



Viertes Kapitel

Die Mathematik des PageRank von Google

Der berühmte und schillernde Mathematiker Paul Erdos (1913-96) sprach über das Große Buch, ein imaginäres Buch, in dem er sich vorstellte, dass Gott die elegantesten und schönsten Beweise der Welt aufbewahrt. Im Jahr 2002 gab Graham Farmelo vom Londoner Science Museum ein ähnliches Buch heraus, ein Buch der schönen Gleichungen, und leistete einen Beitrag dazu. *It Must Be Beautiful: Great Equations of Modern Science* [73] ist eine Sammlung von 11 Aufsätzen über die größten wissenschaftlichen Gleichungen, Gleichungen wie $E = hf$ und $E = mc^2$. Die Autoren wurden aufgefordert, ihre Antworten auf die schwierige Frage zu geben, was eine Gleichung großartig macht. Einer der Autoren, Frank Wilczek, hat ein Zitat von Heinrich Hertz über die Maxwellsche Gleichung aufgenommen:

Man kann sich des Gefühls nicht erwehren, dass diese mathematischen Formeln eine eigenständige Existenz und eine eigene Intelligenz haben, dass sie klüger sind als wir, klüger sogar als ihre Entdecker, dass wir mehr aus ihnen herausholen, als ursprünglich in sie hineingelegt wurde.

Wir wollen nicht behaupten, dass die in diesem Kapitel vorgestellte PageRank-Gleichung,

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T (\boldsymbol{\alpha} \mathbf{S} + (1 - \alpha) \mathbf{E}),$$

einen Platz in Farmelos Buch neben Einsteins Relativitätstheorie verdient hat, finden wir Hertz' Aussage sehr treffend. Mit der obigen einfachen PageRank-Formel kann man viel anfangen - Google hat es jedenfalls getan. Da Schönheit im Auge des Betrachters liegt, lassen wir Sie entscheiden, ob die PageRank-Formel das Adjektiv "*schön*" verdient oder nicht. Wir hoffen, dass die nächsten Kapitel Sie davon überzeugen werden, dass es so ist.

In Kapitel 3 haben wir die PageRank-These mit Worten dargestellt: eine Seite ist wichtig, wenn andere wichtige Seiten auf sie verweisen. Nun ist es an der Zeit, diese Worte in mathematische Gleichungen zu übersetzen. Diese Übersetzung zeigt, dass die PageRank-Wichtigkeitswerte eigentlich die stationären Werte einer riesigen Markov-Kette sind, und folglich erklärt die Markov-Theorie viele interessante Eigenschaften des von Google verwendeten, elegant einfachen PageRank-Modells.

Dies ist das erste der mathematischen Kapitel. Viele der mathematischen Begriffe in jedem Kapitel werden im Kapitel Mathematik (Kapitel 15) erklärt. Wenn Begriffe, die im Kapitel Mathematik vorkommen, im Text eingeführt werden, sind sie kursiv gedruckt, um Sie daran zu erinnern, dass Definitionen und weitere Informationen in Kapitel 15 zu finden sind.

4.1 DIE URSPRÜNGLICHE SUMMENFORMEL FÜR DEN PAGERANK

Brin und Page, die Erfinder des PageRank,¹ begannen mit einer einfachen Summationsgleichung, deren Wurzeln eigentlich aus der bibliometrischen Forschung stammen, der Analyse der Zitationsstruktur unter wissenschaftlichen Arbeiten. Der PageRank einer Seite P_i , bezeichnet als $r(P_i)$, ist die Summe der PageRanks aller Seiten, die auf P_i verweisen.

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|}, \tag{4.1.1}$$

wobei B_{P_i} die Menge der Seiten ist, die auf P_i verweisen (in den Worten von Brin und Page: Backlinks zu P_i), und $|P_j|$ die Anzahl der Outlinks von der Seite P_j ist. Beachten Sie, dass der PageRank der Seiten, die auf P_j verweisen, in Gleichung (4.1.1) durch die Anzahl der Empfehlungen von P_j , bezeichnet als $|P_j|$, gemildert wird. Das Problem bei Gleichung (4.1.1) ist, dass die $r(P_j)$ Werte, die PageRanks der Seiten, die auf die Seite P_i verlinken, nicht bekannt sind. Um dieses Problem zu umgehen, verwendeten Brin und Page ein iteratives Verfahren. Das heißt, sie gingen davon aus, dass zu Beginn alle Seiten den gleichen PageRank haben (z. B. $1/n$, wobei n die Anzahl der Seiten im Google-Index des Webs ist). Nun wird die Regel in Gleichung (4.1.1) befolgt, um $r(P_i)$ für jede Seite P_i im Index zu berechnen. Die Regel in Gleichung (4.1.1) wird nacheinander angewandt, wobei die Werte der vorherigen Iteration in $r(P_j)$ eingesetzt werden. Wir führen weitere Begriffe ein, um ~~dieses~~ *iterative Verfahren* zu definieren. Sei $r_{k+1}(P_i)$ der PageRank der Seite P_i bei Iteration $k + 1$. Dann,

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|}. \tag{4.1.2}$$

Dieser Prozess wird mit $r_0(P_i) = 1/n$ für alle Seiten P_i eingeleitet und wiederholt, in der Hoffnung, dass die PageRank-Werte schließlich zu einigen endgültigen stabilen Werten konvergieren werden. Wendet man Gleichung (4.1.2) auf das winzige Netz in Abbildung 4.1 an, erhält man nach einigen Iterationen die folgenden Werte für die PageRanks.

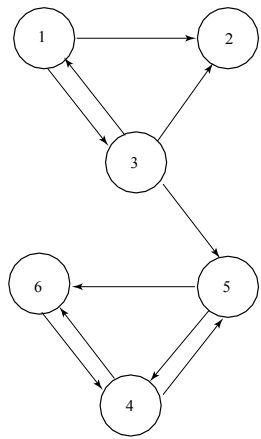


Abbildung 4.1 Gerichtetes Diagramm, das ein Netz von sechs Seiten darstellt

¹Das Patent für PageRank wurde 1998 von Larry Page angemeldet und 2001 erteilt (US-Patent Nr. 6285999), so dass der Name PageRank einen doppelten Bezug sowohl zu Webseiten als auch zu einem seiner Gründer hat.

