

NLA: Het QR algoritme - een introductie

Nithin Govindarajan

KU Leuven

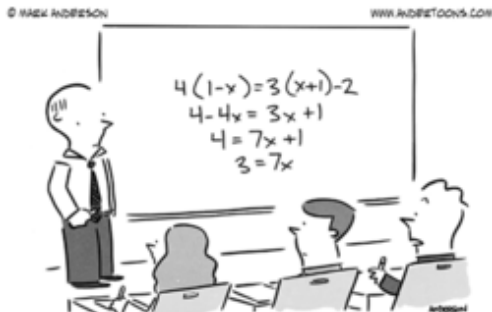
August 24, 2023

Wat wordt er behandeld?

- ▶ Elementen van Lecture 28 in Trefethen&Boyd.
- ▶ Lecture 26 in Trefethen&Boyd.
- ▶ Extra materiaal: afstanden tussen deelruimtes
- ▶ Een selectie van oefeningen.

Probeer het eerst zelf te oplossen!

Notatie: afstanden tussen twee deelruimtes



"Wouldn't it be more efficient to just find who's complicating equations and ask them to stop?"

Notatie: afstanden tussen twee deelruimtes

Beschouw twee deelruimtes:

$$\mathcal{A} = \text{span}\{a_1, \dots, a_r\} \subset \mathbb{R}^n, \quad \mathcal{B} = \text{span}\{b_1, \dots, b_r\} \subset \mathbb{R}^n.$$

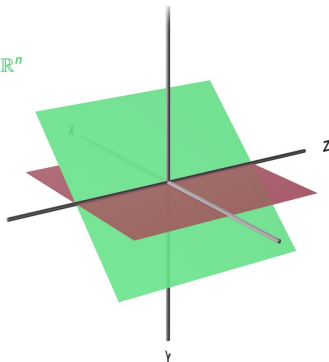
Om ons leven *makkelijker* te maken, introduceren we de volgende definitie om de afstand tussen twee deelruimtes te beschrijven:

$$d(\mathcal{A}, \mathcal{B}) = \sup_{v \in \mathcal{A}, \|v\|_2=1} \left(\inf_{w \in \mathcal{B}, \|w\|_2=1} \|v - w\|_2 \right)$$

Notatie: afstanden tussen twee deelruimtes

DISTANCE BETWEEN TWO SUBSPACES

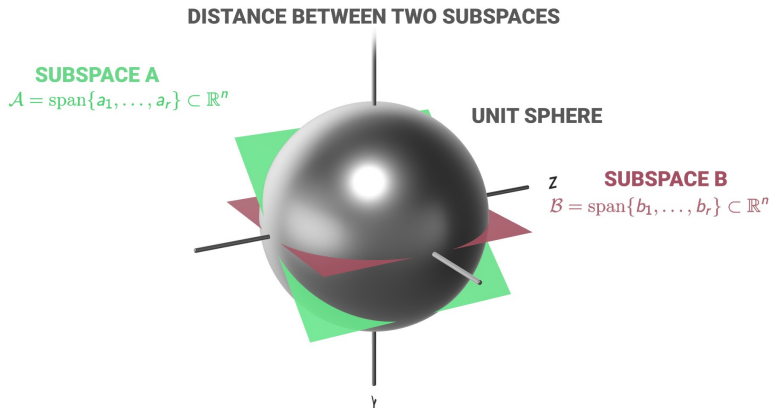
SUBSPACE A
 $\mathcal{A} = \text{span}\{a_1, \dots, a_r\} \subset \mathbb{R}^n$



