

NLA: Arnoldi methode voor eigenwaarde berekeningen

Nithin Govindarajan

KU Leuven

August 24, 2023

Wat wordt er behandeld?

- ▶ Lecture 33 and 34 van Trefethen & Bau
- ▶ Het wordt verwacht dat je Lecture 32 van Trefethen & Bau in je vrije tijd zelf een keer doorleest
- ▶ Een selectie van oefeningen.

Probeer het eerst zelf te oplossen!

Unitaire reductie naar een Hessenberg matrix

We hebben gezien dat, met behulp van Householder reflecties, elke matrix $A \in \mathbb{C}^{n \times n}$ gereduceerd kan worden naar een Hessenberg matrix:

$$Q^* A Q = \begin{bmatrix} * & * & * & \cdots & * & * \\ * & * & * & \cdots & * & * \\ & * & * & \cdots & * & * \\ & & * & \cdots & * & * \\ & & & \ddots & \vdots & \vdots \\ & & & & * & * \end{bmatrix} =: H$$

Arnoldi procedure

- ▶ Herschrijf naar $AQ = QH$.
- ▶ Kolom-voor-kolom hebben we dan de vergelijking:

$$Aq_k = h_{1k}q_1 + h_{2k}q_2 + \dots + h_{(k+1)k}q_{k+1}$$

- ▶ Met voorkennis van q_j voor $j = 1, 2, \dots, k$, is het héél makkelijk om de rest te achterhalen:

$$\begin{aligned}h_{jk} &= q_j^* Aq_k, \\h_{(k+1)k}q_{k+1} &= Aq_k - \sum_{j=1}^k h_{jk}q_j.\end{aligned}$$

- ▶ Dit suggereert een alternatieve manier om de Hessenberg reductie uit te voeren...

Arnoldi procedure

Herken de overeenkomsten en verschillen met (modified) Gram-Schmidt!

- 1: Kies een willekeurige $b \in \mathbb{C}^n$ en stel $q_1 = b / \|b\|_2$.
- 2: **for** $k = 1, 2, \dots, n$ **do**
- 3: $w := Aq_k$
- 4: **for** $j = 1, 2, \dots, k$ **do**
- 5: $h_{jk} := q_j^* w$
- 6: $w = w - h_{jk} q_j$
- 7: **end for**
- 8: $h_{(k+1)k} = \|w\|_2$
- 9: $q_{k+1} = w / h_{(k+1)k}$
- 10: **end for**

Huiswerk: Het algoritme loopt vast als $h_{(k+1)k} = 0$. Hoe moet het uitgebreid worden?

Arnoldi procedure

We kunnen ook ergens tussenin (bijv. $p < n$) stoppen...

- 1: Kies een willekeurige $b \in \mathbb{C}^n$ en stel $q_1 = b / \|b\|_2$.
- 2: **for** $k = 1, 2, \dots, p$ **do**
- 3: $w := Aq_k$
- 4: **for** $j = 1, 2, \dots, k$ **do**
- 5: $h_{jk} := q_j^* w$
- 6: $w = w - h_{jk} q_j$
- 7: **end for**
- 8: $h_{(k+1)k} = \|w\|_2$
- 9: $q_{k+1} = w / h_{(k+1)k}$
- 10: **end for**

Huiswerk: Het algoritme loopt vast als $h_{(k+1)k} = 0$. Hoe moet het uitgebreid worden?

Partiële reductie naar een Hessenberg matrix

Als we stoppen bij $p < n$,

$$A \begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & 0 & h_{(p+1)p} q_{p+1} \end{bmatrix} =$$
$$\begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1(p-1)} & h_{1p} \\ h_{21} & h_{22} & \cdots & h_{2(p-1)} & h_{2p} \\ & h_{32} & \cdots & h_{3(p-1)} & h_{3p} \\ & & \ddots & \vdots & \vdots \\ & & & h_{p(p-1)} & h_{pp} \end{bmatrix}$$

In korte notatie:

$$AQ_p - (h_{(p+1)p} q_{p+1}) e_p^T = Q_p H_p$$

Partiële reductie naar een Hessenberg matrix

Als we stoppen bij $p < n$,

$$A \begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & 0 & h_{(p+1)p} q_{p+1} \end{bmatrix} =$$
$$\begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1(p-1)} & h_{1p} \\ h_{21} & h_{22} & \cdots & h_{2(p-1)} & h_{2p} \\ & h_{32} & \cdots & h_{3(p-1)} & h_{3p} \\ & & \ddots & \vdots & \vdots \\ & & & h_{p(p-1)} & h_{pp} \end{bmatrix}$$