Tabelle 4.1 Erste Iterationen mit (4.1.2) auf Abbildung 4.1

Iteration 0	Iteration 1	Iteration 2	Rang bei Iter. 2
$r_0(P_1) = 1/6$	$r_1(P_1) = 1/18$	$r_2(P_1) = 1/36$	5
$r_0(P_2) = 1/6$	$r_1(P_2) = 5/36$	$r_2(P_2) = 1/18$	4
$r_0(P_3) = 1/6$	$r_1(P_3) = 1/12$	$r_2(P_3) = 1/36$	5
$r_0(P_4) = 1/6$	$r_1(P_4) = 1/4$	$r_2(P_4) = 17/72$	1
$r_0(P_5) = 1/6$	$r_1(P_5) = 5/36$	$r_2(P_5) = 11/72$	3
$r_0(P_6) = 1/6$	$r_1(P_6) = 1/6$	$r_2(P_6) = 14/72$	2

4.2 MATRIXDARSTELLUNG DER SUMMATIONSGLEICHUNGEN

Die Gleichungen (4.1.1) und (4.1.2) berechnen den PageRank für jede Seite einzeln. Mit Hilfe von *Matrizen* ersetzen wir das langwierige Symbol und berechnen bei jeder Iteration einen PageRank-Vektor, der einen einzigen $1 \times n$ -Vektor verwendet, um die PageRank-Werte für alle Seiten im Index zu speichern. Zu diesem Zweck führen wir eine $n \times n$ -Matrix \mathbf{H} und einen $1 \times n$ -Zeilenvektor $\boldsymbol{\pi}^T$ ein. Die Matrix \mathbf{H} ist eine zeilennormierte Hyperlink-Matrix mit $H_{ij} = 1/|P_i|$, wenn es einen Link von Knoten i zu Knoten j gibt, und ansonsten 0. Obwohl \mathbf{H} die gleiche Nicht-Null-Struktur hat wie die binäre *Adjazenzmatrix* für den Graphen (in der Matlab Crawler m-Datei auf Seite 17 als \mathbf{L} bezeichnet), sind ihre Nicht-Null-Elemente Wahrscheinlichkeiten. Betrachten wir noch einmal den kleinen Webgraphen aus Abbildung 4.1.

Die H-Matrix für diesen Graphen lautet

$$\mathbf{H} = \begin{matrix} & \begin{matrix} P_1 & P_2 & P_3 & P_4 & P_5 & P_6 \end{matrix} \\ \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \end{matrix} & \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1/2 \end{pmatrix} \end{matrix}.$$

Die Nicht-Null-Elemente der Zeile i entsprechen den verlinkenden Seiten der Seite i , während die Nicht-Null-Elemente der Spalte i den verlinkenden Seiten der Seite i entsprechen. Wir führen nun einen Zeilenvektor $\boldsymbol{\pi}^{(k)T}$ ein, der den PageRank-Vektor in der Iteration k^{th} darstellt. Unter Verwendung dieser Matrixnotation kann Gleichung (4.1.2) kompakt geschrieben werden als

$$\boldsymbol{\pi}^{(k+1)T} = \boldsymbol{\pi}^{(k)T} \mathbf{H}. \tag{4.2.1}$$

Wenn Sie möchten, können Sie am Beispiel von Abbildung 4.1 überprüfen, ob die Iterationen von Gleichung (4.2.1) mit denen von Gleichung (4.1.2) übereinstimmen.

Die Matrixgleichung (4.2.1) führt zu einigen unmittelbaren Beobachtungen.

1. Jede Iteration von Gleichung (4.2.1) beinhaltet eine Vektor-Matrix-Multiplikation, die im Allgemeinen $O(n^2)$ -Rechnungen erfordert, wobei n die Größe der quadratischen Matrix \mathbf{H} ist.
2. \mathbf{H} ist eine sehr *spärliche* Matrix (ein großer Teil ihrer Elemente ist 0), da die meisten Webseiten nur auf eine Handvoll anderer Seiten verweisen. Spärliche Matrizen, wie die in Abbildung 4.2 gezeigte, sind aus mehreren Gründen willkommen. Erstens benötigen sie nur minimalen Speicherplatz, da spärliche Speichersysteme, die nur die Nicht-Null-Elemente der

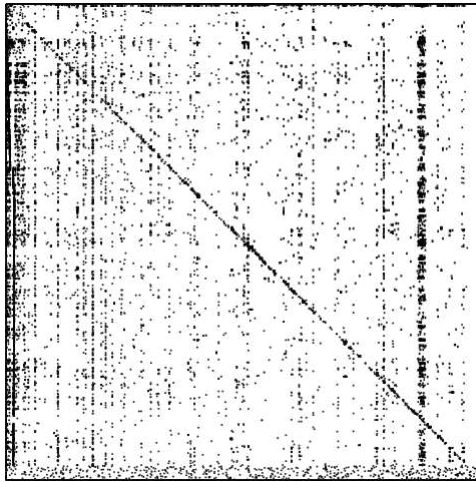


Abbildung 4.2 Beispiel für eine dünnbesetzte Matrix. Die Elemente, die nicht Null sind, werden durch Pixel dargestellt.

Matrix und deren Lage [145], existieren. Zweitens erfordert die Vektor-Matrix-Multiplikation mit einer spärlichen Matrix viel weniger Aufwand als die $O(n^2)$ dichte Berechnung. Tatsächlich erfordert sie $O(nnz(\mathbf{H}))$, wobei $nnz(\mathbf{H})$ die Anzahl der Nicht-Nullen in

\mathbf{H} . Schätzungen zeigen, dass die durchschnittliche Webseite etwa 10 Outlinks hat, was bedeutet, dass \mathbf{H} etwa $10n$ Nonzeros hat, im Gegensatz zu den n^2 Nonzeros in einer vollständig dichten Matrix. Das bedeutet, dass die Vektor-Matrix-Multiplikation von Gleichung (4.2.1) auf $O(n)$ Aufwand reduziert wird.

