A Cayley–Hamilton tétel

Az alábbiakban a Cayley–Hamilton tételre adunk négy (részben vázlatos) bizonyítást.

1. Tétel (Cayley–Hamilton)**.** *Legyen $A \in K^{n\times n}$ egy négyzetes mátrix és $k_A(x) \in K[x]$ a karakterisztikus polinomja. Ekkor $k_A(A) = 0$. Speciálisan $m_A(x) \mid k_A(x)$, ahol $m_A(x)$ a minimálpolinom.*

Jöjjön először a klasszikus *hibás* bizonyítás.

Hibás bizonyítás. Mivel $k_A(x) = \det(A - xI)$, itt x helyére A-t írva $k_A(A) = \det(A - AI) = \det 0 = 0$. □

A fenti bizonyítással az a gond, hogy x helyére A-t írva $A - xI$-ben nem a 0 mátrixot kapjuk, hanem azt, aminek elemei mátrixok: pl. a bal felső elem $a_{11}I - A$. Nem világos, hogy ilyennek van-e egyáltalán determinánsa és ha igen, akkor az micsoda (szakértőknek: valami értelme van, a determináns nemkommutatív gyűrűk fölötti mátrixokra való általánosítása az algebrai K-elméletben a K_1). Ugyanakkor az alábbi bizonyítás lényegében ezt a trükköt teszi precízzé.

1. Bizonyítás. A következő lemma önmagában is hasznos, tanulmányaink során többször is fogunk még vele találkozni (pl. az algebrai egészek témakörében).

2. Lemma. *Legyen R egy kommutatív egységelemes gyűrű és $M \in R^{n\times n}$ egy R-beli együtthatós négyzetes mátrix. Ekkor $M\widehat{M} = \widehat{M}M = \det(M) \cdot I$, ahol I az egységmátrix, $\widehat{M}$ pedig az a mátrix, melyben az i-edik sor j-edik eleme az M mátrix j-edik sorához és i-edik oszlopához tartozó előjeles aldeterminánsa.*

Bizonyítás. Ez nem más, mint a (ferde) kifejtési tétel (inverz mátrix képlete). Azt kell csak meggondolni, hogy ez akkor is érvényes, ha az együtthatók nem testből valók. □

Legyen $M := A - xI \in K[x]^{n\times n}$ polinomelemű mátrix. Ekkor definíció szerint $\det M = k_A(x) \in K[x]$ a karakterisztikus polinom. A fenti lemma szerint $M\widehat{M} = k_A(x)I$, ezt fejtjük ki az alábbiakban: $\widehat{M}$ egy olyan mátrix, melynek minden eleme egy legfeljebb $n-1$-edfokú polinom. Csoportosítva az elemeket x hatványai szerint $\widehat{M} = B_0 + B_1x + \cdots + B_{n-1}x^{n-1}$ alakba írható alkalmas *konstans* $B_0, \ldots, B_{n-1} \in K^{n\times n}$ mátrixokkal. Ha $k_A(x) = a_0 + a_1x + \cdots + a_nx^n$, akkor az $M\widehat{M} = k_A(x)I$ egyenlet az

$$(A - xI)(B_0 + B_1x + \cdots + B_{n-1}x^{n-1}) = a_0I + a_1Ix + \cdots + a_nIx^n$$

alakot ölti. Itt x^i együtthatóját $i = 0, \ldots, n$-re összehasonlítva a két oldalon az

$$\begin{aligned} AB_0 &= a_0I \\ AB_1 - B_0 &= a_1I \\ &\vdots \\ -B_{n-1} &= a_nI \end{aligned}$$

egyenleteket kapjuk. Az i-edik egyenletet balról megszorozva A^i-nel és az összeset összeadva a $0 = k_A(A)$ azonosságot kapjuk (a bal oldalon teleszkopikus összeg keletkezik). □

A fenti bizonyításnak az az előnye, hogy könnyű leírni, lépésről lépésre megérteni és megtanulni. Ugyanakkor hiányérzetünk marad, mert nem világos, miért működik. Jöjjön most egy indukciós.

