FLEURIANNE BERTRAND, CHRISTIAN CLASON WS 2016/17
BLATT 8

NUMERISCHE MATHEMATIK FÜR DAS LEHRAMT

GOOGLES PAGERANK

Google sortiert die Ergebnisse einer Suchanfrage (unter anderem) nach der *Wichtigkeit* der Seiten. Dabei ist die Wichtigkeit x_i der Seite S_i bestimmt durch die Wichtigkeit aller Seiten, die einen Link auf die Seite S_i setzen. Hat die Seite S_j also die Wichtigkeit x_j, und führt einer von insgesamt n_j Links auf die Seite S_i, so erhält S_i den Beitrag $\frac{x_j}{n_j}$. Die Summe über alle diese Beiträge ergibt die Wichtigkeit x_i. Definiert man eine Matrix A mit Hilfe von

$$a_{ij} = \begin{cases} \frac{1}{n_j} & S_j \text{ verlinkt } S_i\,, \\ 0 & S_j \text{ verlinkt nicht } S_i\,, \end{cases}$$

so ist der Vektor der Wichtigkeiten Lösung des Eigenwertproblems $Ax = x$. Hat jede Seite mindestens einen Link auf eine andere Seite, so summiert sich jede Spalte der Matrix A zu 1; solche Matrizen nennt man *(Spalten-)stochastisch.*

Hausaufgaben (Bearbeitung bis 11.01.2017, pünktlich zum Beginn der Übung)

H 6.1 Zeigen Sie, dass jede stochastische Matrix als betragsgrößten Eigenwert $\lambda = 1$ hat.

Hinweis: Setzen Sie $e = (1, \dots, 1)^T$ *und betrachten Sie* $A^T e$.

Der zugehörige normalisierte Eigenvektor v kann also mit Hilfe der Vektoriteration

$$v^{(k)} = \frac{A^k x}{\|A^k x\|_1}$$

aus einem (zufälligen) Startvektor x berechnet werden.

H 6.2 Gegeben sei eine Menge von Webseiten S_i, $1 \leqslant i \leqslant 6$ mit folgenden Links:

Seite	S_1	S_2	S_3	S_4	S_5	S_6
hat Link auf Seite	S_2, S_5	S_3, S_4	S_4, S_5, S_6	S_1	S_1	S_5

Skizzieren Sie die Seiten und ihre Verknüpfungen. Berechnen Sie den zugehörigen Vektor der Wichtigkeiten mit Hilfe der Vektoriteration. Wie unterscheidet sich die Reihenfolge der Seiten von derjenigen, die nur die Anzahl der Links auf die jeweilige Seite zählt?

Eine Schwierigkeit dieses Ansatzes ist, dass die Existenz eines eindeutigen (normierten) Eigenvektors nicht garantiert ist. Zerfällt das Netz der Seiten in Teile, die nicht untereinander verbunden sind (ist es also nicht möglich, von jeder Seite durch Folgen von Links zu jeder anderen zu gelangen), so hat der Eigenraum zum Eigenwert 1 die Dimension $k > 1$.

H 6.3 Betrachten Sie die vier Seiten S_1, S_2, S_3, S_4, wobei sich lediglich S_1 und S_2 sowie S_3 und S_4 gegenseitig verlinken. Finden Sie zwei linear unabhängige Eigenvektoren zum Eigenwert 1.

Hinweis: Betrachten Sie die disjunkten "Teilnetze" $\{S_1, S_2\}$ *und* $\{S_3, S_4\}$ *separat.*

Tatsächlich sind einer Studie aus dem Jahr 2000 zufolge ein Fünftel aller Webseiten nicht Teil einer zusammenhängenden Struktur. Ein sinnvolles Verfahren muss also auch in diesem Falle einen eindeutigen Wichtigkeitsvektor finden. Folgende Modifikation garantiert dies: Statt A verwendet man für $0 \leqslant m \leqslant 1$ und $e = (1, \dots, 1)^T$ die Matrix

$$M = (1 - m)A + m\frac{1}{n}ee^T\,.$$

(ee^T ist also eine Matrix mit den Einträgen $e_{ij} = 1$.)

H 6.4 Zeigen Sie, dass für eine stochastische Matrix A auch M stochastisch ist.

Da die Matrix M nur positive Einträge hat, folgt aus dem Satz von Perron–Frobenius, dass zum betragsgrößten Eingewert nur ein linear unabhängiger Eigenvektor existiert, der ausschließlich positive Komponenten hat.

H 6.5 Betrachten Sie noch einmal die Matrix A aus H 6.3. Welchen (normierten) Eigenvektor zum Eigenwert 1 hat die Matrix $M = \frac{1}{2}A + \frac{1}{2n}ee^T$?

P 6.6 Zum Abschluss sollen Sie dieses Verfahren anhand einer Sammlung realer Webseiten prüfen. Gehen Sie wie folgt vor:

a) Laden Sie sich die Datei `webseiten.mat` von der Vorlesungs-Webseite herunter.

b) Starten Sie Matlab/Octave, und wechseln Sie in das Verzeichnis, in dem Sie die Datei abgespeichert haben.

c) Erzeugen Sie durch den Befehl `load webseiten.mat` eine Matrix `A` der gewichteten Verknüpfungen und ein Feld `U` von URLs.

d) Implementieren Sie die Vektoriteration, um den Eigenvektor zum betragsgrößten Eigenwert von $M = (1-m)A + m\frac{1}{n}ee^T$ mit $m = 0.15$ zu berechnen. (Sie sollten die Iteration solange durchführen, bis $\|v^{(k)} - v^{(k-1)}\| < 10^{-6}$ gilt.)

e) Die Matrix M ist voll besetzt, so dass eine Matrix-Vektor-Multiplikation den Aufwand $\mathcal{O}(n^2)$ hat. Dagegen ist A dünn besetzt; die Multiplikation ist also in $\mathcal{O}(n)$ machbar. Können Sie die Matrixmultiplikation im Iterationsschritt berechnen, ohne die Matrix M explizit aufstellen zu müssen?
Hinweis: Die Iterierten $v^{(k)}$ sind bezüglich $\|\cdot\|_1$ normiert.

f) Berechnen Sie den Vektor der Wichtigkeiten, und finden Sie die wichtigste Webseite. (Der Befehl `[wert,index]=max(x)` liefert Wert und Index des größten Eintrags von `x`; `U(index)` gibt dann die URL der zugehörigen Webseite aus.)