Cayley-Hamilton Theorem: Let A be an $n \times n$ matrix with entries in a commutative ring R. Then A is the root of monic polynomial in $R[x]$ of degree n.

Proof: We will eventually prove that $p(A) = 0$ where $p(X) = \det(XI - A)$. This requires expanding the determinant expression, evaluating at A, and showing that each of the entries in the resulting matrix polynomial combine to zero. We will give a combinatorial description of these cancellations. We first consider some concepts from graph theory.

A directed graph is a collection of vertices and directed edges. We will regard the numbers $1, 2, \ldots, n$ as vertices. An (i, j)-walk of length k is a series of transitions from a boxed vertex i to a boxed vertex j through k directed edges. For example,

$$w = \boxed{2} \to 3 \to 3 \to 1 \to \boxed{5}$$

is a $(2, 5)$-walk of length 4 from vertex 2 to vertex 5. There are no (i, j)-walks of length 0 when $i \neq j$, and an (i, i)-walk of length 0 consists of the boxed vertex $\boxed{\text{i}}$. The set $W_{i,j}(k)$ contains all (i, j)-walks of length k.

A k-cycle ($k \geq 1$) is a circular arrangement of k vertices and directed edges. For example,

$$c = 2 \to 1 \to 3 \to 2$$

is a 3-cycle and

$$c = 2 \to 2$$

is a 1-cycle. It doesn't matter which vertex the cycle begins with, so a k-cycle has k equivalent representations.

A compound k-cycle is a set of vertex-disjoint cycles with a total of k edges. For example,

$$cc = \{1 \to 2 \to 1, 3 \to 7 \to 4 \to 3, 5 \to 5\}$$

is a compound 6-cycle. We will regard a collection of vertices with no edges as a compound 0-cycle. The set $CC(k)$ contains all compound k-cycles.

Next, some concepts from algebraic combinatorics. Assume that the matrix A has entries a_{ij}. The weight of a walk is the product of the matrix entries corresponding to its edges:

$$\text{weight}(\ \boxed{2} \to 3 \to 3 \to 1 \to \boxed{5}\ ) = a_{23}a_{33}a_{31}a_{15}.$$

The weight of an (i,i)-walk of length 0 is 1:

$$\text{weight}(\boxed{\text{i}}) = 1.$$

The weight of a cycle is minus one times the product of the matrix entries corresponding to its edges:

$$\text{weight}(2 \to 1 \to 3 \to 2) = -a_{21}a_{13}a_{32},$$

$$\text{weight}(2 \to 2) = -a_{22}.$$

The weight of a compound cycle is the product of the weights of its cycles:

$$\text{weight}(\{1 \to 2 \to 1, 3 \to 7 \to 4 \to 3, 5 \to 5\}) = (-1)^3 a_{12}a_{21}a_{37}a_{74}a_{43}a_{55}.$$

The weight of the compound 0-cycle is 1.

Lemma 1: The i,j-entry of A^k is

$$\sum_{w \in W_{i,j}(k)} \text{weight}(w).$$

Proof: For $k \geq 1$, the (i,j)-entry of A^k is

$$\sum_{i_1,i_2,\ldots,i_{k-1}} a_{ii_1}a_{i_1i_2}\cdots a_{i_{k-1}j} =$$

$$\sum_{i_1,i_2,\ldots,i_{k-1}} \text{weight}(\boxed{\text{i}} \to i_1 \to i_2 \to \cdots \to i_{k-1} \to \boxed{\text{j}}) =$$

$$\sum_{w \in W_{i,j}(k)} \text{weight}(w).$$

For $k = 0$ the (i,j)-entry of $A^0 = I$ is 0 or 1 depending on whether $i \neq j$ or $i = j$. This is consistent with

$$\sum_{w \in W_{i,j}(0)} \text{weight}(w)$$

if we interpret the sum of weights over an empty set to be zero (when $i \neq j$).

Definition 2: For each k, $0 \le k \le n$, we define

$$p_k = \sum_{cc \in CC(k)} \text{weight}(cc).$$

Example 3: $CC(0)$ contains only the empty compound cycle, therefore $p_0 = 1$.

Example 4: $CC(1)$ contains only compound cycles of the form $\{i \to i\}$, therefore $p_1 = -\sum_{i=1}^{n} a_{ii}$.