2. Bizonyítás. Feltesszük, hogy $K \leq \mathbb{C}$ és használjuk az algebra alaptételét. (Megjegyzés: valójában nincs *egyikre sem* szükség, mert minden test benne van egy alkalmas algebrailag zárt testben: ez a tétel is szerepel bizonyítással az Algebra, számelmélet és matematika alapjai blokk *Testek és algebrák* c. tárgyában.) Ha $n = 1$, akkor az állítás triviális, tehát legyen $n > 1$ és tegyük föl, hogy már beláttuk a Cayley–Hamilton tételt az $(n-1) \times (n-1)$-es mátrixokra.

Tehát $A \in \mathbb{C}^{n\times n}$ és $k_A[x] \in \mathbb{C}[x]$-nek van egy $\lambda_1 \in \mathbb{C}$ gyöke. Ezért A-nak van egy λ_1-hez tartozó $0 \neq b_1 \in \mathbb{C}^n$ sajátvektora: $Ab_1 = \lambda_1 b_1$. Kiegészítjük b_1-et $\mathbb{C}^n$ egy bázisává a $b_2, \dots, b_n \in \mathbb{C}^n$ vektorokkal. Legyen az A-hoz tartozó lineáris leképezés mátrixa az új $b_1, \dots, b_n$ bázisban B, azaz $B = S^{-1}AS$, ahol S az áttérési mátrix. Vegyük észre, hogy A-ra pontosan akkor teljesül a Cayley–Hamilton tétel állítása, ha B-re, hiszen $k_B(x) = \det(S^{-1}AS - xI) = \det(S^{-1}(A - xI)S) = \det S^{-1} \det(A - xI) \det S = k_A(x)$ és

$$k_B(B) = k_A(B) = k_A(S^{-1}AS) = a_0 I + a_1 S^{-1}AS + \cdots + a_n(S^{-1}AS)^n = S^{-1}k_A(A)S \ .$$

Másrészt B első oszlopának első eleme λ_1, a többi pedig 0, továbbá jelölje C az első sor és első oszlop elhagyásával B-ből kapott mátrixot. Ekkor $k_B(x) = (\lambda_1 - x)k_C(x)$ adódik a $B - xI$ mátrix determinánsát az első oszlop szerint kifejtve. Ebbe B-t behelyettesítve $k_B(B) = (\lambda_1 I - B)k_C(B)$ adódik (gondoljuk meg, hogy ezt tényleg szabad: B felcserélhető a skalármátrixokkal). Node nyilván $\lambda_1 I - B$ első oszlopának minden eleme 0. Másrészt a $k_C(B)$ mátrix első sora kivételével minden eleme 0, hiszen C-re igaz a Cayley–Hamilton tétel. (Itt azt kell észrevenni, hogy blokk-felsőháromszög mátrix blokkfőátlójában tagonként kell elvégezni a műveleteket.) A mátrixszorzás szabályai miatt $(\lambda_1 I - B)k_C(B) = 0$. $\square$

A fentinek egy variánsa a következő, ami nem használ indukciót.

3. Bizonyítás. Továbbra is legyen $K \leq \mathbb{C}$ és használjuk az algebra alaptételét (ahogy az előző bizonyításban, itt sincs erre igazán szükség). Írjuk az A minimálpolinomját $m_A(x) = (x - \lambda_1)^{\alpha_1} \cdots (x - \lambda_r)^{\alpha_r}$ gyöktényezős alakban (egyelőre nem tudjuk, hogy minden λ_i sajátérték!).

