# The Cayley-Hamilton theorem

We've seen a lot of concepts:

- Eigenvectors and eigenvalues
- Eigenbasis
- Characteristic polynomial
- Minimal polynomial

We've seen a lot of concepts:

- Eigenvectors and eigenvalues
  A vector $\vec{v}$ with $A\vec{v} = \lambda\vec{v}$ is called an eigenvector. We call $\lambda$ the eigenvalue.

- Eigenbasis
  A basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ of eigenvectors. If $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are nonzero eigenvectors with distinct eigenvalues, they must be an eigenbasis.

- Characteristic polynomial
  The polynomial $\chi_A(x) = \det(x\mathrm{Id} - A)$. Its roots are the eigenvalues.

- Minimal polynomial
  The lowest degree polynomial $m(x)$ with $m(A) = 0$. If $f(x)$ is any other polynomial with $f(A) = 0$, then $m(x)$ divides $f(x)$.

Let $V$ be a finite dimensional vector space, $A : V \to V$ a linear transformation, $\chi_A(x)$ the characteristic polynomial and $m_A(x)$ the minimal polynomial.

**Theorem:** Let $\lambda$ be a scalar. Then the following are equivalent:

(1) $\lambda$ is an eigenvalue of $A$.

(2) $\lambda$ is a root of $\chi_A(x)$.

(3) $\lambda$ is a root of $m_A(x)$.

**Question:** Which parts have we already proved? How did we prove them?

**Theorem:** Let $\lambda$ be a scalar. Then the following are equivalent:

(1) $\lambda$ is an eigenvalue of $A$.

(2) $\lambda$ is a root of $\chi_A(x)$.

(3) $\lambda$ is a root of $m_A(x)$.

(1) $\iff$ (2): We have $\chi_A(\lambda) = 0$ if and only if $\det(A - \lambda \mathrm{Id}) \neq 0$ if and only if there is a vector $\vec{v}$ with $(A - \lambda \mathrm{Id})\vec{v} = \vec{0}$ if and only if there is a vector $\vec{v}$ with $A\vec{v} = \lambda\vec{v}$.

**Theorem:** Let $\lambda$ be a scalar. Then the following are equivalent:

(1) $\lambda$ is an eigenvalue of $A$.

(2) $\lambda$ is a root of $\chi_A(x)$.

(3) $\lambda$ is a root of $m_A(x)$.

(1) $\iff$ (2): We have $\chi_A(\lambda) = 0$ if and only if $\det(A - \lambda \mathrm{Id}) \neq 0$ if and only if there is a vector $\vec{v}$ with $(A - \lambda \mathrm{Id})\vec{v} = \vec{0}$ if and only if there is a vector $\vec{v}$ with $A\vec{v} = \lambda\vec{v}$.

(1) $\implies$ (3): Let $\vec{v}$ be a nonzero $\lambda$-eigenvector. For any polynomial $f(x)$, we have $f(A)\vec{v} = f(\lambda)\vec{v}$. In particular, $m_A(A)\vec{v} = m_A(\lambda)\vec{v}$. But $m_A(A)\vec{v} = 0\vec{v} = \vec{0}$, so $m_A(\lambda) = 0$.

(3) $\implies$ (1): Since $\lambda$ is a root of $m_A(x)$, we can write $m_A(x) = (x - \lambda)f(x)$. Suppose that $\lambda$ is not an eigenvalue of $A$. Then $A - \lambda\mathrm{Id}$ is invertible. We have $m_A(A) = (A - \lambda\mathrm{Id})f(A) = 0$. So $f(A) = (A - \lambda\mathrm{Id})^{-1}0 = 0$. But $\deg f < \deg m_A$, a contradiction.

**Theorem:** Let $\lambda$ be a scalar. Then the following are equivalent:

(1) $\lambda$ is an eigenvalue of $A$.

(2) $\lambda$ is a root of $\chi_A(x)$.

(3) $\lambda$ is a root of $m_A(x)$.