Example 5: $CC(2)$ contains compound cycles of the form $\{i \to i, j \to j\}$ and $\{i \to j \to i\}$ where $i < j$. Therefore

$$p_2 = \sum_{i<j} (a_{ii}a_{jj} - a_{ij}a_{ji}).$$

Theorem 6: With notation as above,

$$\sum_{k=0}^{n} p_{n-k} A^k = 0.$$

Therefore A is a root of the monic degree-n polynomial

$$p(x) = \sum_{k=0}^{n} p_{n-k} x^k.$$

Example 7: Let A be a 2×2 matrix. Then

$$\begin{aligned}
p_2 I &= (a_{11}a_{22} - a_{12}a_{21}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11}a_{22} - a_{12}a_{21} & 0 \\ 0 & a_{11}a_{22} - a_{12}a_{21} \end{bmatrix} \\
p_1 A &= -(a_{11} + a_{22}) \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} -a_{11}^2 - a_{11}a_{22} & -a_{11}a_{12} - a_{12}a_{22} \\ -a_{11}a_{21} - a_{21}a_{22} & -a_{11}a_{22} - a_{22}^2 \end{bmatrix} \\
p_0 A^2 &= \begin{bmatrix} a_{11}^2 + a_{12}a_{21} & a_{11}a_{12} + a_{12}a_{22} \\ a_{21}a_{11} + a_{22}a_{21} & a_{21}a_{12} + a_{22}^2 \end{bmatrix}
\end{aligned}$$

$$p_2 I + p_1 A + p_0 A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

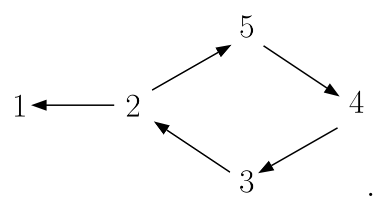
Example 8: Let $P(\mathbb{Z}_5)$ denote the set of polynomial functions from $\mathbb{Z}_5$ to $\mathbb{Z}_5$. Since $x^5 = x$, a spanning set for $P(\mathbb{Z}_5)$ is $\{1, x, x^2, x^3, x^4\}$. These functions are linearly indepenent over $\mathbb{Z}_5$: Suppose $a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 = 0$. Evaluating at 0 through 4 mod 5 we obtain a system of equations which in matrix form is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & 2 & 1 \\ 1 & 4 & 1 & 4 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since the determinant of the coefficient matrix is 3, $a_0 = a_1 = a_2 = a_3 = a_4 = 0$. Now consider the linear operator $T : P(\mathbb{Z}_5) \to P(\mathbb{Z}_5)$ defined by $T(f(x)) = xf(x)$. The matrix representation of T with respect to the basis $\{1, x, x^2, x^3, x^4\}$ is

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

A directed graph representing non-zero edge weights based on the information in A is



The only non-trivial compound cycle through the vertex set $\{1, 2, 3, 4, 5\}$ that contributes a non-zero weight is

$$5 \to 4 \to 3 \to 2 \to 5,$$

which implies that the only non-zero coefficients of $p(x)$ are $p_0 = 1$ and $p_4 = -1$. Hence $p(x) = x^5 - x$.

Proof of Theorem 6: We will make an argument that each of the n^2 entries of $\sum_{k=0}^{n} p_{n-k} A^k$ is equal to 0. By Lemma 1 and Definition 2, the (i, j)-entry of this expression is

$$\sum_{k=0}^{n}\left(\sum_{cc\in CC(n-k)}\text{weight}(cc)\sum_{w\in W_{i,j}(k)}\text{weight}(w)\right).$$

Let $D(i,j)$ be the set of all ordered pairs of the form (cc, w), where cc is an arbitrary compound cycle and w is an (i,j)-walk and the total number of edges contributed by cc and w is equal to n. The sum above can be more simply expressed as

$$\sum_{(cc,w)\in D(i,j)}\text{weight}(cc)\text{weight}(w).$$

We will argue that the terms in this expression can be paired off in such a way that each pair has a sum equal to zero. This will imply that the entire sum is equal to zero.

We will partition $D(i,j)$ into $D_e(i,j)$ and $D_o(i,j)$, where $D_e(i,j)$ is the set of all $(cc,w)\in D(i,j)$ where cc contains an even number of cycles and $D_o(i,j)$ is the set of all $(cc,w)\in D(i,j)$ where cc has contains an odd number of cycles. We will then produce a one-to-one correspondence between $D_e(i,j)$ and $D_o(i,j)$ such that if

$$(cc,w)\leftrightarrow(cc',w')$$

then

$$\text{weight}(cc')\text{weight}(w') = -\text{weight}(cc)\text{weight}(w).$$

This accomplishes the task described in the previous paragraph.