SUBSPACE B
 $\mathcal{B} = \text{span}\{b_1, \dots, b_r\} \subset \mathbb{R}^n$

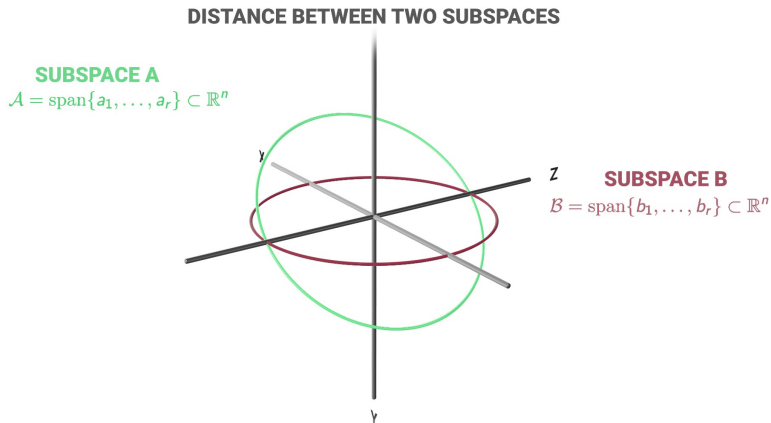
Visualization by Rob Zink

Notatie: afstanden tussen twee deelruimtes



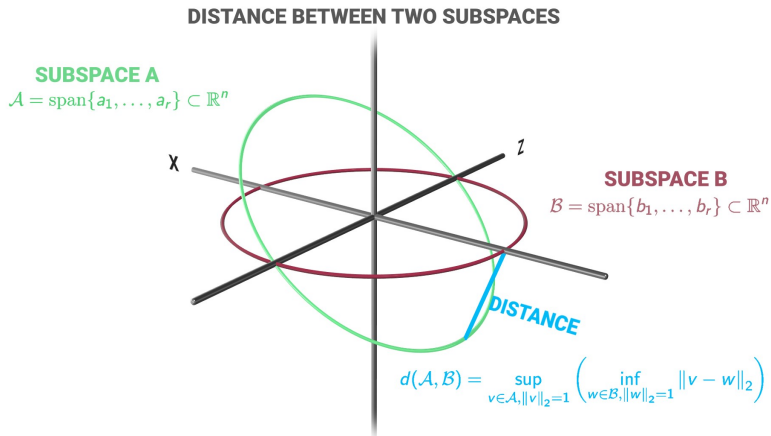
Visualization by Rob Zink

Notatie: afstanden tussen twee deelruimtes



Visualization by Rob Zink

Notatie: afstanden tussen twee deelruimtes



Visualization by Rob Zink

Stellingen van vorige les in de nieuwe notatie

Theorem (Convergentie methoden van machten)

Neem aan dat A reëel en symmetrisch is. De iteraties $q^{(k)} \in \mathbb{R}^n$ convergeren naar de eigenvector u_1 met snelheid:

$$\left\| q^{(k)} - (\pm u_1) \right\|_2 = O \left(\left| \frac{\lambda_2}{\lambda_1} \right|^k \right)$$

Hierboven heeft \pm de betekenis de bovengrens geldt voor $+u_1$ of $-u_1$. De teken kan mogelijk veranderen als een functie van k .

Stellingen van vorige les in de nieuwe notatie

Theorem (Convergentie methoden van machten)

Neem aan dat A reëel en symmetrisch is. De deelruimtes gedefinieerd door de vectoren $q^{(k)} \in \mathbb{R}^n$ convergeren naar de eigenruimte van de eigenvector u_1 :

$$d\left(\text{span}\{u_1\}, \text{span}\{q^{(k)}\}\right) = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right).$$

QR algoritme

- ▶ Neem aan dat $A \in \mathbb{R}^{n \times n}$ symmetrisch is.
- ▶ De basis QR algoritme werkt heel simpel. Initialiseer $A^{(0)} = A$ en doe de volgende iteraties:

$$A^{(k)} = R^{(k)} Q^{(k)}, \quad A^{(k-1)} = Q^{(k)} R^{(k)}$$

- ▶ Dit is equivalent aan de iteratie:

$$A^{(k)} = Q^{(k)T} A^{(k-1)} Q^{(k)}, \quad A^{(0)} = A.$$

- ▶ Onder bepaalde milde condities geldt:

$$\lim_{k \rightarrow \infty} A^{(k)} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Waarom werkt het QR algoritme?

Om dat te begrijpen, kijken we eerst naar een *generalisatie* van de methoden van machten...

Gelijktijdige iteraties

Gegeven een initiële vector $q^{(0)} \in \mathbb{R}^n$ met $\|q^{(0)}\|_2 = 1$, doe de volgende iteraties voor $k = 1, 2, \dots$:

$$w^{(k)} = Aq^{(k-1)} \quad (1)$$

$$q^{(k)} = \frac{w^{(k)}}{\|w^{(k)}\|_2} \quad (2)$$

$$\lambda^{(k)} = q^{(k)T} A q^{(k)} \quad (3)$$

Gelijktijdige iteraties

Gegeven een initiële matrix¹ $\hat{Q}^{(0)} \in \mathbb{R}^{n \times p}$ met $\hat{Q}^{(0)T} \hat{Q}^{(0)} = I_p$, doe de volgende iteraties voor $k = 1, 2, \dots$:

$$W^{(k)} = A\hat{Q}^{(k-1)} \quad (1)$$

$$\hat{Q}^{(k)} R^{(k)} = W^{(k)} \quad (2)$$

$$\Lambda^{(k)} = \hat{Q}^{(k)T} A \hat{Q}^{(k)} \quad (3)$$

¹Het gebruik van het dakje wordt later duidelijk.

