The characteristic polynomial of $A$ is

$$\det\text{`}(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -2 & -1 \\ -3 & \lambda - 2 & 1 \\ 2 & -4 & \lambda - 2 \end{vmatrix} = \lambda^3 - 5\lambda^2 + 8\lambda - 16.$$

For the adjugate matrix we have

$$\begin{aligned} \operatorname{adj}\text{`}(\lambda I - A) &= \begin{pmatrix} \lambda^2 - 4\lambda + 8 & 2\lambda & \lambda - 4 \\ 3\lambda - 4 & \lambda^2 - 3\lambda + 4 & -\lambda + 4 \\ -2\lambda + 16 & 4\lambda - 8 & \lambda^2 - 3\lambda - 4 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \lambda^2 + \begin{pmatrix} -4 & 2 & 1 \\ 3 & -3 & -1 \\ -2 & 4 & -3 \end{pmatrix} \lambda + \begin{pmatrix} 8 & 0 & -4 \\ -4 & 4 & 4 \\ 16 & -8 & -4 \end{pmatrix}. \end{aligned} \tag{10}$$

Equation (6) then becomes

$$(\lambda^3 - 5\lambda^2 + 8\lambda - 16)I$$
$$= \left( \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \lambda^2 + \begin{pmatrix} -4 & 2 & 1 \\ 3 & -3 & -1 \\ -2 & 4 & -3 \end{pmatrix} \lambda + \begin{pmatrix} 8 & 0 & -4 \\ -4 & +4 & 4 \\ 16 & -8 & -4 \end{pmatrix} \right) (\lambda I - A).$$

Applying the isomorphism $\phi$ to both sides of this equation, we obtain

$$I\Lambda^3 - 5I\Lambda^2 + 8I\Lambda - 16I$$
$$= \left( \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \Lambda^2 + \begin{pmatrix} -4 & 2 & 1 \\ 3 & -3 & -1 \\ -2 & 4 & -3 \end{pmatrix} \Lambda + \begin{pmatrix} 8 & 0 & -4 \\ -4 & +4 & 4 \\ 16 & -8 & -4 \end{pmatrix} \right) (\Lambda - A).$$

Note that the first factor on the right-hand side cannot be written as a single matrix in the middle member of (10) with $\Lambda$ replacing $\lambda$, since, intuitively, the formal variable $\lambda$ is a scalar while $\Lambda$ is a matrix. Substituting $\Lambda = A$ in this equation, we obtain that $A^3 - 5IA^2 + 8IA - 16I = 0$, i.e., that

$$A^3 - 5A^2 + 8A - 16I = 0.$$

# 7 Final comments

## 7.1 The adjugate polynomial also commutes

In using Lemma 1 to prove Theorem 2, we needed to make sure that in equation (9) the coefficients of the second factor, i.e., that of $\phi\text{`}(\lambda I - A) = \Lambda - A$ commute with $A$. This is of course obvious, the only coefficients being $I$ and $-A$. While it was not needed in the proof, it is easy to establish that the coefficients of the first factor, $\phi\text{`}\Big(\big(\text{adj}\,\text{`}(\lambda I - A)\big)\Big)$ also commute with $A$. This is because, according to (3) we also have

$$\big(\det\text{`}(\lambda I - A)\big)I = (\lambda I - A)\big(\text{adj}\,\text{`}(\lambda I - A)\big),$$

and so

$$\phi\text{`}\Big(\big(\det\text{`}(\lambda I - A)\big)I\Big) = \phi\text{`}(\lambda I - A)\ \phi\text{`}\Big(\big(\text{adj}\,\text{`}(\lambda I - A)\big)\Big)$$

in addition to equations (6) and (9). Using the last equation together with (9) it easily follows that the coefficients of $\phi\text{`}\Big(\big(\text{adj}\,\text{`}(\lambda I - A)\big)\Big)$ commute with $A$. We will omit the details.

## 7.2 Use versus mention

In the sentences "cat is and animal" and "cat is a three-letter word," the word cat occurs in two different meanings. First it is used the refer to the animal, second the word itself is mentioned. It would be more correct to say that "'cat' is a three-letter word." The source of the confusion is that in the second sentence you use the word cat itself instead of using its name 'cat' (in quotes). Such a confusion is not possible with the first sentence, since one cannot put a cat (the animal itself) in the sentence; one is forced to use the name of the animal. The situation is similar in mathematics. When one writes the equation $3 + 2 = 5$, one does not put the actual number 3 in the equation; one uses its name, which is 3. In other words, one mentions the number 3, i.e., uses its name, and does not use the number itself. Such a distinction is important in some mathematical contexts.

### 7.3 What are polynomials, really?

In Subsection 2.1 we described the formal definition of a polynomial, but a polynomial is really not what was described in that subsection. A polynomial is really an expression, or a *term* in an appropriate language. Given a ring $R$, we can more appropriately define a polynomial as follows. First, we need to give a name to every element of the ring $R$, and we also need to give a name $\lambda$ to a fixed new variable. Then a polynomial is an expression of the form

$$\sum_{k=0}^{n} a_k \lambda^k,$$

where $a_1$, $a_2$, ... $a_n$ are names of elements of the ring $R$. In the definition of polynomials we will also assume that the variable $\lambda$ commutes with every element of the ring $R$. The evaluation operator substitutes the name of an element of $R$ for the variable $\lambda$. Lemma 1 describes conditions under which certain polynomial identities are preserved under such a substitution.

Such a definition more appropriately represents what is meant by a polynomial. In such a definition, polynomials belong to the *syntax* of the language describing the ring. When one studies algebraic properties of polynomials, one needs to introduce an algebraic formalism to study the syntax. This was done in the formal definition of polynomials above in a way that the syntactic aspect of polynomials was only mentioned as an aside.

## References


[1] K. Gödel. *The Consistency of Axiom of Chocie and of the Generalized Continuum-hyposthesis with the Axioms of Set Theory*. Princeton University Press, Princeton, NJ, 1940.

[2] K. Hoffman and R. Kunze. *Linear Algebra*. Prentice Hall, Englewood Cliffs, NY, second edition, 1971. Available free of charge at
`https://docs.google.com/file/d/0B-87D9cxiLfibzN1UTRkWTdHVFk/edit?pli=1`
.

[3] Attila Máté. Introduction to numerical analysis with C programs, August 2013.
`http://www.sci.brooklyn.cuny.edu/~mate/nml/numanal.pdf`
.

[4] Hans Schneider and George Philip Barker. *Matrices and Linear Algebra, 2nd ed.* Dover Publications, New York, 1973.