3. Der iterative Prozess der Gleichung (4.2.1) ist ein einfacher *linearer stationärer Prozess*, wie er in den meisten Vorlesungen zur numerischen Analyse untersucht wird [82, 127]. Tatsächlich handelt es sich um die klassische *Potenzmethode*, die auf \mathbf{H} angewendet wird.
4. \mathbf{H} sieht einer *stochastischen Übergangswahrscheinlichkeitsmatrix* für eine *Markov-Kette* sehr ähnlich. Die "**Dangling Nodes**" des Netzes, d. h. die Knoten ohne "Outlinks", bilden 0 Zeilen in der Matrix. Alle anderen Zeilen, die den nicht-verknüpften Knoten entsprechen, bilden stochastische Zeilen. Daher wird \mathbf{H} als *substochastisch* bezeichnet.

Diese vier Beobachtungen sind wichtig für die Entwicklung und Ausführung des PageRank-Modells, und wir werden im Laufe des Kapitels auf sie zurückkommen. Im Folgenden werden wir uns mit der iterativen Matrixgleichung (4.2.1) beschäftigen.

4.3 PROBLEME MIT DEM ITERATIVEN PROZESS

Die Gleichung (4.2.1) hat wahrscheinlich bei den Lesern, insbesondere bei unseren mathematischen Lesern, einige Fragen aufgeworfen. Zum Beispiel,

- Wird sich dieser iterative Prozess endlos fortsetzen oder wird er konvergieren?

- Unter welchen Umständen oder Eigenschaften von \mathbf{H} ist die Konvergenz garantiert?

- Wird sie zu etwas konvergieren, das im Zusammenhang mit dem PageRank-Problem sinnvoll ist?
- Konvergiert sie zu nur einem Vektor oder zu mehreren Vektoren?
- Hängt die Konvergenz vom Startvektor $\pi^{(0)T}$ ab?
- Wenn es irgendwann konvergiert, wie lange ist "irgendwann"? Das heißt, wie viele Iterationen können wir bis zur Konvergenz erwarten?

Wir werden diese Fragen in den nächsten Abschnitten beantworten. Unsere Antworten hängen jedoch davon ab, wie Brin und Page einige der Probleme gelöst haben, auf die sie bei ihrer Gleichung (4.2.1) gestoßen sind.

Brin und Page begannen den Iterationsprozess ursprünglich mit $\pi^{(0)T} = 1/n \mathbf{e}^T$, wobei \mathbf{e}^T der Zeilenvektor aller 1en ist. Bei der Verwendung von Gleichung (4.2.1) mit diesem Ausgangsvektor stießen sie sofort auf mehrere Probleme. Zum Beispiel gibt es das Problem der "**Rank Sinks**", also der Seiten, die bei jeder Iteration mehr und mehr PageRank anhäufen, die Punkte monopolisieren und sich weigern zu teilen. Im einfachen Beispiel von Abbildung 4.3 ist der baumelnde Knoten 3 eine Ranksenke. In dem komplizierteren Beispiel von Abbildung 4.1 ist die Gruppe der Knoten 4, 5,

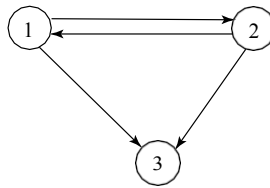


Abbildung 4.3 Einfacher Graph mit Ranksenke

und 6 verschwören sich, um den PageRank zu horten. Nach nur 13 Iterationen von Gleichung (4.2.1), $\pi^{(13)T} = (0 \ 0 \ 0 \ 2/3 \ 1/3 \ 1/5)$. Diese Verschwörung kann böswillig oder unabsichtlich sein. (Siehe die Anmerkungen zur Suchmaschinenoptimierung und zu Linkfarmen auf den Seiten 43 bzw. 52). Das Beispiel mit $\pi^{(13)T}$ zeigt auch ein weiteres Problem, das durch Senken verursacht wird. Wenn Knoten PageRank horten, kann es sein, dass einige Knoten keinen mehr haben. Daher ist es schwierig, Knoten nach ihren PageRank-Werten zu ordnen, wenn die Mehrheit der Knoten einen PageRank von 0 hat. Idealerweise sollte der PageRank-Vektor positiv sein, d. h. alle positiven Werte enthalten.

Außerdem gibt es das Problem der **Zyklen**. Betrachten Sie den einfachsten Fall in Abbildung 4.4. Seite

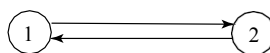


Abbildung 4.4 Einfacher Graph mit Zyklus

1 verweist nur auf Seite 2 und umgekehrt, wodurch eine Endlosschleife oder ein Zyklus entsteht. Angenommen, der iterative Prozess von Gleichung (4.2.1) wird mit $\pi^{(0)T} = (1 \ 0)$ ausgeführt. Die Iterate werden nicht konvergieren, egal wie lange der Prozess läuft. Die Iterate $\pi^{(k)T}$ wechselt unendlich oft zwischen $(1 \ 0)$, wenn k

gerade ist, und $(0, 1)$, wenn k ungerade ist.