So $m_A(x)$ and $\chi_A(x)$ have the same roots, but maybe with different multiplicities. Our main result for today improves (3) $\implies$ (2):

**The Cayley-Hamilton theorem** We have $\chi_A(A) = 0$ and, therefore, $m_A(x)$ divides $\chi_A(x)$.

**The Cayley-Hamilton theorem** We have $\chi_A(A) = 0$ and, therefore, $m_A(x)$ divides $\chi_A(x)$.

**The Cayley-Hamilton theorem** We have $\chi_A(A) = 0$ and, therefore, $m_A(x)$ divides $\chi_A(x)$.

Let's start with an example, to make it clear what we are proving. Let $A = \left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]$. The characteristic polynomial is $(x-1)(x-4) - 2 \cdot 3 = x^2 - 5x + 4 - 6 = x^2 - 5x - 2$.

We have

$$\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right]^2 - 5\left[\begin{smallmatrix} 1 & 2 \\ 3 & 4 \end{smallmatrix}\right] - 2\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 7 & 10 \\ 15 & 22 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 5 & 10 \\ 15 & 20 \end{smallmatrix}\right] - \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right].$$

For diagonalizable matrices:

Suppose that $A = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$.

Then $\chi_A(x) = \det(x\mathrm{Id} - A) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$.

For each eigenvector $\vec{v_i}$, we have $(A-\lambda_1)(A-\lambda_2)\cdots(A-\lambda_n)\vec{v}_i = (\lambda_i-\lambda_1)(\lambda_i-\lambda_2)\cdots(\lambda_i-\lambda_i)\cdots(\lambda_i-\lambda_n)\vec{v}_i = 0\vec{v}_i = \vec{0}$. So $\chi_A(A)\vec{v}_i = \vec{0}$

## Another important case

Suppose that there is a vector $\vec{v}$ such that $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, $\ldots$, $A^{n-1}\vec{v}$ is a basis of $V$. In this case, we must have

$$A^n\vec{v} = c_0\vec{v} + c_1 A\vec{v} + c_2 A^2\vec{v} + \cdots + c_{n-1}A^{n-1}\vec{v}.$$

**Question:** What is the matrix of $A$ in the basis $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, $\ldots$, $A^{n-1}\vec{v}$?

## Another important case

Suppose that there is a vector $\vec{v}$ such that $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, $\ldots$, $A^{n-1}\vec{v}$ is a basis of $V$. In this case, we must have

$$A^n\vec{v} = c_0\vec{v} + c_1 A\vec{v} + c_2 A^2\vec{v} + \cdots + c_{n-1}A^{n-1}\vec{v}.$$

**Question:** What is the matrix of $A$ in the basis $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, $\ldots$, $A^{n-1}\vec{v}$?

$$\begin{bmatrix} & & & & & c_0 \\ 1 & & & & & c_1 \\ & 1 & & & & c_2 \\ & & \ddots & & & \vdots \\ & & & 1 & & c_{n-2} \\ & & & & 1 & c_{n-1} \end{bmatrix}$$

What is the characteristic polynomial?

We expand along the top row:

$$\det \begin{bmatrix} x & & & & & -c_0 \\ -1 & x & & & & -c_1 \\ & -1 & x & & & -c_2 \\ & & \ddots & \ddots & & \vdots \\ & & & -1 & x & -c_{n-2} \\ & & & & -1 & x-c_{n-1} \end{bmatrix} =$$

$$x \det \begin{bmatrix} x & & & & -c_1 \\ -1 & x & & & -c_2 \\ & \ddots & \ddots & & \vdots \\ & & -1 & x & -c_{n-2} \\ & & & -1 & x-c_{n-1} \end{bmatrix} + (-1)^n c_0 \det \begin{bmatrix} -1 & x & & & \\ & -1 & x & & \\ & & \ddots & \ddots & \\ & & & -1 & x \\ & & & & -1 \end{bmatrix} =$$

$$x \det \begin{bmatrix} x & & & & -c_1 \\ -1 & x & & & -c_2 \\ & \ddots & \ddots & & \vdots \\ & & -1 & x & -c_{n-2} \\ & & & -1 & x-c_{n-1} \end{bmatrix} - c_0.$$

