Institutionen för matematik

SF1681
Linjär algebra, fk
HT20

SF1681 LINJÄR ALGEBRA, FORTSÄTTNINGSKURS FÖRELÄSNING 4

DAVID RYDH

4. Samtidig diagonalisering och Jordans normalform

Målet för idag.
- Minimalpolynomet och Cayley–Hamiltons sats
- Samtidig diagonalisering
- Exponentialavbildningen och andra analytiska funktioner av matriser

Spår, determinant och egenvärden.

Definition 4.1. Spåret (en: *trace*) av en $n \times n$-matris A är summan av diagonalelementen.

$$\operatorname{tr} A = \sum_{i=1} a_{ii}$$

Man kan visa att $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ och därmed att $\operatorname{tr}(PAP^{-1}) = \operatorname{tr}(AP^{-1}P) = \operatorname{tr}(A)$. Alltså beror inte spåret på val av bas och vi kan definiera spåret av en operator på ett ändligtdimensionellt vektorrum. Alternativt kan man visa att det karakteristiska polynomet är

$$p_A(x) = x^n - \operatorname{tr}(A)x^{n-1} + a_{n-2}x^{n-2} \cdots + a_1 x + (-1)^n \det(A)$$

och eftersom det karakteristiska polynomet är invariant under konjugering så är även spåret det.

Om $k = \mathbb{C}$ så är $p_A(x) = (x-\lambda_1)(x-\lambda_2)\dots(x-\lambda_n)$. Detta ger att

- Spåret är summan av egenvärdena $\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.
- Determinanten är produkten av egenvärdena $\det(A) = \lambda_1\lambda_2\cdots\lambda_n$.

Detta kan också ses direkt genom att först överföra matrisen till övertriangulär form eftersom egenvärdena då är diagonalelementen.

Exempel 4.2. För en 2×2-matris $A = (a_{ij})$ med egenvärden λ_1, λ_2 har vi att det karakteristiska polynomet är:

$$\begin{aligned}
p_A(x) &= x^2 - \operatorname{tr}(A)x + \det(A) \\
&= x^2 - (a_{11} + a_{22})x + (a_{11}a_{22} - a_{21}a_{12}) \\
&= x^2 - (\lambda_1 + \lambda_2)x + \lambda_1\lambda_2
\end{aligned}$$

Datum: 2021-01-03.

Minimalpolynomet och Cayley–Hamiltons sats. Om $p(x) = a_d x^n + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$ är ett polynom så kan vi evaluera det i en $n\times n$-matris A på följande vis:

$$p(A) = a_d A^d + a_{d-1}A^{d-1} + \cdots + a_1 A + a_0 I$$

och får en $n\times n$-matris. Observera här att det konstanta monomet $a_0 = a_0 x^0$ blir $a_0 A^0 = a_0 I$ där I är identitetsmatrisen av storlek $n\times n$.

Definition 4.3 (Minimalpolynomet)**.** Om A är en $n\times n$-matris är *minimalpolynomet* $q_A(x)$ det *moniska*[1] polynom av lägsta grad så att $q_A(A) = 0$.

Anmärkning 4.4. Graden av minimalpolynomet är det lägsta heltal d så att $\{I, A, A^2, \ldots, A^d\}$ är linjärt beroende eftersom $(x^d + a_d x^{d-1} + \cdots + a_0)(A) = 0 \iff (A^d + \cdots + a_0 I) = 0 \iff A^d$ är en linjärkombination av $\{I, A, A^2, A^3, \ldots, A^{d-1}\}$. Vi noterar att $\deg q_A(x) \le n^2$ eftersom $\{A^i\}_{i=0}^{n^2} = \{I, A, A^2, \ldots, A^{n^2}\}$ är n^2+1 matriser i ett vektorrum av dimension n^2 och därmed linjärt beroende. Nästa sats ger att $\deg q_A(x) \le n$.

Sats 4.5 (Cayley–Hamilton, LADR 8.37)**.** $p_A(A) = 0$.

Exempel 4.6. För 2×2-matriser har vi att $p_A(x) = x^2 - \operatorname{tr}(A)x + \det(A)$ och Cayley–Hamiltons sats säger

$$\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}^2 - (x_{11}+x_{22}) \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} + (x_{11}x_{22} - x_{12}x_{21}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

vilket går att verifiera direkt.