Vermenigvuldig links met Q_p^* :

$$Q_p^* A Q_p - Q_p^* (h_{(p+1)p} q_{p+1}) e_p^T = Q_p^* Q_p H_p$$

Partiële reductie naar een Hessenberg matrix

Als we stoppen bij $p < n$,

$$A \begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & 0 & h_{(p+1)p}q_{p+1} \end{bmatrix} =$$
$$\begin{bmatrix} q_1 & q_2 & \cdots & q_{p-1} & q_p \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1(p-1)} & h_{1p} \\ h_{21} & h_{22} & \cdots & h_{2(p-1)} & h_{2p} \\ & h_{32} & \cdots & h_{3(p-1)} & h_{3p} \\ & & \ddots & \vdots & \vdots \\ & & & h_{p(p-1)} & h_{pp} \end{bmatrix}$$

Ofwel:

$$Q_p^* A Q_p = H_p$$

De “Ritz” eigenwaarden en eigenvectoren

Het eigenwaarde-eigenvector paar (μ_i, x_i) van de Hessenberg matrix:

$$H_p x_i = \mu_i x_i$$

noemt men de Ritz eigenwaarden en eigenvectoren.

voor $p \ll n$, is het relatief weinig werk om aan (μ_i, x_i) te komen, dus hoe zijn de Ritz eigenwaarden en eigenvectoren gerelateerd aan A ?

Een eerste analyse

- Sowieso voor $p = n$ zijn de Ritz eigenwaarden gelijk aan de eigenwaarden van A . We hebben dan ook:

$$A(Q_n x) = \mu(Q_n x).$$

- In het algemeen, hebben we:

$$\begin{aligned} A Q_p x - (h_{(p+1)p} q_{p+1}) e_p^T x &= Q_p H_p x \\ \left(A - \underbrace{(h_{(p+1)p} q_{p+1}) e_p^T Q_p^*}_{:=E} \right) (Q_p x) &= \mu(Q_p x) \end{aligned}$$

- *Bauer-Fike* geeft aan dat μ een goede benadering is van de eigenwaarde van A als $\|E\|$ klein is en A normaal.

Een eerste analyse

We kunnen een diepere analyse doen, maar daar hebben we eerst een paar tools voor nodig.

Cayley-Hamilton

Theorem

Definieer $c(x) := \det(\lambda I - A)$. De matrix $A \in \mathbb{C}^{n \times n}$ annuleert zijn eigen karakteristieke veelterm:

$$c(A) = \alpha_0 I + \alpha_1 A + \dots + A^n = 0.$$

Cayley-Hamilton

Proof.

Als A diagonaliseerbaar is, hebben we een makkelelijk bewijs:

$$\begin{aligned}c(A) &= \alpha_0 V \Lambda V^{-1} + \alpha_1 V \Lambda V^{-1} + \dots + V \Lambda^n V^{-1} \\&= V (\alpha_0 I + \alpha_1 \Lambda + \dots + \alpha_n \Lambda^n) V^{-1} \\&= V \begin{bmatrix} c(\lambda_1) & & \\ & \ddots & \\ & & c(\lambda_n) \end{bmatrix} V^{-1}\end{aligned}$$

Maar zelfs zonder enige gebruik van eigendecomposities, is deze stelling te bewijzen (we besparen de details). □

Oefening 1

Stel dat $b, Ab, \dots, A^{p-1}b$ een lineaire onafhankelijke collectie van vectoren vormt. Toon aan dat de vectoren $q_1, q_2, \dots, q_p \in \mathbb{C}^n$ van de arnoldi procedure een basis vormen voor de zogehete *Krylov* ruimte:

$$\mathcal{K}(A, b, p) := \text{span} \{ b, Ab, \dots, A^{p-1}b \} .$$

oefening 1 - antwoord

Met behulp van inductie, tonen we aan dat

$$A^r b = \alpha_1 q_1 + \alpha_2 q_2 + \dots + \alpha_{r+1} q_{r+1}, \quad \alpha_{r+1} \neq 0$$

voor $r = 0, 1, \dots, p-1$. Dit heeft als gevolg dat¹:

$$\text{span}\{b, Ab, \dots, A^r b\} = \text{span}\{q_1, q_2, \dots, q_{r+1}\}$$

¹Ga voor jezelf na!

oefening 1 - antwoord

De *basis-stap* is triviaal, omdat $q_1 = b / \|b\|_2$ volgens de Arnoldi procedure.

oefening 1 - antwoord

Voor de *inductie-stap*, nemen we aan dat (inductie-hypothese):

$$A^{r-1}b = \alpha_1 q_1 + \alpha_2 q_2 + \dots + \alpha_r q_r, \quad \alpha_r \neq 0.$$

en dat $\text{span} \{b, Ab, \dots, A^{r-1}b\} = \text{span} \{q_1, q_2, \dots, q_r\}$.