Gelijktijdige iteraties

Gegeven een initiële matrix¹ $\hat{Q}^{(0)} \in \mathbb{R}^{n \times p}$ met $\hat{Q}^{(0)T} \hat{Q}^{(0)} = I_p$, doe de volgende iteraties voor $k = 1, 2, \dots$:

$$W^{(k)} = A\hat{Q}^{(k-1)} \quad (1)$$

$$\hat{Q}^{(k)} R^{(k)} = W^{(k)} \quad (2)$$

$$\Lambda^{(k)} = \hat{Q}^{(k)T} A \hat{Q}^{(k)} \quad (3)$$

Huiswerk. Overtuig jezelf dat gelijktijdige iteraties in essentie reduceert naar de methoden van machten voor $p = 1$.

¹Het gebruik van het dakje wordt later duidelijk.

Gelijktijdige iteraties - convergentie invariante deelruimten

- ▶ Neem weer aan dat A reëel en symmetrisch is.
- ▶ Laat $\{u_k\}_{k=1}^n$ de eigenvectoren zijn van A met $\|u_k\|_2 = 1$.
Neem aan dat eigenwaarden gesorteerd zijn op grootte:

$$|\lambda_1| > \dots > |\lambda_p| > |\lambda_{p+1}| \geq \dots \geq |\lambda_n| \geq 0$$

- ▶ Druk $\hat{Q}^{(0)} \in \mathbb{R}^{n \times p}$ uit in termen van de eigenvectoren:

$$\hat{Q}^{(0)} = \left[\begin{array}{ccc|ccc} u_1 & \cdots & u_p & u_{p+1} & \cdots & u_n \end{array} \right] \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

- ▶ We nemen aan dat $C_1 \in \mathbb{R}^{p \times p}$ inverteerbaar is.
- ▶ Het is aan te tonen² dat $\text{Im}(\hat{Q}^{(k)}) = \text{Im}(A^k \hat{Q}^{(0)})$. Dus,

$$\hat{Q}^{(k)} = A^k \hat{Q}^{(0)} D_k$$

voor een bepaalde inverteerbare $D_k \in \mathbb{R}^{p \times p}$.

²vraag in practicum

Gelijktijdige iteraties - convergentie invariante deelruimten

- ▶ Neem weer aan dat A reëel en symmetrisch is.
- ▶ Laat $\{u_k\}_{k=1}^n$ de eigenvectoren zijn van A met $\|u_k\|_2 = 1$.
Neem aan dat eigenwaarden gesorteerd zijn op grootte:

$$|\lambda_1| > \dots > |\lambda_p| > |\lambda_{p+1}| \geq \dots \geq |\lambda_n| \geq 0$$

- ▶ Druk $\hat{Q}^{(0)} \in \mathbb{R}^{n \times p}$ uit in termen van de eigenvectoren:

$$\hat{Q}^{(0)} = [U_1 \quad U_2] \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

- ▶ We nemen aan dat $C_1 \in \mathbb{R}^{p \times p}$ inverteerbaar is.
- ▶ Het is aan te tonen² dat $\text{Im}(\hat{Q}^{(k)}) = \text{Im}(A^k \hat{Q}^{(0)})$. Dus,

$$\hat{Q}^{(k)} = A^k \hat{Q}^{(0)} D_k$$

voor een bepaalde inverteerbare $D_k \in \mathbb{R}^{p \times p}$.

²vraag in practicum

Gelijktijdige iteraties - convergentie invariante deelruimten

We leiden het volgende af:

$$\begin{aligned}\hat{Q}^{(k)} &= A^k \hat{Q}^{(0)} D_k \\&= \left[\lambda_1^k u_1 \quad \cdots \quad \lambda_p^k u_p \mid \lambda_{p+1}^k u_{p+1} \quad \cdots \quad \lambda_n^k u_n \right] \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} D_k \\&= \left(\left[\left(\frac{\lambda_1}{\lambda_p} \right)^k u_1 \quad \cdots \quad u_p \right] + \right. \\&\quad \left. \left[\left(\frac{\lambda_{p+1}}{\lambda_p} \right)^k u_{p+1} \quad \cdots \quad \left(\frac{\lambda_n}{\lambda_p} \right)^k u_n \right] C_2 C_1^{-1} \right) \lambda_p^k C_1 D_k \\&= \left(\left[\left(\frac{\lambda_1}{\lambda_p} \right)^k u_1 \quad \cdots \quad u_p \right] + O \left(\left(\frac{\lambda_{p+1}}{\lambda_p} \right)^k \right) \right) \lambda_p^k C_1 D_k\end{aligned}$$