4.4 EIN WENIG MARKOV-KETTEN-THEORIE

Bevor wir zu den Anpassungen von Brin und Page an Gleichung (4.2.1) kommen, die die Probleme des vorigen Abschnitts lösen, halten wir inne, um eine kleine Theorie für Markov-Ketten einzuführen. (Lesern, die mit Markov-Ketten weniger vertraut sind, empfehlen wir dringend, das Kapitel Mathematik, Kapitel 15, zu lesen, bevor sie fortfahren.) In den Beobachtungen 3 und 4 haben wir festgestellt, dass die Gleichung (4.2.1) der Potenzmethode ähnelt, die auf eine Markov-Kette mit der *Übergangswahrscheinlichkeitsmatrix* \mathbf{H} angewendet wird. Diese Beobachtungen sind sehr hilfreich, denn die Theorie der Markov-Ketten ist gut entwickelt,² und sehr gut auf das PageRank-Problem anwendbar. Mit der Markov-Theorie können wir Anpassungen an Gleichung (4.2.1) vornehmen, die wünschenswerte Ergebnisse, Konvergenzeigenschaften und ermutigende Antworten auf die Fragen auf Seite 34 gewährleisten. Insbesondere wissen wir, dass die Potenzmethode, die auf eine Markov-Matrix \mathbf{P} angewandt wird, für jeden beliebigen Startvektor zu einem einzigen positiven Vektor konvergiert, der als *stationärer Vektor* bezeichnet wird, solange \mathbf{P} *stochastisch*, *irreduzibel* und *aperiodisch* ist. (Aperiodizität plus Irreduzibilität impliziert Primitivität.) Daher können die PageRank-Konvergenzprobleme, die durch Senken und Zyklen verursacht werden, überwunden werden, wenn \mathbf{H} leicht modifiziert wird, so dass es eine Markov-Matrix mit diesen gewünschten Eigenschaften ist.

Markov-Eigenschaften, die den PageRank beeinflussen

Es gibt einen eindeutigen positiven PageRank-Vektor, wenn die Google-Matrix stochastisch und irreduzibel ist. Mit der zusätzlichen Eigenschaft der Aperiodizität konvergiert die Power-Methode zu diesem PageRank-Vektor, unabhängig vom Startvektor des iterativen Prozesses.

4.5 FRÜHE ANPASSUNGEN DES GRUNDMODELLS

Tatsächlich haben Brin und Page genau das getan. Sie beschreiben ihre Anpassungen am grundlegenden PageRank-Modell in ihren Originalarbeiten von 1998. Interessant ist, dass in keinem ihrer Papiere der Begriff "Markov-Kette" vorkommt, nicht ein einziges Mal. Auch wenn sie sich dessen 1998 nicht bewusst waren, wissen sie jetzt sicherlich, dass ihr ursprüngliches Modell mit Markov-Ketten in Verbindung steht, da Markov-Ketten-Forscher begeistert und stetig auf den PageRank-Zug aufgesprungen sind und eifrig an dem arbeiten, was manche die große Anwendung von Markov-Ketten nennen.

Anstatt Markov-Ketten und ihre Eigenschaften zur Beschreibung ihrer Anpassungen zu verwenden, nutzen Brin und Page den Begriff des **zufälligen Surfers**. Stellen Sie sich einen Websurfer vor, der wahllos der Hyperlinkstruktur des Webs folgt. Das heißt, wenn er auf eine Seite mit mehreren Links stößt, wählt er einen zufällig aus, verlinkt zu dieser neuen Seite und setzt diesen Prozess der zufälligen Entscheidung unbegrenzt fort. Langfristig ist der Anteil der Zeit, die der zufällige Surfer auf einer bestimmten Seite verbringt, ein Maß für die relative Bedeutung dieser Seite. Wenn er einen großen Teil seiner Zeit auf einer bestimmten Seite verbringt, dann muss er beim zufälligen Verfolgen der Hyperlink-Struktur des Webs immer wieder zu dieser Seite zurückgekehrt sein. Seiten, die er oft besucht, müssen wichtig sein, denn auf sie müssen andere wichtige Seiten verweisen. Leider stößt dieser zufällige Surfer auf einige Probleme. Er bleibt immer dann hängen, wenn er einen hängenden Knotenpunkt betritt.

Und im Web gibt es jede Menge solcher Knoten, z. B. PDF-Dateien, Bilddateien, Datentabellen usw. Um dieses Problem zu lösen, definieren Brin und Page

² Vor fast 100 Jahren, 1906, erfand Andrei Andrejewitsch Markow die Ketten, die ab 1926 seinen Namen trugen [20].

ihre erste Anpassung, die wir als **Stochastizitätsanpassung** bezeichnen, weil die 0^T Zeilen von \mathbf{H} durch $1/n \mathbf{e}^T$ ersetzt werden, wodurch \mathbf{H} stochastisch wird. Infolgedessen kann der zufällige Surfer, nachdem er einen Dangling-Knoten eingegeben hat, nun einen beliebigen Hyperlink zu einer beliebigen Seite herstellen. Für das winzige 6-Knoten-Netz in Abbildung 4.1 ist die **stochastische Matrix** mit der Bezeichnung \mathbf{S}

$$\mathbf{S} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Die mathematische Formulierung dieses Stochastik-Fixes zeigt, dass \mathbf{S} aus einer *Rang-1-Aktualisierung* von \mathbf{H} entsteht, d. h. $\mathbf{S} = \mathbf{H} + \mathbf{a}(1/n \mathbf{e}^T)$, wobei $a_i = 1$ ist, wenn die Seite i ein "Dangling Node" ist, und sonst 0. Der binäre Vektor \mathbf{a} wird als **"Dangling Node"-Vektor bezeichnet**. \mathbf{S} ist eine Kombination aus der rohen ursprünglichen Hyperlink-Matrix \mathbf{H} und einer Rang-Eins-Matrix $1/n \mathbf{a} \mathbf{e}^T$.

Diese Anpassung garantiert, dass \mathbf{S} stochastisch ist und somit die Übergangswahrscheinlichkeitsmatrix für eine Markov-Kette darstellt. Sie allein kann jedoch nicht die gewünschten Konvergenzergebnisse garantieren. (Das heißt, dass ein eindeutiges positives $\boldsymbol{\pi}^T$ existiert und dass die Gleichung (4.2.1) schnell zu diesem $\boldsymbol{\pi}^T$ konvergiert.) Brin und Page benötigten eine weitere Anpassung, diesmal eine **Primitivitätsanpassung**. Mit dieser Anpassung ist die resultierende Matrix stochastisch und *primitiv*. Eine primitive Matrix ist sowohl irreduzibel als auch aperiodisch. Der stationäre Vektor der Kette (in diesem Fall der PageRank-Vektor) existiert also, ist eindeutig und kann durch eine einfache Potenziteration gefunden werden. Brin und Page verwenden erneut den Random Surfer, um diese Markov-Eigenschaften zu beschreiben.