3. Lemma. *Legyen* $V_i := \operatorname{Ker}(A - \lambda_i I)^{\alpha_i}$. *Ekkor* $V = \bigoplus_{i=1}^r V_i$.

Bizonyítás. Legyen $f_i(x) := \frac{m_A(x)}{(x-\lambda_i)^{\alpha_i}}$ $(i = 1, \dots, r)$. Ekkor az $f_1, \dots, f_r$ polinomoknak nincs közös gyöke, azaz relatív prímek. Az euklideszi algoritmusból következik (r szerinti indukcióval), hogy a konstans 1 polinom (ami a legnagyobb közös osztó) előáll

$$1 = f_1(x)q_1(x) + \cdots + f_r(x)q_r(x)$$

alakban alkalmas $q_i(x) \in \mathbb{C}[x]$ polinomokkal. Ebbe az egyenletbe $x \mapsto A$-t helyettesítve azt kapjuk, hogy

$$I = f_1(A)q_1(A) + \cdots + f_r(A)q_r(A) \ ,$$

majd tetszőleges $v \in \mathbb{C}^n$ vektorra

$$v = f_1(A)q_1(A)v + \cdots f_r(A)q_r(A)v \ .$$

Vegyük észre, hogy $f_i(A)q_i(A)v \in V_i = \operatorname{Ker}(A - \lambda_i)^{\alpha_i}$, hiszen

$$(A - \lambda_i)^{\alpha_i} f_i(A)q_i(A)v = m_A(A)q_i(A)v = 0 \ .$$

Így $V = V_1 + \cdots + V_r$. Tegyük fel, hogy az összeg nem direkt, azaz

$$(V_1 + \cdots + V_{i-1}) \cap V_i \neq 0$$

valamely $2 \leq i \leq r$-re. Ekkor léteznek olyan $v_1 \in V_1, \dots v_{i-1} \in V_{i-1}$ vektorok, melyekre

$$0 \neq v_i := v_1 + \cdots v_{i-1} \in V_i \ .$$

Sőt, mivel $(x - \lambda_k)^{\alpha_k} \mid f_j(x)$, ezért $f_j(A)v_k = 0$ minden $j \neq k$-ra ($1 \leq k \leq i$ és $1 \leq j \leq r$). Speciálisan

$$v_i = f_1(A)q_1(A)v_i + \cdots f_r(A)q_r(A)v_i = f_i(A)q_i(A)v_i = f_i(A)q_i(A)(v_1 + \cdots + v_{i-1}) = 0 \ ,$$

ellentmondás (itt használtuk hallgatólagosan, hogy $f_j(A)$ és $q_j(A)$ felcserélhető, ezért minden $j \neq i$-re $f_j(A)q_j(A)v_i = q_j(A)f_j(A)v_i = q_j(A)0 = 0$). $\square$

A következő lemma triviális következménye a Cayley–Hamilton tételnek, de közvetlenül is igazolható és így használható a tétel bizonyítására.

4. Lemma. *Legyen φ egy nilpotens lineáris transzformáció egy n-dimenziós V vektortéren. Ekkor $\varphi^n = 0$.*

Bizonyítás. Indukció n szerint. Ha $n = 1$, akkor az állítás világos, hiszen ha $\varphi \neq 0$, akkor $\dim \operatorname{Ker}(\varphi) < 1$, azaz φ injektív.

Mivel φ nilpotens, nem lehet injektív, ezért $\operatorname{Ker}(\varphi) \neq 0$, így $\dim \operatorname{Im}(\varphi) = n - \dim \operatorname{Ker}(\varphi) < n$. Node $\operatorname{Im}(\varphi)$ egy φ-invariáns altér, melyre megszorítva φ-t ugyancsak egy nilpotens transzformációt kapunk. Az indukciós feltevést használva $\varphi_{|\operatorname{Im}(\varphi)}$-nek a $\dim \operatorname{Im}(\varphi)$-edik hatványa 0, speciálisan $\varphi^{n-1}_{|\operatorname{Im}(\varphi)} = 0$. Tehát $\varphi^n(v) = \varphi^{n-1}(\varphi(v)) = 0$, hiszen $\varphi(v) \in \operatorname{Im}(\varphi)$ és ezen φ^{n-1} azonosan 0. □