Gelijktijdige iteraties - convergentie invariante deelruimten

- ▶ Omdat $\lambda_p^k C_1 D_k$ niet-singulier is, geldt:

$$\operatorname{Im} \left(\hat{Q}^{(k)} \right) = \operatorname{Im} \left(\left[\left(\frac{\lambda_1}{\lambda_p} \right)^k u_1 \quad \cdots \quad u_p \right] + O \left(\left(\frac{\lambda_{p+1}}{\lambda_p} \right)^k \right) \right)$$

- ▶ Hoe ver ligt de invariante deelruimte $\operatorname{Im} (U_1)$ van $\operatorname{Im} \left(\hat{Q}^{(k)} \right)$?
- ▶ Omdat $\left(\frac{\lambda_i}{\lambda_p} \right)^k > 1$ voor $i = 1, \dots, p-1$, zal de verstoring het grootst zijn op u_p .
- ▶ Hiermee kan aangetoond worden dat:

$$d \left(\operatorname{Im} (U_1), \operatorname{Im} \left(\hat{Q}^{(k)} \right) \right) = O \left(\left(\frac{\lambda_{p+1}}{\lambda_p} \right)^k \right)$$

Gelijktijdige iteraties - convergentie naar eigenruimtes

- De kolommen van:

$$\hat{Q}^{(k)} = \begin{bmatrix} \hat{q}_1^{(k)} & \hat{q}_2^{(k)} & \dots & \hat{q}_p^{(k)} \end{bmatrix}$$

beschrijven een deelruimte die in zekere zin convergeren naar de invariante deelruimte $\text{Im}(U_1)$.

- Maar wat kan er gezegd worden over de individuele vectoren $\hat{q}_i^{(k)}$ zelf?

Die convergeren naar de eigenvectoren!

Gelijktijdige iteraties - convergentie naar eigenruimtes

De redenering gaat als volgt:

- ▶ We kunnen gelijktijdige iteraties ook toepassen op een sub-matrix van $\hat{Q}^{(0)}$, neem bijv. de eerste $p - 1$ kolommen:

$$\hat{Q}^{(0)} = \begin{bmatrix} \tilde{Q}^{(0)} & \hat{q}_p^{(k)} \end{bmatrix}$$

- ▶ De iteraties $\tilde{Q}^{(k)}$ zullen identiek zijn aan de eerste $p - 1$ kolommen van $\hat{Q}^{(k)}$ (op het \pm teken na dan...)
- ▶ Een herhaling van de voorgaande analyse zal aangeven dat $\text{Im}(\tilde{Q}^{(k)})$ convergeert (onder specifieke voorwaarden) naar de invariante deelruimte beschreven door de eerste $p - 1$ eigenvectoren: u_1, u_2, \dots, u_p .
- ▶ Trek de zelfde conclusies voor de eerste $p - 2$, eerste $p - 3$ vectoren, etc...

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$\begin{aligned}W^{(k)} &= A\hat{Q}^{(k-1)} \\ \hat{Q}^{(k)}R^{(k)} &= W^{(k)} \\ \Lambda^{(k)} &= \hat{Q}^{(k)T}A\hat{Q}^{(k)}\end{aligned}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$\begin{aligned}W^{(k)} &= \hat{Q}^{(k-1)}\Lambda^{(k-1)} \\ \hat{Q}^{(k)}R^{(k)} &= W^{(k)} \\ \Lambda^{(k)} &= \hat{Q}^{(k)T}A\hat{Q}^{(k)}\end{aligned}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$W^{(k)} = \hat{Q}^{(k-1)} \Lambda^{(k-1)}$$