Das Argument der Zufallssurfer für die Primitivitätsanpassung lautet folgendermaßen. Es stimmt zwar, dass Surfer der Hyperlink-Struktur des Webs folgen, aber manchmal wird ihnen langweilig und sie verlassen die Hyperlink-Methode des Surfers, indem sie ein neues Ziel in die URL-Zeile des Browsers eingeben. Wenn dies geschieht, "teleportiert" sich der zufällige Surfer wie eine Star Trek-Figur auf die neue Seite, wo er wieder mit dem Hyperlink-Surfen beginnt, bis zur nächsten Teleportation usw. Um diesen Vorgang mathematisch zu modellieren, erfanden Brin und Page eine neue Matrix \mathbf{G} , die wie folgt lautet

$$\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) 1/n \mathbf{e} \mathbf{e}^T,$$

wobei α ein Skalar zwischen 0 und 1 ist. \mathbf{G} wird die **Google-Matrix** genannt. In diesem Modell ist α ein Parameter, der den Anteil der Zeit steuert, in der der zufällige Surfer den Hyperlinks folgt, anstatt zu teleportieren. Nehmen wir an, $\alpha = 0,6$. Dann folgt der zufällige Surfer zu 60 % der Zeit der Hyperlink-Struktur des Webs und die anderen 40 % der Zeit teleportiert er zu einer zufälligen neuen Seite. Das Teleportieren ist zufällig, weil die Teleportationsmatrix $\mathbf{E} = 1/n \mathbf{e} \mathbf{e}^T$ gleichmäßig ist, d. h. der Surfer hat die gleiche Wahrscheinlichkeit, beim Teleportieren zu einer beliebigen Seite zu springen.

Die Primitivitätsanpassung hat mehrere Konsequenzen.

- **G** ist *stochastisch*. Es ist die *konvexe Kombination* der beiden stochastischen Matrizen **S** und $\mathbf{E} = 1/n \mathbf{e} \mathbf{e}^T$.

- **G** ist *irreduzibel*. Jede Seite ist direkt mit jeder anderen Seite verbunden, so dass die Irreduzibilität trivialerweise erzwungen ist.
- **G** ist *aperiodisch*. Die Selbstschleifen ($G_{ii} > 0$ für alle i) erzeugen Aperiodizität.
- **G** ist *primitiv*, weil $G^k > 0$ für einige k . (Tatsächlich gilt dies für $k = 1$.) Dies impliziert, dass ein eindeutiges positives π^T existiert, und die auf **G** angewandte Potenzmethode konvergiert garantiert zu diesem Vektor.
- **G** ist völlig *dicht*, was aus rechnerischer Sicht sehr schlecht ist. Glücklicherweise kann **G** als eine *Rang-Eins-Aktualisierung* der sehr spärlichen Hyperlink-Matrix **H** geschrieben werden. Dies ist rechnerisch vorteilhaft, wie wir später in Abschnitt 4.6 zeigen.

$$\begin{aligned}\mathbf{G} &= \alpha \mathbf{S} + (1 - \alpha) \mathbf{I} / n \mathbf{e} \mathbf{e}^T \\ &= \alpha (\mathbf{H} + \mathbf{I} / n \mathbf{a} \mathbf{e}^T) + (1 - \alpha) \mathbf{I} / n \mathbf{e} \mathbf{e}^T \\ &= \alpha \mathbf{H} + (\alpha \mathbf{a} + (1 - \alpha) \mathbf{e}) \mathbf{I} / n \mathbf{e}^T.\end{aligned}$$

- **G** ist insofern künstlich, als die rohe Hyperlink-Matrix **H** zweimal verändert wurde, um die gewünschten Konvergenzeigenschaften zu erreichen. Einen stationären Vektor (also einen PageRank-Vektor) gibt es für **H** nicht, also haben Brin und Page kreativ gemogelt, um ihr gewünschtes Ergebnis zu erreichen. Für das zweifach modifizierte **G** gibt es einen eindeutigen PageRank-Vektor, und wie sich herausstellt, ist dieser Vektor bemerkenswert gut darin, Webseiten einen globalen Bedeutungswert zu geben.

Notation für das PageRank-Problem

H	sehr spärliche, rohe substochastische Hyperlink-
S	Matrix spärliche, stochastische,
G	höchstwahrscheinlich reduzierbare Matrix
E	vollständig dichte, stochastische, primitive Matrix, genannt Google-
n	Matrix vollständig dichte, Rang-Eins-Teleportationsmatrix
α	Anzahl der Seiten im Index der Suchmaschine = Reihenfolge von H , S ,
π^T	G , E
α^T	Skalierungsparameter zwischen 0 und 1
	stationärer Zeilenvektor von G , genannt PageRank-Vektor,
	binärer Dangling-Node-Vektor

Zusammenfassend lässt sich sagen, dass die angepasste PageRank-Methode von Google

$$\pi^{(k+1)T} = \pi^{(k)T} \mathbf{G}, \quad (4.5.1)$$

die einfach die auf **G** angewandte *Potenzmethode* ist.

Wir schließen diesen Abschnitt mit einem Beispiel. Kehren wir noch einmal zu Abbildung 4.1 zurück, für $\alpha = 0,9$,

die stochastische, primitive Matrix

\mathbf{G} ist

$$\mathbf{G} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{1}{9} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{1}{12} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1/60 & 7/15 & 7/15 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 19/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 7/15 & 1/60 & 7/15 \\ 1/60 & 1/60 & 1/60 & 1/60 & 11/12 & 1/60 \end{pmatrix}$$

Der PageRank-Vektor von Google ist der stationäre Vektor von \mathbf{G} und ist gegeben durch

$$\boldsymbol{\pi}^T = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \\ .03721 & .05396 & .04151 & .3751 & .206 & .2862 \end{matrix}$$

Die Interpretation von $\pi_1 = .03721$ ist, dass der zufällige Surfer in 3,721 % der Fälle die Seite 1 sieht. Daher können die Seiten in diesem winzigen Web nach ihrer Wichtigkeit wie folgt geordnet werden (4 6 5 2 3 1), d. h. Seite 4 ist die wichtigste Seite und Seite 1 die unwichtigste Seite, gemäß der PageRank-Definition von Wichtigkeit.