## What is the characteristic polynomial?

$$\det\begin{bmatrix} x & & & & & -c_0 \\ -1 & x & & & & -c_1 \\ & -1 & x & & & -c_2 \\ & & \ddots & \ddots & & \vdots \\ & & & -1 & x & -c_{n-2} \\ & & & & -1 & x-c_{n-1} \end{bmatrix} =$$

$$x\det\begin{bmatrix} x & & & & -c_1 \\ -1 & x & & & -c_2 \\ & \ddots & \ddots & & \vdots \\ & & -1 & x & -c_{n-2} \\ & & & -1 & x-c_{n-1} \end{bmatrix} + (-1)^n c_0 \det\begin{bmatrix} -1 & x & & & \\ & -1 & x & & \\ & & \ddots & \ddots & \\ & & & -1 & x \\ & & & & -1 \end{bmatrix} =$$

$$x\det\begin{bmatrix} x & & & & -c_1 \\ -1 & x & & & -c_2 \\ & \ddots & \ddots & & \vdots \\ & & -1 & x & -c_{n-2} \\ & & & -1 & x-c_{n-1} \end{bmatrix} - c_0.$$

Continuing:

$$x^2\det\begin{bmatrix} x & & -c_2 \\ \ddots & \ddots & \vdots \\ & -1 \; x & -c_{n-2} \end{bmatrix} - c_1 x - c_0 = \cdots$$

$$= x^n - c_{n-1}x^{n-1} - \cdots - c_2 x^2 - c_1 x - c_0.$$

To summarize the previous computation: If there is a vector $\vec{v}$ such that $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, …, $A^{n-1}\vec{v}$ is a basis of $V$, and

$$A^n\vec{v} = c_0\vec{v} + c_1A\vec{v} + c_2A^2\vec{v} + \cdots + c_{n-1}A^{n-1}\vec{v}$$

then the characteristic polynomial of $A$ is

$$x^n - c_{n-1}x^{n-1} - \cdots - c_2x^2 - c_1x - c_0.$$

To summarize the previous computation: If there is a vector $\vec{v}$ such that $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, …, $A^{n-1}\vec{v}$ is a basis of $V$, and

$$A^n\vec{v} = c_0\vec{v} + c_1A\vec{v} + c_2A^2\vec{v} + \cdots + c_{n-1}A^{n-1}\vec{v}$$

then the characteristic polynomial of $A$ is

$$x^n - c_{n-1}x^{n-1} - \cdots - c_2x^2 - c_1x - c_0.$$

So we want to show that

$$A^n - c_{n-1}A^{n-1} - \cdots - c_2A^2 - c_1A - c_0 = 0.$$

To summarize the previous computation: If there is a vector $\vec{v}$ such that $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, ..., $A^{n-1}\vec{v}$ is a basis of $V$, and

$$A^n\vec{v} = c_0\vec{v} + c_1 A\vec{v} + c_2 A^2\vec{v} + \cdots + c_{n-1}A^{n-1}\vec{v}$$

then the characteristic polynomial of $A$ is

$$x^n - c_{n-1}x^{n-1} - \cdots - c_2 x^2 - c_1 x - c_0.$$

So we want to show that

$$A^n - c_{n-1}A^{n-1} - \cdots - c_2 A^2 - c_1 A - c_0 = 0.$$

We just need to check that $A^n - c_{n-1}A^{n-1} - \cdots - c_2 A^2 - c_1 A - c_0$ kills each basis vector. We certainly have

$$\left(A^n - c_{n-1}A^{n-1} - \cdots - c_2 A^2 - c_1 A - c_0\right)\vec{v} = \\ A^n\vec{v} - c_{n-1}A^{n-1}\vec{v} - \cdots - c_2 A^2\vec{v} - c_1 A\vec{v} - c_0\vec{v} = \vec{0}.$$

We just need to check that $A^n - c_{n-1}A^{n-1} - \cdots - c_2A^2 - c_1A - c_0$ kills each basis vector. We certainly have

$$\left(A^n - c_{n-1}A^{n-1} - \cdots - c_2A^2 - c_1A - c_0\right)\vec{v} = \\ A^n\vec{v} - c_{n-1}A^{n-1}\vec{v} - \cdots - c_2A^2\vec{v} - c_1A\vec{v} - c_0\vec{v} = \vec{0}.$$

Working a bit harder,

$$\left(A^n - c_{n-1}A^{n-1} - \cdots - c_2A^2 - c_1A - c_0\right)(A^j\vec{v}) = \\ A^j\left(A^n - c_{n-1}A^{n-1} - \cdots - c_2A^2 - c_1A - c_0\right)\vec{v} = A^j\vec{0} = \vec{0}.$$

Why did we spend so long on this special case? Because it is going to be the heart of our proof of the general case. We now prove the Cayley-Hamilton theorem by induction on $n$.