Example 9: Assume $n=12$, $i=10$, $j=4$,

$$\begin{aligned} cc &= \{2\to3\to2, 5\to10\to9\to5, 1\to1\} \\ w &= \boxed{10}\to6\to5\to11\to5\to5\to\boxed{4}. \end{aligned}$$

We will remove the cycle $5\to10\to9\to5$ from cc to create cc' and add this cycle to w to create w'. The result is

$$\begin{aligned} cc' &= \{2\to3\to2, 1\to1\} \\ w' &= \boxed{10}\to6\to5\to10\to9\to5\to11\to5\to5\to\boxed{4}. \end{aligned}$$

Since (cc, w) and (cc', w') have exactly the same collection of edges, and since cc' has one less cycle in it than cc does,

$$\text{weight}(cc', w') = -\text{weight}(cc, w).$$

Example 10: Assume $n = 12$, $i = 10$, $j = 4$,

$$\begin{aligned} cc &= \{5 \to 7 \to 5, 11 \to 9 \to 8 \to 1 \to 11\} \\ w &= \boxed{10} \to 2 \to 3 \to 4 \to 2 \to 5 \to \boxed{4}. \end{aligned}$$

We will remove the cycle $2 \to 3 \to 4 \to 2$ from w to create w' and add this cycle to cc to create cc'. The result is

$$\begin{aligned} cc' &= \{5 \to 7 \to 5, 11 \to 9 \to 8 \to 1 \to 11, 2 \to 3 \to 4 \to 2\} \\ w' &= \boxed{10} \to 2 \to 5 \to \boxed{4}. \end{aligned}$$

Since (cc, w) and (cc', w') have exactly the same collection of edges, and since cc' has one more cycle in it than cc does,

$$\text{weight}(cc', w') = -\text{weight}(cc, w).$$

To resume the proof of Theorem 6, we are going to establish the correspondence between $D_e(i, j)$ and $D_o(i, j)$ as follows: given a pair (cc, w), we are going to exchange a cycle between cc and w to create (cc', w'). This guarantees $\text{weight}(cc')\text{weight}(w') = -\text{weight}(cc)\text{weight}(w)$. Examples 10 and 11 illustrate the difficulties to overcome using this approach: How do you decide whether to take a cycle from cc and add it to w or vice versa? Which cycle do you choose? When you take a cycle from w and add it to cc, you must be careful not to create overlapping cycles in cc'. Finally, how do you know that this cycle-transfer technique creates a one-to-one correspondence?

The construction we will describe is based on the following two observations:

Observation 1. Given (cc, w) with a total of n edges, if none of the vertices in w appear more than once then w and cc share a vertex. Reason: If cc

and w have no vertex in common, then if cc has k edges and w has $n-k$ edges then cc has k vertices and w has $n-k+1$ vertices, giving rise to $n+1$ distinct vertices. This is not possible, given that there are only n vertices.

Observation 2. Given any (cc, w) with n edges, examine each vertex x in w in the order it appears along the path. Label it with C if it appears in cc and label it with D if it doesn't appear in cc. Label it with F it is appearing for the first time in w and label it with G if it is not appearing for the first time in w. Then every vertex receives one of the four compound labels CF, DF, CG, DG. By Observation 1, the labels cannot all be DF. Let x be the first vertex along w that receives a label in $\{CF, CG, DG\}$. Then x cannot have the label CG because it is not being encountered for the first time in w. So in fact x receives a label in $\{CF, DG\}$, and the labels along W through x form one of two sequences: $DF, DF, \ldots, DF, CF$ or $DF, DF, \ldots, DF, DG$.