4.6 BERECHNUNG DES PAGERANK-VEKTORS

Das PageRank-Problem kann auf zwei Arten erklärt werden:

1. Lösen Sie das folgende Eigenvektorproblem für $\boldsymbol{\pi}^T$.

$$\begin{aligned} \boldsymbol{\pi}^T &= \boldsymbol{\pi}^T \mathbf{G}, \\ \boldsymbol{\pi}^T \mathbf{e} &= 1. \end{aligned}$$

2. Lösen Sie das folgende *lineare homogene System* für $\boldsymbol{\pi}^T$.

$$\begin{aligned} \boldsymbol{\pi}^T (\mathbf{I} - \mathbf{G}) &= \mathbf{0}^T, \\ \boldsymbol{\pi}^T \mathbf{e} &= 1. \end{aligned}$$

Im ersten System besteht das Ziel darin, den normalisierten *dominanten linken Eigenvektor* von \mathbf{G} zu finden, der dem *dominanten Eigenwert* $\lambda_1 = 1$ entspricht. (\mathbf{G} ist eine stochastische Matrix, daher ist $\lambda_1 = 1$.) Im zweiten System besteht das Ziel darin, den normalisierten linken Nullvektor von $\mathbf{I} - \mathbf{G}$ zu finden. Für beide Systeme gilt die Normalisierungsgleichung $\boldsymbol{\pi}^T \mathbf{e} = 1$, die sicherstellt, dass $\boldsymbol{\pi}^T$ ein Wahrscheinlichkeitsvektor ist. In dem Beispiel in Abschnitt 4.5 ist \mathbf{G} eine 6×6 -Matrix, so dass wir den eig-Befehl von Matlab verwendet haben, um $\boldsymbol{\pi}^T$ zu lösen, und dann das Ergebnis normalisiert haben (indem wir den Vektor durch seine Summe dividiert haben), um den PageRank-Vektor zu erhalten. Für eine webgroße Matrix wie die von Google ist dies jedoch nicht ausreichend. Es müssen andere, fortschrittlichere und rechenintensivere Methoden verwendet werden. Natürlich ist $\boldsymbol{\pi}^T$ der stationäre Vektor einer Markov-Kette mit der Übergangsmatrix \mathbf{G} , und es wurde viel über die Berechnung des stationären Vektors für eine allgemeine Markov-Kette geforscht. Siehe William J. Stewarts Buch

Introduction to the Numerical Solution of Markov Chains [154], das über ein Dutzend Methoden zur Ermittlung von π^T enthält. Allerdings sind die besonderen Merkmale der

PageRank-Matrix \mathbf{G} machen eine numerische Methode, die Power-Methode, zum klaren Favoriten. In diesem Abschnitt besprechen wir die Potenzmethode, die ursprünglich von Brin und Page für die Ermittlung des PageRank-Vektors vorgeschlagen wurde. Andere fortgeschrittenere Methoden werden in Kapitel 9 beschrieben.

Die größte Matrixberechnung der Welt

Cleve Moler, der Gründer von Matlab, schrieb einen Artikel [131] für seinen Newsletter *Matlab News* vom Oktober 2002, in dem er PageRank als "The World's Largest Matrix Computation" bezeichnete. Damals wendete Google die Potenzmethode auf eine dünnbesetzte Matrix der Ordnung 2,7 Milliarden an. Jetzt sind es bis zu 8,1 Milliarden!

Die Potenzmethode ist eine der ältesten und einfachsten iterativen Methoden zur Ermittlung des *dominanten Eigenwerts und Eigenvektors* einer Matrix.³ Daher kann sie verwendet werden, um den stationären Vektor einer Markov-Kette zu finden. (Der stationäre Vektor ist einfach der dominante linke Eigenvektor der Markov-Matrix.) Allerdings ist die Potenzmethode für ihre schildkrötenartige Geschwindigkeit bekannt. Von den verfügbaren iterativen Methoden (Gauß-Seidel, Jacobi, GMRES mit Neustart, BICGSTAB usw. [18]) ist die Potenzmethode im Allgemeinen die langsamste. Warum haben Brin und Page also eine Methode gewählt, die für ihre Langsamkeit bekannt ist? Es gibt mehrere gute Gründe für ihre Wahl.

Erstens ist die Leistungsmethode einfach. Die Implementierung und Programmierung sind elementar. (Siehe Kasten auf Seite 42 für eine Matlab-Implementierung der PageRank-Power-Methode.) Darüber hinaus kann die auf \mathbf{G} angewandte Potenzmethode (Gleichung (4.5.1)) tatsächlich in Form des sehr dünn besetzten \mathbf{H} ausgedrückt werden.

$$\begin{aligned}\boldsymbol{\pi}^{(k+1)T} &= \boldsymbol{\pi}^{(k)T} \mathbf{G} \\ &= \alpha \boldsymbol{\pi}^{(k)T} \mathbf{S} + \frac{1 - \alpha}{n} \boldsymbol{\pi}^{(k)T} \mathbf{e} \mathbf{e}^T \\ &= \alpha \boldsymbol{\pi}^{(k)T} \mathbf{H} + (\alpha \boldsymbol{\pi}^{(k)T} \mathbf{a} + 1 - \alpha) \mathbf{e}^T / n.\end{aligned}\quad (4.6.1)$$

Die Vektor-Matrix-Multiplikationen ($\boldsymbol{\pi}^{(k)T} \mathbf{H}$) werden mit dem extrem dünn besetzten \mathbf{H} durchgeführt, und \mathbf{S} und \mathbf{G} werden nie gebildet oder gespeichert, sondern nur ihre Rang-Eins-Komponenten, \mathbf{a} und \mathbf{e} , werden benötigt. Es sei daran erinnert, dass jede Vektor-Matrix-Multiplikation $O(n)$ ist, da \mathbf{H} etwa 10 Nullen pro Zeile hat. Dies ist wahrscheinlich der Hauptgrund dafür, dass Brin und Page im Jahr 1998 die Potenzmethode verwendeten. Aber warum ist die Power-Methode auch heute noch die vorherrschende Methode in PageRank-Forschungsarbeiten, und warum waren die meisten Verbesserungen neuartige Modifikationen der PageRank-Power-Methode und nicht Experimente mit anderen Methoden? Die weiteren Vorteile der PageRank-Power-Methode geben Antwort auf diese Fragen.