A 4. Lemmát alkalmazva az $A - \lambda_i I$ mátrixszal való szorzás V_i-re való megszorítására azt kapjuk, hogy $(A - \lambda_i I)^{\dim V_i}$ azonosan 0 a V_i altéren. Így a $(A - \lambda_1 I)^{\dim V_1} \cdots (A - \lambda_r I)^{\dim V_r}$ szorzat azonosan 0 az egész V-n, hiszen az a 3. Lemma szerint a V_i-k direkt összege. Speciálisan $m_A(x) \mid (x - \lambda_1)^{\dim V_1} \cdots (x - \lambda_r)^{\dim V_r}$, azaz $\alpha_i \leq \dim V_i$ minden $i = 1, \ldots, r$-re. Végezetül előjeltől eltekintve az $(x - \lambda_1)^{\dim V_1} \cdots (x - \lambda_r)^{\dim V_r}$ polinom az A karakterisztikus polinomja, hiszen V_i-n az A-nak egyetlen sajátértéke van (jelesül λ_i), mert az $A_{|V_i}$ megszorítás minimálpolinomja $(x - \lambda_i)^{\alpha_i}$ és minden sajátérték gyöke a minimálpolinomnak. □

5. Következmény. *Egy $A \in \mathbb{C}^{n \times n}$ mátrix pontosan akkor diagonalizálható, ha minimálpolinomja különböző gyöktényezők szorzatára bomlik (azaz minden sajátérték egyszeres gyöke).*

Bizonyítás. Tegyük fel, hogy $m_A(x) = (x - \lambda_1) \cdots (x - \lambda_r)$. Ekkor a 3. Lemma szerint $V = V_1 \oplus \cdots \oplus V_r$, ahol $V_i = \operatorname{Ker}(A - \lambda_i I)$ nem más, mint a λ_i-hez tartozó sajátaltér. $V_1, V_2, \ldots, V_r$ bázisait összeuniózva V-nek egy sajátvektorokból álló bázisát kapjuk, azaz A diagonalizálható.

Megfordítva tegyük fel, hogy A diagonalizálható. Mivel a minimálpolinom nem függ a bázis választásától, ezért azt is feltehetjük, hogy A diagonális. Node diagonális mátrix minimálpolinomja valóban a különböző sajátértékekhez tartozó gyöktényezők szorzata, hiszen ennek is gyöke a mátrix. □

Az utolsó bizonyítás azon alapszik, hogy a Cayley–Hamilton-tétel *diagonális* mátrixokra triviális, ráadásul szinte minden mátrix diagonalizálható.

4. Bizonyítás. Itt is feltesszük, hogy $K \leq \mathbb{C}$ (most sincs rá szükség). Ha $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ egy diagonális mátrix, akkor a főátlóbeli elemeket tagonként kell a $k_A(x)$ polinomba helyettesíteni, de mivel mindegyik sajátérték, ezért $k_A(A) = 0$. Ahogy a 2. bizonyításban is láttuk, ebből következik az állítás tetszőleges diagonalizálható mátrixra is. Most belátjuk, hogy a diagonalizálható mátrixok sűrűn vannak az összes mátrix között (azaz tetszőleges A mátrixra és $\varepsilon > 0$ számra van olyan B diagonalizálható mátrix, melyre $A - B$ minden elemének abszolútértéke $< \varepsilon$). Ebből következik a Cayley–Hamilton tétel tetszőleges A mátrixra, hiszen konvergálhatunk A-hoz diagonalizálható mátrixokkal és $k_M(M)$ folytonos függvénye M-nek.

