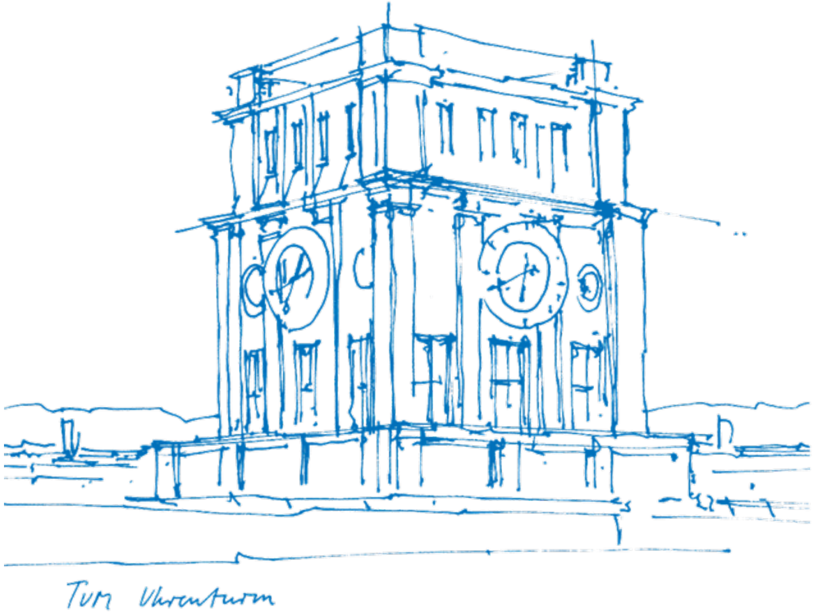
Lineare Algebra für Informatik - Woche 12

Cosmin Aprodu

Technische Universität München

Online, 01 Juli 2021



Diagonalisierung

Wiederholung: Sei $A \in K^{n \times n}$ diagonalisierbar und $S \in GL_n(K)$, dann:

- $A = S \cdot D \cdot S^{-1}$ oder auch $D = S^{-1} \cdot A \cdot S$.
- Das charakteristische Polynom χ_A zerfällt in Linearfaktoren.
- Für alle Eigenwerte λ_i gilt: $m_g(\lambda_i) = m_a(\lambda_i)$

$\rightarrow$ *Wie bestimmt man die Matrizen **S** und **D**?*

(1) Die Basisvektoren der Eigenräume sind die *Spalten* der Matrix ***S***.

(2) Die Matrix ***D*** enthält auf der *Hauptdiagonale* die Eingenwerte (Anzahl durch m_a gegeben), sonst reiner Nullen.
Alternativ, $S^{-1} \cdot A \cdot S$ berechnen (dauert länger ☹).

Bestimmung der Google-Matrix

Wir betrachten die Seiten $P_1, \dots, P_n$ der Internets und definieren die **Weblink-Matrix** (Adjazenz-Matrix des Web-Graphen) $W = (w_{i,j}) \in \mathbb{R}^{n\times n}$ durch:

$$w_{i,j} = \begin{cases} 1, & \text{falls } P_i \text{ einen Link auf } P_j \text{ enthält} \\ 0, & \text{sonst} \end{cases}$$

(1) Wir definieren die Matrix $H = (h_{i,j}) \in \mathbb{R}^{n\times n}$, indem wir in W jede Zeile durch die Anzahl der Einträge gleich Eins teilen, also:

$$h_{i,j} = \begin{cases} \frac{1}{\mathrm{outdeg}(i)}, & \text{falls } P_i \text{ einen Link auf } P_j \text{ enthält} \\ 0, & \text{sonst} \end{cases}$$

Wenn wir die Wichtigkeit der Seite P_i mit α_i bezeichnen, dann erfüllt der Vektor $p := (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^{1\times n}$ die Gleichung:

$$p \cdot H = p \;\Leftrightarrow\; H^T \cdot p^T = p^T$$

Also, p^T ist Eigenvektor von H^T zum Eigenwert 1. Die Wichtigkeit ist dann eindeutig festgelegt, wenn $\sum_i \alpha_i = 1$

Bestimmung der Google-Matrix (2)

(**2**) Wir definieren die Matrix $S = (s_{i,j}) \in \mathbb{R}^{n \times n}$, indem wir alle Nullzeilen mit Eins durch die Anzahl der Einträge ersetzen, also:

$$s_{i,j} = \begin{cases} \frac{1}{\operatorname{outdeg}(i)}, & \text{falls } P_i \text{ einen Link auf } P_j \text{ enthält} \\ 0, & \text{falls } \operatorname{outdeg}(i) > 0 \\ \frac{1}{n}, & \text{sonst} \end{cases}$$

(**3**) Daraus erhalten wir die **PageRank-Matrix** (auch **Google-Matrix**) $G \in \mathbb{R}^{n \times n}$:

$$G := (1-\alpha) \cdot S + \frac{\alpha}{n} \cdot \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}, \text{ wobei Google } \alpha = 0.15 \text{ gewählt hat.}$$

Als Interpretation in den Surfer-Zufallsmodell, wird es mit Wahrscheinlichkeit $(1-\alpha)$ ein Link auf der Seite geklickt und mit Wahrscheinlichkeit α auf eine beliebige Seite gesprungen.

Bestimmung des PageRank-Vektors

Der **PageRank-Vektor** p kann man erhalten, indem man den Eigenvektor zum Eigenwert 1 von G^T ausrechnet.
$\rightarrow$ Aufgrund der Konstruktion der Matrix G sind wir sicher, dass die Matrix G^T immer 1 als Eigenwert haben wird.

Bemerkung: Berechnungsmöglichkeiten für p:

- Durch *Gauß-Elimination* (ungeeignet für große n, $O(n^3)$ Operationen werden benötigt)
- Folgende Beobachtung verwenden:

$$\lim_{k\to\infty} G^k = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}$$