Arnoldi produceert $Aq_k = h_{1k}q_1 + h_{2k}q_2 + \dots + h_{(k+1)k}q_{k+1}$. Dus:

$$\begin{aligned} A^r b &= A(A^{r-1}b) = \alpha_1 Aq_1 + \alpha_2 Aq_2 + \dots + \alpha_r Aq_r \\ &= \alpha'_1 q_1 + \alpha'_2 q_2 + \dots + \alpha'_r q_r + \alpha'_{r+1} q_{r+1} \end{aligned}$$

Omdat $\{b, Ab, \dots, A^r b\}$ lineair onafhankelijk zijn, moet $\alpha'_{r+1} \neq 0$. Dit leidt ook naar:

$$\text{span} \{b, Ab, \dots, A^r b\} = \text{span} \{q_1, q_2, \dots, q_{r+1}\}.$$

Een inzichtrijke stelling

Theorem

De karakteristieke veelterm $c_p(\mu) := \det(\mu I - H_p)$ is de oplossing van het minimalisatie probleem:

$$\min_{\deg(v) \leq p} \|v(A)b\|_2 = \|c_p(A)b\|_2$$

Proof.

Deze stelling volgt uit Cayley-Hamilton en het aangetoonde resultaat in oefening 1. De details staan in het boek van Trefethen & Bau. □

Interpretatie

- ▶ Een generieke $b \in \mathbb{C}^n$ heeft componenten in alle eigenvectoren:

$$b = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_n u_n, \quad \alpha_i \neq 0.$$

- ▶ De norm $\|c_p(A)b\|_2$ wordt min-of-meer bepaald door de waarden van $c_p(\mu)$ geëvalueerd op de eigenwaarden van A . Dit volgt uit:

$$\|c_p(A)b\|_2 = \|U c_p(\Lambda) U^{-1} b\|_2 = \|U c_p(\Lambda) \alpha\|_2.$$

- ▶ Als $\|c_p(A)b\|_2$ klein is, dan is $c_p(\mu)$ geëvalueerd op de eigenwaarden van A ook klein.
- ▶ De nulpunten (i.e. Ritz waarden) van $c_p(\mu)$ en de eigenwaarden van A liggen dan dicht op elkaar.
- ▶ In het extreme geval zijn ze exact hetzelfde.

De dominante eigenwaarden worden eerst gevonden!

- ▶ Arnoldi heeft de neiging om eerst te convergeren naar de dominante eigenwaarden.
- ▶ Stel dat $A \in \mathbb{C}^{n \times n}$ normaal is en:

$$\lambda_1 = 1000, \quad |\lambda_i| \approx 1 \quad \text{voor } i = 2, \dots, n.$$

- ▶ Omdat A normaal is, hebben we de bovengrens:

$$\|c_p(A)b\|_2 = \sqrt{\sum_{i=1}^n |\alpha_i c_p(\lambda_i)|^2}$$

- ▶ $c_p(\cdot)$ heeft maar p nulpunten (= Ritz waarden μ_i).
- ▶ Als $|c_p(\lambda_i)|$ heel klein is dan zit λ_i dicht in de buurt van een bepaalde ritz waarde μ_j .

Voorbeeld

Stel dat:

$$A = \begin{bmatrix} 1000 & & & \\ & 1 & & \\ & & e^{\pi/3} & \\ & & & e^{2\pi/3} \end{bmatrix}, \quad b = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

De eigenwaarde $\lambda = 1000$ schiet uit van de rest...

Welke quadratische functie minimalizeert $\|p(A)b\|_2$ het beste?

- ▶ Slechte keuze: $p(x) = (x - 1)(x - e^{\pi/3})$
- ▶ Betere keuze: $p(x) = (x - 1)(x - 1000)$

De optimale veelterm heeft de tendentie om de uit-schietende eigenwaarde als allereerst te benaderen.

Numeriek Voorbeeld uit Trefethen's boek

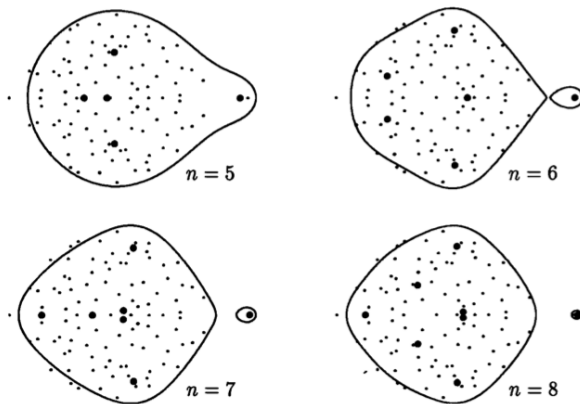


Figure 34.3. *Arnoldi lemniscates (34.4)–(34.5) at steps $n = 5, 6, 7, 8$ for the same matrix A . The small dots are the eigenvalues of A , and the large dots are the eigenvalues of H_n , i.e., the Ritz values. One component of the Arnoldi lemniscate first “swallows” the outlier eigenvalue, and in subsequent iterations it then shrinks to a point at a geometric rate.*