Korollarium 4.7 (LADR, 8.46)**.** *$q_A(x)$ är en delare i $p_A(x)$.*

Bevis. Vi utför polynomdivision av $p_A(x)$ med $q_A(x)$. Detta ger

$$p_A(x) = s(x)q_A(x) + r(x),$$

där antingen $r(x) = 0$ eller $\deg r(x) < \deg q_A(x)$. Sätter vi $x = A$ så är $r(A) = p_A(A) - s(A)q_A(A) = 0$, vilket ger $r(x) = 0$ eftersom $q_A(x)$ per definition är det polynom av lägst grad för vilket $q_A(A) = 0$. □

Exempel 4.8. Låt $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ med $k = \mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Då har vi

$$A^2 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{och} \quad A^3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I = A^2 + A.$$

Karakteristiska polynomet är $p_A(x) = x^2 + x + 1$ och eftersom $\{I, A\}$ är linjärt oberoende är detta också minimalpolynomet.

Exempel 4.9. Om vi har koefficienter i $k = \mathbb{F}_2$ får vi

$$B = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \Longrightarrow \quad B^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Det karakteristiska polynomet är $p_B(x) = x^4$ medan minimalpolynomet är $q_B(x) = x^2$.

Hur blir det med koefficienter i $\mathbb{Q}$ för samma matris?

$$B = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B^2 = \begin{bmatrix} 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B^3 = \begin{bmatrix} 0 & 4 & 4 & 0 \\ 0 & 4 & 4 & 0 \\ 0 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Vi får $p_B(x) = x^4 - 2x^3$ och $q_B(x) = x^3 - 2x^2$.

[1] med ledande koefficient ett

Anmärkning 4.10. Vi kan ordna en matris med ett givet karakteristiskt polynom $p_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ genom

$$A = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}$$

Här är minimalpolynomet lika med det karakteristiska polynomet ty de första $n-k$ kolumnerna av A^k är $\mathbf{e}_{k+1}, \mathbf{e}_{k+2}, \ldots, \mathbf{e}_n$ och man ser därför lätt att $I, A, A^2, \ldots, A^{n-1}$ är linjärt oberoende i $M_{n,n}$.

Sats 4.11 (LADR, 8.49)**.** *Nollställena till minimalpolynomet $q_A(x)$ är precis egenvärdena till A.*

Bevis. Eftersom $q_A(x)$ är en delare till det karakteristiska polynomet $p_A(x)$ så måste varje nollställe till $q_A(x)$ vara ett nollställe till $p_A(x)$, dvs ett egenvärde.

Omvänt, om λ är ett egenvärde så finns en egenvektor $\boldsymbol{\xi}$ så att $A\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$. Om $q_A(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ så är

$$\begin{aligned} 0 = q_A(A)\boldsymbol{\xi} &= A^n\boldsymbol{\xi} + a_{n-1}A^{n-1}\boldsymbol{\xi} + \cdots + a_1A\boldsymbol{\xi} + a_0\boldsymbol{\xi} \\ &= \lambda^n\boldsymbol{\xi} + a_{n-1}\lambda^{n-1}\boldsymbol{\xi} + \cdots + a_1\lambda\boldsymbol{\xi} + a_0\boldsymbol{\xi} \\ &= q_A(\lambda)\boldsymbol{\xi}. \end{aligned}$$

Eftersom $\boldsymbol{\xi} \neq \mathbf{0}$ så är alltså $q_A(\lambda) = 0$. □

Bevis av Cayley–Hamiltons sats. Låt $k = \mathbb{C}$. Till att börja med kan vi se att $P_A(A) = 0$ gäller för diagonalmatriser.

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad \Rightarrow \quad p_D(D) = \begin{bmatrix} p_D(\lambda_1) & 0 & \cdots & 0 \\ 0 & p_D(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_D(\lambda_n) \end{bmatrix} = 0$$

eftersom $p_D(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$. Vi har att alla matriser i $M_n(\mathbb{C})$ kan approximeras godtyckligt nära av diagonaliserbara matriser eftersom alla matriser med distinkta nollställen till det karakteristiska polynomet är diagonaliserbara.

Eftersom $A \mapsto P_A(A)$ är en kontinuerlig funktion som är noll för alla diagonaliserbara matriser måste den vara noll för alla matriser.

Om vi skriver upp identiteten $P_A(A) = 0$ i variablerna $x_{11}, \ldots, x_{nn}$ som i Exempel 4.6 är det en polynomidentitet med heltalskoefficienter. Därmed kommer den att gälla för alla kroppar k. □

Blockmatriser. En *blockmatris* är en $(m+n) \times (a+b)$ matris på formen

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

där B och C har m rader, D och E har n rader, B och D har a kolumner och C och E har b kolumner.