Tegyük fel indirekten, hogy van egy $A \in \mathbb{C}^{n \times n}$ mátrix, melynek valamely $\varepsilon > 0$ valós számra az ε sugarú környezetében egyik mátrix sem diagonalizálható. Azt fogjuk használni, hogy ha $k_B(x)$-nek nincs többszörös gyöke, akkor B diagonalizálható, hiszen $k_B(x)$ minden gyökéhez tartozik egy-egy sajátvektor, melyek lineárisan függetlenek. Vegyünk egy $Y = ((y_{ij}))$ $n \times n$-es mátrixot, melynek minden eleme egy-egy formális változó, azaz $Y \in \mathbb{C}[y_{11}, \ldots, y_{nn}]^{n \times n}$ egy (n^2-változós) polinomokból álló mátrix. Ennek a karakterisztikus polinomja $k_Y(x) = \det(Y - xI) \in \mathbb{C}[x, y_{11}, \ldots, y_{nn}]$ egy $n^2 + 1$-változós polinom. Egy polinomnak pontosan akkor van többszörös gyöke (x-ben), ha van közös gyöke a deriváltjával. Vegyük tehát a $k_Y(x)$ polinom x-szerinti $\frac{\partial}{\partial x} k_Y(x)$ deriváltját (ez még mindig egy $n^2 + 1$-változós polinom). Két normált polinomnak pontosan akkor van közös gyöke, ha a rezultánsuk 0: egy polinom és a deriváltja esetében ez – konstans szorzótól eltekintve – a polinom *diszkriminánsa*. Képezzük tehát $k_Y(x)$-nek (az x szerinti) diszkriminánsát. Ez $k_Y(x)$ együtthatóinak polinomja, azaz

$\mathbb{C}[y_{11},\dots,y_{nn}]$-nek egy $D(y_{11},\dots,y_{nn})$ eleme. Vegyük észre, hogy D nem lehet az azonosan 0 polinom, hiszen *van olyan behelyettesítése*, ami nem 0, hiszen van olyan mátrix, aminek az összes sajátértéke különböző (pl. legyen $y_{ii}=i$ és $y_{ij}=0$ ha $i\neq j$). Az már elemi analízis/algebra, hogy ekkor $D\colon \mathbb{C}^{n\times n}\to\mathbb{C}$ nem tűnhet el egy A mátrix teljes ε sugarú környezetén: ellenkező esetben itt az összes parciális derivált is 0 lenne, alternatíve ez következik a polinomok azonossági tételéből is. $\square$

6. Megjegyzés. *A fenti bizonyításban valójában nincs szükség „analízisre" és a $\mathbb{C}$ topológiájára: $\mathbb{C}$ helyett tetszőleges K testre működik, csak az algebrai geometriában használatos ún.* Zariski-topológiát *kell tekinteni $K^{n\times n}$-en a $\mathbb{C}^{n\times n}$-en szokásos topológia helyett. A Zariski-topológiában a zárt halmazok véges sok polinom közös zérushelyeiként állnak elő. Azon mátrixok Zariski-nyílt halmazt alkotnak, melyeknek minden sajátértéke különböző, sőt, lényegében a polinomgyűrű nullosztómentességén múlik, hogy ez nemcsak nyílt, hanem sűrű is. Ez a módszer meglepően hatékony, rengetegszer használják a modern algebrai számelméletben, geometriában, többek között a 2-dimenziós p-adikus Langlands-megfeleltetés Colmez-féle – 2010-ben megjelent – bizonyítása is tartalmaz egy hasonló típusú érvelést. Ezekben a modernebb bizonyításokban is az a közös, hogy egy tételt először bizonyos speciális esetekben igazolunk, majd belátjuk, hogy a speciális esetek valamilyen értelemben sűrűn vannak az összes eset között, a tételünk igazságértéke pedig folytonosan változik az „esetek" halmazán. Ahhoz, hogy a „valamilyen értelemben sűrűn vannak" kifejezést precízzé tegyük, a Zariski-topológiának messzemenő (absztrakt) általánosításaira is szükség van. A lineáris transzformációkat tudjuk n^2 darab számmal tökéletesen jellemezni (melyek között semmilyen összefüggés nincs), de más, nehezebb tételekhez esetleg bizonyos típusú síkgörbék, vagy akár adott tulajdonságú csoportreprezentációk izomorfia osztályainak paraméterezésére lehet szükség. Az ilyen paramétertereket a geometriában* modulustérnek *(angolul moduli space) nevezik.*