$$\hat{Q}^{(k)} R^{(k)} = W^{(k)}$$

$$\Lambda^{(k)} = \hat{Q}^{(k)T} \hat{Q}^{(k-1)} \hat{Q}^{(k-1)T} A \hat{Q}^{(k-1)} \hat{Q}^{(k-1)T} \hat{Q}^{(k)}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$\begin{aligned}W^{(k)} &= \hat{Q}^{(k-1)}\Lambda^{(k-1)} \\ \hat{Q}^{(k)}R^{(k)} &= W^{(k)} \\ \Lambda^{(k)} &= \left(\hat{Q}^{(k-1)T}\hat{Q}^{(k)}\right)^T \Lambda^{(k-1)} \left(\hat{Q}^{(k-1)T}\hat{Q}^{(k)}\right)\end{aligned}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$\begin{aligned}\hat{Q}^{(k)} R^{(k)} &= \hat{Q}^{(k-1)} \Lambda^{(k-1)} \\ \Lambda^{(k)} &= \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right)^T \Lambda^{(k-1)} \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right)\end{aligned}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $\hat{Q}^{(0)} = I_n$

$$\left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right) R^{(k)} = \Lambda^{(k-1)}$$

$$\Lambda^{(k)} = \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right)^T \Lambda^{(k-1)} \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right)$$

Plot twist: Gelijktijdige iteraties = QR iteraties

We zien dus dat $\Lambda^{(k)}$ aan de volgende iteraties voldoet:

$$\begin{aligned} \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right) R^{(k)} &= \Lambda^{(k-1)} \\ \Lambda^{(k)} &= R^{(k)} \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)} \right) \end{aligned}$$

Plot twist: Gelijktijdige iteraties = QR iteraties

$$\begin{aligned}\left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)}\right) R^{(k)} &= \Lambda^{(k-1)} \\ \Lambda^{(k)} &= R^{(k)} \left(\hat{Q}^{(k-1)T} \hat{Q}^{(k)}\right)\end{aligned}$$

- ▶ De startwaarde $\Lambda^{(0)} = A$ levert dezelfde iteraties op als met $\hat{Q}^{(0)} = I_n$.
- ▶ Definieer: $Q^{(k)} = \hat{Q}^{(k-1)T} \hat{Q}^{(k)}$.
- ▶ Her-definieer: $\Lambda^{(k)} = A^{(k)}$

Plot twist: Gelijktijdige iteraties = QR iteraties

Start met $A^{(0)} = A$. En doe de iteraties:

$$\begin{aligned} Q^{(k)} R^{(k)} &= A^{(k-1)} \\ A^{(k)} &= R^{(k)} Q^{(k)} \end{aligned}$$

Oefening 1

Toon aan dat:

$$A^k = \hat{Q}^{(k)} \hat{R}^{(k)}$$

waarbij:

$$\hat{Q}^{(k)} = Q^{(1)} Q^{(2)} \dots Q^{(k)}, \quad \hat{R}^{(k)} = R^{(k)} R^{(k-1)} \dots R^{(1)}.$$

Oefening 1 - antwoord

- Om aan te tonen dat $\hat{Q}^{(k)} = Q^{(1)} Q^{(2)} \dots Q^{(k)}$:

$$\begin{aligned}\hat{Q}^{(k)} &= \hat{Q}^{(k-1)} \hat{Q}^{(k-1)T} \hat{Q}^{(k)} \\ &= \hat{Q}^{(k-1)} Q^{(k)} \\ &= Q^{(1)} Q^{(2)} \dots Q^{(k)}\end{aligned}$$

- Vervolgens hebben we:

$$\begin{aligned}\hat{Q}^{(1)} R^{(1)} &= A I_n \\ A \hat{Q}^{(1)} R^{(1)} &= A^2 \\ \hat{Q}^{(2)} R^{(2)} R^{(1)} &= A^2 \\ &\vdots \\ \hat{Q}^{(k)} \underbrace{R^{(k)} \dots R^{(2)} R^{(1)}}_{\hat{R}^{(k)}} &= A^k\end{aligned}$$

Een naïeve implementatie van het QR algoritme

- ▶ Een directe toepassing van de basis QR algoritme:

$$A^{(k)} = R^{(k)}Q^{(k)}, \quad A^{(k-1)} = Q^{(k)}R^{(k)}$$

is heel langzaam

- ▶ In elke cyclus wordt een QR-decompositie berekend. Dit kost $O(n^3)$.
- ▶ Laten we aannemen dat we ongeveer $O(n)$ iteraties nodig hebben voor een goede benadering.
- ▶ Dan komen snel uit op $O(n^4)$ in totaal!
- ▶ Het kan beter...

Hoe maken we het QR algoritme effectief?

Een paar elementen:

- ▶ Reductie naar een Hessenberg matrix
- ▶ Het gebruik van “shifts”
- ▶ Het gebruik van deflatie (wordt niet besproken)

Meer volgende les!