Om $\mathbf{x} = \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix}$ är en kolonnvektor med $a+b$ kolumner så blir

$$A\mathbf{x} = \begin{bmatrix} B & C \\ D & E \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} B\mathbf{v} + C\mathbf{w} \\ D\mathbf{v} + E\mathbf{w} \end{bmatrix}$$

Om vi betraktar $\mathbb{R}^{a+b} = \mathbb{R}^a \oplus \mathbb{R}^b = V \oplus W$ och $\mathbb{R}^{m+n} = \mathbb{R}^m \oplus \mathbb{R}^n$ som inre direkta summor och låter $L \colon \mathbb{R}^{a+b} \longrightarrow \mathbb{R}^{m+n}$ vara den linjära avbildning som ges av A så har vi

$$L((\mathbf{v}, \mathbf{w})) = (B\mathbf{v} + C\mathbf{w}, D\mathbf{v} + E\mathbf{w})$$

där $\mathbf{v} \in \mathbb{R}^a$ och $\mathbf{w} \in \mathbb{R}^b$.

Vi säger att A är *block-övertriangulär* om $D = 0$. Då är $L((\mathbf{v}, \mathbf{0})) = (B\mathbf{v}, \mathbf{0})$ och L tar alltså $\mathbb{R}^a$ på $\mathbb{R}^m$.

Det viktigaste fallet är då $a = m$ och $b = n$. Då är A, B, E kvadratiska och L är en operator. Enkla beräkningar ger:

(1) $\det(A) = \det(B)\det(E)$.
(2) $\operatorname{tr}(A) = \operatorname{tr}(B) + \operatorname{tr}(E)$.
(3) $p_A(x) = p_A(x)p_E(x)$.
(4) Egenvärdena till A är egenvärdena av B och E tillsammans.
(5) Egenvektorer till B är egenvektorer till A. Detta gäller dock ej för egenvektorer till E.

Exempel 4.12. Betrakta

$$A = \begin{bmatrix} 1 & -2 & 7 \\ 1 & 4 & 5 \\ 0 & 0 & 8 \end{bmatrix}$$

som en övertriangulär block-matris där övre vänstra blocket har storlek 2×2, dvs

$$B = \begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$$

i notationen ovan. Matrisen B har determinant 6 och spår 5 och alltså egenvärden $\{2, 3\}$. Alltså har A egenvärden $\{2, 3, 8\}$. Egenvektorerna till B är

- En egenvektor till $\lambda = 2$ är $(2, -1)$.
- En egenvektor till $\lambda = 3$ är $(1, -1)$.

Detta ger egenvektorerna $(2, -1, 0)$ och $(1, -1, 0)$ till A. Men $(0, 0, 1)$ är inte en egenvektor till egenvärdet $\lambda = 8$ utan lite beräkningar ger egenvektorn $(3, 7, 5)$.

Exempel 4.13. Matrisen

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

är en övertriangulär block-matris. Även om B och E är diagonaliserbara (de är diagonala!) så är inte A diagonaliserbar.

Vi säger att A är *block-diagonal* om $C = D = 0$. Då kan vi dela upp L som en operator $L|_V \colon V \longrightarrow V$ som ges av matrisen B och en operator $L|_W \colon W \longrightarrow W$ som ges av matrisen E. Egenvektorerna till L är då precis egenvektorerna till $L|_V$ och $L|_W$. Och A är diagonaliserbar precis när B och E är diagonaliserbara.

Operatorer och block-övertriangulära matriser. Om A är block-övertriangulär så är V *invariant* under L, dvs $L(\mathbf{v}) \in V$ för alla $\mathbf{v} \in V$. Vi får en operator $L|_V \colon V \longrightarrow V$ som är L begränsad till V och som ges av matrisen A. T ex är en egenvektor till $L|_V$ en egenvektor till L. Däremot kan vi inte begränsa L till en operator på W eftersom $L(\mathbf{w})$ inte ligger i W för alla $\mathbf{w} \in W$ såvida inte $C = 0$.

Istället har vi en inducerad operator $U/V \longrightarrow U/V$ på *kvotrummet* där $U = \mathbb{R}^{a+b} = V \oplus W$ på följande vis. Lägg först märke till att $W \to U \to U/V$ är en isomorfi eftersom W är ett komplement till V. Vi kan alltså representera varje vektor i U/V unikt som $\mathbf{w} + V$ där $\mathbf{w} \in W$. Applicerar vi L på en vektor $\mathbf{w} + \mathbf{v}$ får vi:

$$L(\mathbf{w} + \mathbf{v}) = E\mathbf{w} + (B\mathbf{v} + C\mathbf{w}) = E\mathbf{w} + V$$

Så L inducerar en avbildning $U/V \longrightarrow U/V$ som ges av matrisen E.

Samtidig diagonalisering och kommuterande operatorer.