A könnyebb érthetőség kedvéért a 4. bizonyítást részletezzük a 2×2-es mátrixok példáján. Először azt akarjuk látni, hogy $\mathbb{C}^{2\times 2}$-ben sűrűn vannak azon $\begin{pmatrix} a & b\\ c & d\end{pmatrix}$ mátrixok, melyeknek különböző a két sajátértéke, hiszen ezen mátrixokra teljesül a Cayley–Hamilton tétel. A karakterisztikus polinom $\det\begin{pmatrix} a-x & b\\ c & d-x\end{pmatrix} = x^2-(a+d)x+(ad-bc)$. A két sajátérték pontosan akkor különböző, ha a $D(a,b,c,d):=(a+d)^2-4(ad-bc)=a^2+d^2-2ad+4bc$ diszkriminánsnem 0. Ez a $D(a,b,c,d)$ négyváltozós polinom nem a 0 polinom (ránézésre, de a fenti érvelés ebben a speciális esetben azt használja, hogy a konkrét $\begin{pmatrix} 1 & 0\\ 0 & 2\end{pmatrix}$ diagonális mátrixra $D(1,0,0,2)=1\neq 0$). A következő lépés, hogy nincs olyan $\begin{pmatrix} a_0 & b_0\\ c_0 & d_0\end{pmatrix}\in\mathbb{C}^{2\times 2}$ mátrix és $\varepsilon>0$ valós, hogy $D(a,b,c,d)=0$ minden olyan $\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\mathbb{C}^{2\times 2}$ mátrixra, melyre $\left\|\begin{pmatrix} a & b\\ c & d\end{pmatrix}-\begin{pmatrix} a_0 & b_0\\ c_0 & d_0\end{pmatrix}\right\|_\infty := \max(|a-a_0|,|b-b_0|,|c-c_0|,|d-d_0|)<\varepsilon$. Rögzítsük ugyanis $b=b_0$, $c=c_0$, $d=d_0$-at, viszont a fusson az $a_0\in\mathbb{C}$ szám ε sugarú nyílt környezetén. Mivel $D(a,b_0,c_0,d_0)$ másodfokú a-ban, ezért csak két olyan a érték lehet, melyre $D(a,b_0,c_0,d_0)=0$, azaz nyilván nem minden olyan a, amire $|a-a_0|<\varepsilon$. Végezetül legyen $M_0=\begin{pmatrix} a_0 & b_0\\ c_0 & d_0\end{pmatrix}\in\mathbb{C}^{2\times 2}$ tetszőleges és minden $n\geq 1$ egészre válasszunk olyan $M_n\in\mathbb{C}^{2\times 2}$, melyre $\|M_n-M_0\|_\infty<\frac{1}{n}$ és M_n-nek két különböző sajátértéke van. Így M_n-re teljesül a Cayley–Hamilton tétel, azaz $k_{M_n}(M_n)=0$ $(n\geq 1)$. Viszont $\lim_{n\to\infty}M_n=M_0$ miatt $k_{M_0}(M_0)=\lim_{n\to\infty}k_{M_n}(M_n)=0$, hiszen a

$$\begin{array}{ccc} \mathbb{C}^{2\times 2} & \to & \mathbb{C}^{2\times 2}\\ M=\begin{pmatrix} a & b\\ c & d\end{pmatrix} & \mapsto & k_M(M)=\begin{pmatrix} a & b\\ c & d\end{pmatrix}^2-(a+d)\begin{pmatrix} a & b\\ c & d\end{pmatrix}+(ad-bc)I\end{array}$$

hozzárendelés egy folytonos függvény. Ez pedig pontosan azt jelenti, hogy M_0-ra is teljesül a Cayley–Hamilton tétel állítása.