Die Potenzmethode ist, wie viele andere iterative Methoden, matrixfrei, ein Begriff, der sich auf die Speicherung und Handhabung der Koeffizientenmatrix bezieht. Bei matrixfreien Methoden wird auf die Koeffizientenmatrix nur über die Vektor-Matrix-Multiplikationsroutine zugegriffen. Es findet keine Manipulation der Matrix statt. Dies steht im Gegensatz zu direkten Methoden, bei denen Elemente der Matrix bei jedem Schritt manipuliert werden. Ändern und Speichern von Elementen der Google

³Die Potenzmethode geht mindestens auf das Jahr 1913 zurück. Mit Hilfe von James H. Wilkinson wurde die Potenzmethode in den 1960er Jahren zur Standardmethode für die Ermittlung der Eigenwerte und Eigenvektoren einer Matrix mit einem Digitalrechner [152, S. 69-70].

Matrix nicht machbar ist. Auch wenn \mathbf{H} sehr spärlich ist, schließen seine enorme Größe und sein Mangel an Struktur die Verwendung direkter Methoden aus. Stattdessen werden matrixfreie Methoden, wie die Klasse der iterativen Methoden, bevorzugt.

Die Potenzmethode ist auch speicherfreundlich. Neben der dünnbesetzten Matrix \mathbf{H} und dem Knotenvektor \mathbf{a} muss nur ein Vektor, der aktuelle Iterus $\boldsymbol{\pi}^{(k)T}$, gespeichert werden. Dieser Vektor ist vollständig dicht, d. h. es müssen n reelle Zahlen gespeichert werden. Bei Google ist $n = 8,1$ Milliarden, so dass man die sparsame Mentalität des Unternehmens verstehen kann, wenn es um die Speicherung geht. Andere iterative Methoden, wie GMRES oder BICGSTAB, sind zwar schneller, erfordern aber die Speicherung mehrerer Vektoren. Ein neu gestartetes GMRES(10) erfordert beispielsweise bei jeder Iteration die Speicherung von 10 Vektoren der Länge n , was dem Speicherbedarf der gesamten \mathbf{H} -Matrix entspricht, da $\text{nnz}(\mathbf{H}) \approx 10n$.

Der letzte Grund für die Verwendung der Power-Methode zur Berechnung des PageRank-Vektors ist die Anzahl der erforderlichen Iterationen. Brin und Page berichteten in ihren Veröffentlichungen von 1998, und andere haben dies bestätigt, dass nur 50-100 Power-Iterationen erforderlich sind, bevor die Iterate konvergieren und eine zufriedenstellende Annäherung an den exakten PageRank-Vektor liefern. Es sei daran erinnert, dass jede Iteration der Potenzmethode $O(n)$ Aufwand erfordert, weil \mathbf{H} so spärlich ist. Daher ist es schwer, eine Methode zu finden, die 50 $O(n)$ -Power-Iterationen übertreffen kann. Algorithmen, deren Laufzeit und Rechenaufwand linear (oder sublinear) mit der Problemgröße sind, sind sehr schnell und selten.

Die nächste logische Frage lautet: Warum benötigt die auf \mathbf{G} angewandte Potenzmethode nur etwa 50 Iterationen, um zu konvergieren? Gibt es etwas an der Struktur von \mathbf{G} , das auf diese schnelle Konvergenz hindeutet? Die Antwort ergibt sich aus der Theorie der Markov-Ketten. Im Allgemeinen hängt die *asymptotische Konvergenzrate* der auf eine Matrix angewandten Potenzmethode vom Verhältnis der beiden größten Eigenwerte ab, die mit λ_1 und λ_2 bezeichnet werden. Genauer gesagt ist die asymptotische Konvergenzrate die Rate, bei der $|\lambda_2 / \lambda_1|^{21^k} \rightarrow 0$. Für stochastische Matrizen wie \mathbf{G} ist $\lambda_1 = 1$, so dass $|\lambda_2|$ die Konvergenz bestimmt. Da \mathbf{G} ebenfalls primitiv ist, ist $|\lambda_2| < 1$. Im Allgemeinen erfordert die numerische Bestimmung von λ_2 für eine Matrix einen Rechenaufwand, den man nicht aufwenden möchte, nur um eine Schätzung der asymptotischen Konvergenzrate zu erhalten. Glücklicherweise ist es für das PageRank-Problem einfach zu zeigen [127, S. 502], [90, 108], dass, wenn die jeweiligen Spektren $\sigma(\mathbf{S}) = \{1, \mu_2, \dots, \mu_n\}$ und $\sigma(\mathbf{G}) = \{1, \lambda_2, \dots, \lambda_n\}$ sind, dann

$$\lambda_k = \alpha \mu_k \text{ für } k = 2, 3, \dots, n.$$

(Ein kurzer Beweis dieser Aussage findet sich am Ende dieses Kapitels.) Außerdem ist es aufgrund der Linkstruktur des Webs sehr wahrscheinlich, dass $|\mu_2| = 1$ (oder zumindest $|\mu_2| \approx 1$) ist, was bedeutet, dass $|\lambda_2(\mathbf{G})| = \alpha$ (oder $|\lambda_2(\mathbf{G})| \approx \alpha$) ist. Folglich erklärt der konvexe Kombinationsparameter α die berichtete Konvergenz nach nur 50 Iterationen. In ihren Papieren verwenden die Google-Gründer Brin und Page $\alpha = .85$, und beim letzten Bericht ist dies immer noch der von Google verwendete Wert. $\alpha^{50} = .85^{50} \approx .000296$, was bedeutet, dass man bei der 50. Iteration eine *Genauigkeit* von etwa 2-3 Stellen im ungefähren PageRank-Vektor erwarten kann. Dieser Genauigkeitsgrad ist für Googles Ranking-Bedürfnisse offenbar ausreichend. Mathematisch gesehen könnten zehn Stellen Genauigkeit erforderlich sein, um zwischen den Elementen des PageRank-Vektors zu unterscheiden (siehe Abschnitt 8.3), aber wenn PageRank-Bewertungen mit Inhaltsbewertungen kombiniert werden, ist

eine hohe Genauigkeit möglicherweise weniger wichtig.