We now prove the Cayley-Hamilton theorem by induction on $n$.

The base case is $n = 1$, so $A = [\,a\,]$, the characteristic polynomial is $x - a$, and we have $A - a\mathrm{Id} = 0$ as desired.

We now prove the Cayley-Hamilton theorem by induction on $n$.

The base case is $n = 1$, so $A = [\,a\,]$, the characteristic polynomial is $x - a$, and we have $A - a\mathrm{Id} = 0$ as desired.

Now, choose any nonzero vector $\vec{v}$ and start computing $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$ until the first point where we have a linear dependency

$$A^m\vec{v} = c_{m-1}A^{m-1}\vec{v} + \cdots + c_2A^2\vec{v} + c_1A\vec{v} + c_0\vec{v}.$$

Completing $\vec{v}$, $A\vec{v}$, $A^2\vec{v}$, $\dots$, $A^{m-1}\vec{v}$ to a basis, the operator $A$ looks like

$$\begin{bmatrix} P & Q \\ 0 & R \end{bmatrix} \text{ where } P = \begin{bmatrix} & & & & c_0 \\ 1 & & & & c_1 \\ & 1 & & & c_2 \\ & & \ddots & & \vdots \\ & & & 1 & c_{m-2} \\ & & & & 1\ c_{m-1} \end{bmatrix}.$$

$$A = \begin{bmatrix} P & Q \\ 0 & R \end{bmatrix} \text{ where } P = \begin{bmatrix} & & & & & c_0 \\ 1 & & & & & c_1 \\ & 1 & & & & c_2 \\ & & \ddots & & & \vdots \\ & & & 1 & & c_{m-2} \\ & & & & 1 & c_{m-1} \end{bmatrix}.$$

By our previous computation, $\chi_P(P) = 0$. If $P$ is the whole matrix $A$, we are done.

$$A = \begin{bmatrix} P & Q \\ 0 & R \end{bmatrix} \text{ where } P = \begin{bmatrix} & & & & & c_0 \\ 1 & & & & & c_1 \\ & 1 & & & & c_2 \\ & & \ddots & & & \vdots \\ & & & 1 & & c_{m-2} \\ & & & & 1 & c_{m-1} \end{bmatrix}.$$

By our previous computation, $\chi_P(P) = 0$. If $P$ is the whole matrix $A$, we are done.

If not, by induction, we also have $\chi_R(R) = 0$. And we have $\chi_A(x) = \chi_P(x)\chi_R(x)$. So

$$\chi_A(A) = \chi_A\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right) = \chi_P\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right)\chi_R\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right) =$$
$$\left[\begin{smallmatrix} \chi_P(P) & * \\ 0 & * \end{smallmatrix}\right]\left[\begin{smallmatrix} * & * \\ 0 & \chi_R(R) \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & * \\ 0 & * \end{smallmatrix}\right]\left[\begin{smallmatrix} * & * \\ 0 & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right].$$

$$A = \begin{bmatrix} P & Q \\ 0 & R \end{bmatrix} \text{ where } P = \begin{bmatrix} & & & & c_0 \\ 1 & & & & c_1 \\ & 1 & & & c_2 \\ & & \ddots & & \vdots \\ & & & 1 & c_{m-2} \\ & & & 1 & c_{m-1} \end{bmatrix}.$$

By our previous computation, $\chi_P(P) = 0$. If $P$ is the whole matrix $A$, we are done.

If not, by induction, we also have $\chi_R(R) = 0$. And we have $\chi_A(x) = \chi_P(x)\chi_R(x)$. So

$$\chi_A(A) = \chi_A\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right) = \chi_P\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right)\chi_R\left(\left[\begin{smallmatrix} P & Q \\ 0 & R \end{smallmatrix}\right]\right) =$$

$$\left[\begin{smallmatrix} \chi_P(P) & * \\ 0 & * \end{smallmatrix}\right]\left[\begin{smallmatrix} * & * \\ 0 & \chi_R(R) \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & * \\ 0 & * \end{smallmatrix}\right]\left[\begin{smallmatrix} * & * \\ 0 & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right].$$

**QED**