Definition 4.14. Två operatorer L_1 och L_2 är *samtidigt diagonaliserbara* om det finns en gemensam bas av egenvektorer, dvs en bas så att *båda* operatorerna representeras av diagonalmatriser.

Anmärkning 4.15. Om L_1 och L_2 är samtidigt diagonaliserbara måste $L_1 \circ L_2 = L_2 \circ L_1$, eftersom detta gäller för diagonalmatriser.

Definition 4.16 (Kommuterande operatorer)**.** Två operatorer L_1 och L_2 *kommuterar* om $L_1 \circ L_2 = L_2 \circ L_1$.

Sats 4.17 (Sadun, Thm. 4.10)**.** *Om* $\dim V < \infty$ *och* L_1 *och* L_2 *är diagonaliserbara är följande ekvivalent*

(a) L_1 *och* L_2 *är samtidigt diagonaliserbara*
(b) L_1 *och* L_2 *kommuterar.*

Att (a) $\Longrightarrow$ (b) är anmärkningen ovan.

Bevisidé för (b) $\Longrightarrow$ *(a).* Om L_1 och L_2 kommuterar är egenrummen för operatorn L_1 är invarianta under L_2 och tvärtom: dvs om $\boldsymbol{\xi}$ är en egenvektor till L_1 med egenvärde λ, så är även $L_2(\boldsymbol{\xi})$ en egenvektor till L_1 med samma egenvärde. Detta följer av:

$$L_1(L_2(\boldsymbol{\xi})) = L_2 \circ L_1(\boldsymbol{\xi}) = L_2(\lambda\boldsymbol{\xi}) = \lambda L_2(\boldsymbol{\xi})$$

Om E_λ är egenrummet till L_1 med egenvärdet λ är alltså $L_2(E_\lambda) \subseteq E_\lambda$. Om vi därför väljer en egenbas för L_1 så blir matrisen för L_2 blockdiagonal i denna bas. Eftersom L_2 är diagonaliserbar måste varje block vara diagonaliserbart. $\square$

Exempel 4.18. Låt $L\colon V \longrightarrow V$ vara en operator och $p(x)$ och $q(x)$ vara polynom. Då kommuterar $p(L)$ med $q(L)$. Detta gäller även om L inte är diagonaliserbar.

Exempel 4.19. Om $L_1, L_2\colon V \longrightarrow V$ är kommuterande operatorer så kan vi definiera $p(L_1, L_2)$ för varje polynom $p(x,y) \in k[x,y]$.

Exponentialfunktionen och andra funktioner av matriser. Om $f(z)$ är en *analytisk* funktion[2] kring $z = 0$, t ex $f(z) = \exp(z)$, kan vi definiera $f(A)$ för en kvadratisk matris A genom att sätta

$$f(A) = \sum_{i=0}^{\infty} a_i A^i$$

förutsatt att serien konvergerar.

Definition 4.20 (Norm)**.** Om $\|\cdot\|$ är den euklidiska normen på $\mathbb{R}^n$ eller $\mathbb{C}^n$ kan vi definiera normen av en $n \times n$-matris som

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

Anmärkning 4.21. Om $\|A\| < \infty$ får vi $\|A^i\| \leq \|A\|^i$ och konvergensen av $\sum_{i\geq 0} a_i A^i$ fås från konvergensen av $f(z) = \sum_{i\geq 0} a_i z^i$. Dvs, om serien för $f(z)$ konvergerar då $|z| < R$ så konvergerar serien för $f(A)$ då $\|A\| < R$.

Analytiska funktioner på diagonaliserbar matriser.

Sats 4.22. *Om* A *är diagonaliserbar med* $A = PDP^{-1}$ *och* $f(z) = \sum_{i\geq 0} a_i z^i$ *en analytisk funktion kring* $z = 0$ *kan vi beräkna*

$$f(A) = Pf(D)P^{-1} = P \begin{bmatrix} f(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & 0 & \cdots & 0 \\ 0 & 0 & f(\lambda_3) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} P^{-1}$$

där serien är konvergent.

Bevis. För all $i \geq 0$ har vi:

$$A^i = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = PD^iP^{-1}$$

vilket ger

$$f(A) = Pf(D)P^{-1}.$$

[2] Detta betyder att Taylorserien i varje punkt konvergerar i en omgivning. Om f är en analytisk funktion definierad på hela $\mathbb{C}$ så konvergerar Taylor-serien överallt.

För en diagonalmatris $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ så är $D^i = \mathrm{diag}(\lambda_1^i, \dots, \lambda_n^i)$ vilket ger

$$f(D) = \mathrm{diag}\big(f(\lambda_1), \dots, f(\lambda_n)\big). \qquad \square$$