Now let $D_{CF}(i,j)$ be the set of those (cc, w) where x receives the label CF and let $D_{DG}(i,j)$ be the set of those (cc, w) where x receives the label DG. These two sets form a partition of $D(i,j)$. Given $(cc, w) \in D_{CF}(i,j)$, we remove the cycle in cc containing x and insert it into w to create $(cc', w') \in D_{DG}(i,j)$ as in Example 10. Given $(cc, w) \in D_{DG}(i,j)$, we remove the cycle in w in which x is visited for the first and second time and add it to cc to create $(cc', w') \in D_{CF}(i,j)$ as in Example 11. So we define $f : D(i,j) \to D(i,j)$ by $f(cc, w) = (cc', w')$, where we apply the appropriate cycle transfer between cc and w according to whether $(cc, w) \in D_{CF}(i,j)$ or $(cc, w) \in D_{DG}(i,j)$. Examples 10 and 11 were generated using this rule. It is not difficult to verify that $f \circ f$ is the identity map. Therefore f is both injective and surjective and establishes a one-to-one correspondence between $D_e(i,j)$ and $D_o(i,j)$. It also has the desired sign-reversing property. (In algebraic combinatorics we call f a sign-reversing involution.) This completes the proof of Theorem 6, hence of the Cayley-Hamilton Theorem.

Theorem 11: Let A be an $n \times n$ matrix with entries a_{ij}. For each k, $0 \le k \le n$, let $CC(k)$ be the set of k compound cycles on the vertex set $\{1, 2, \ldots, n\}$ and define

$$p_k = \sum_{cc \in CC(k)} \text{weight}(cc).$$

Then

$$\det(xI - A) = \sum_{k=0}^{n} p_{n-k} x^k.$$

Proof: For each permutation σ we define the sets $D(\sigma)$ and $F(\sigma)$ via

$$D(\sigma) = \{i : \sigma(i) \neq i\}$$

and

$$F(\sigma) = \{i \in \sigma(i) : \sigma(i) = i\}.$$

Then

$$\det(xI - A) = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i \in D(\sigma)} (-a_{i\sigma(i)}) \prod_{i \in F(\sigma)} (x - a_{ii}).$$

Given that the sign of a k-cycle is $(-1)^{k-1}$, for any permutation σ we have $\operatorname{sgn}(\sigma) = (-1)^{|D(\sigma)|+c(\sigma)}$ where $c(\sigma)$ is the number of non-trivial cycles in σ. Therefore

$$\det(xI - A) = \sum_{\sigma} (-1)^{c(\sigma)} \prod_{i \in D(\sigma)} a_{i\sigma(i)} \prod_{i \in F(\sigma)} (x - a_{ii}).$$

For any subset S of $[n]$, we have

$$\prod_{i \in S} (x - a_{ii}) = \sum_{I \subseteq S} x^{|I|} \prod_{i \in S \setminus I} (-a_{ii}).$$

With this substitution we have

$$\det(xI - A) = \sum_{\sigma} (-1)^{c(\sigma)} \prod_{i \in D(\sigma)} a_{i\sigma(i)} \sum_{I \subseteq F(\sigma)} x^{|I|} \prod_{i \in F(\sigma) \setminus I} (-a_{ii}).$$

Therefore the coefficient of x^k in $\det(xI - A)$ is

$$\sum_{\sigma} \sum_{\substack{I \subseteq F(\sigma) \\ |I| = k}} (-1)^{c(\sigma)} \prod_{i \in D(\sigma)} a_{i\sigma(i)} \prod_{i \in F(\sigma) \setminus I} (-a_{ii}).$$

Note that

$$(-1)^{c(\sigma)} \prod_{i \in D(\sigma)} a_{i\sigma(i)} \prod_{i \in F(\sigma) \setminus I} (-a_{ii})$$

is the weight of the compound cycle with edge set

$$\{i \to \sigma(i) : i \in I^c\}.$$

Let's give this compound cycle the name $cc(\sigma, I)$. As I ranges through all subsets of $F(\sigma)$ of size k, $cc(\sigma, I)$ ranges through all compound cycles in $CC(n-k)$ whose non-trivial cycles are the same as those in σ. Letting σ vary we produce all compound cycles in $CC(n-k)$. Therefore the coefficient of x^k in $\det(xI - A)$ is

$$\sum_{\sigma} \sum_{\substack{I \subseteq F(\sigma) \\ |I|=k}} \text{weight}(cc(\sigma, I)) = \sum_{cc \in C(n-k)} \text{weight}(cc) = p_{n-k}.$$

Comment 12: The interested reader can prove for himself that for $k > 0$ we have

$$p_k = \sum_{\substack{I \subseteq [n] \\ |I|=k}} \det(-A_I)$$

where $-A_I$ denotes the submatrix of $-A$ obtained by using entries from rows and columns in I. Hence A satisfies the polynomial

$$p(x) = x^n + \sum_{\emptyset \neq I \subseteq [n]} \det(-A_I) x^{n-|I|}.$$