Subdominanter Eigenwert der Google-Matrix

Für die Google-Matrix $\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{1}/n \mathbf{e} \mathbf{e}^T$,

$$|\lambda_2(\mathbf{G})| \leq \alpha.$$

- Für den Fall, dass $|\lambda_2(\mathbf{S})| = 1$ ist (was aufgrund der Reduzierbarkeit des Webgraphen häufig vorkommt), ist $|\lambda_2(\mathbf{G})| = \alpha$. Daher ist die asymptotische Konvergenzrate der PageRank-Power-Methode von Gleichung (4.6.1) die Rate, bei der $\alpha^k \rightarrow 0$.

Wir können nun positive Antworten auf die sechs Fragen aus Abschnitt 4.3 geben. Mit den Anpassungen der Stochastizität und der Primitivität konvergiert die auf \mathbf{G} angewandte Potenzmethode garantiert zu einem einzigen positiven Vektor, dem PageRank-Vektor, unabhängig vom Startvektor. Da der resultierende PageRank-Vektor positiv ist, gibt es keine unerwünschten Gleichstände bei 0. Um PageRank-Werte mit einer Genauigkeit von ungefähr τ Stellen zu erzeugen, müssen außerdem $-\tau / \log_{10} \alpha$ Iterationen müssen abgeschlossen werden.

Matlab m-file für PageRank Power Methode

Diese m-Datei ist eine Matlab-Implementierung der PageRank-Power-Methode aus Gleichung (4.6.1).

```
Funktion [pi,time,numiter]=\hbox{PageRank}(pi0,H,n,alpha,epsilon);

% \hbox{PageRank} berechnet den \hbox{PageRank}-Vektor für einen n-mal-n
Markov
    %Matrix H mit Startvektor pi0 (ein Zeilenvektor)
    %und Skalierungsparameter alpha (skalar). Verwendet Leistung
    % Methode.
    %
    % BEISPIEL: [pi,time,numiter]=\hbox{PageRank}(pi0,H,1000,.9,1e-8);
    %
    % INPUT: pi0 = Startvektor bei Iteration 0 (ein Zeilenvektor)
    %H = zeilennormierte Hyperlink-Matrix (n-mal-n dünnbesetzte
Matrix)
    %n = Größe der H-Matrix (skalar)
    %alpha = Skalierungsparameter im \hbox{PageRank}-Modell (skalar)
    %epsilon = Konvergenztoleranz (skalar, z. B. 1e-8)
    %
    % OUTPUT: pi = \hbox{PageRank} Vektor
    %time = für die Berechnung des Vektors \hbox{PageRank}
erforderliche Zeit
    %numiter = Anzahl der Iterationen bis zur Konvergenz
    %
    % Der Startvektor wird in der Regel auf den einheitlichen Vektor gesetzt,
    % pi0=1/n*ones(1,n).
    % HINWEIS: Matlab speichert dünnbesetzte Matrizen spaltenweise und ist daher
schneller
```

zum einige Operationen mit H' , der Transponierten von H , durchzuführen.

```
% "a", den "Dangling Node"-Vektor, erhalten, wobei a(i)=1, wenn Knoten i
%ist baumelnder Knoten und 0, o.w.

rowsumvector=ones(1,n)*H';
nonzerorows=find(rowsumvector);
zerorows=setdiff(1:n,nonzerorows);
l=length(zerorows);
a=sparse(zerorows,ones(l,1),ones(l,1),n,1);

k=0;
residual=1;
pi=pi0;
tic;

while (residual >= epsilon)
    prevpi=pi;
    k=k+1;
    pi=alpha*pi*H + (alpha*(pi*a)+1-
    alpha)*((1/n)*ones(1,n)); residual=norm(pi-prevpi,1);
end
numiter=k;
time=toc;
```

Suche innerhalb einer Site

Im hart umkämpften Geschäft der Suche ist Google manchmal erfrischend großzügig. Zum Beispiel lässt Google Website-Autoren kostenlos seine Technologie für die Suche auf ihrer Website nutzen. (Ein Klick auf die Schaltfläche "Mehr" auf der Google-Startseite führt Sie zu den neuesten Informationen über die Google-Dienste). Bei Suchanfragen innerhalb einer Website schränkt Google die Menge der relevanten Seiten auf die Seiten innerhalb der Website ein. Diese standortinternen relevanten Seiten werden dann anhand der globalen PageRank-Werte eingestuft. Im Wesentlichen wird bei dieser standortinternen Suche die Website aus dem riesigen Google-Index mit Milliarden von Seiten extrahiert und der Teil des Webs herausgefiltert, der zu der Website gehört. Die Betrachtung eines einzelnen Subwebs macht den Hyperlink-Graphen viel übersichtlicher.

ASIDE: Suchmaschinen-Optimierung

Da sich immer mehr Verkäufe ins Internet verlagern, wenden sich große und kleine Unternehmen gleichermaßen an Suchmaschinenoptimierer (SEOs), um ihre Gewinne zu steigern. SEOs stellen Webseiten und Links sorgfältig zusammen, um die Chancen zu "optimieren", dass die Seiten ihrer Kunden auf den ersten Seiten der Suchmaschinenergebnisse erscheinen. SEOs können als ethisch oder unethisch eingestuft werden. Ethische SEOs sind gute Netizens, Bürgerinnen und Bürger des Netzes, die nur gute Ratschläge geben, z. B. wie man Text am besten anzeigt und Bilder und Tags kennzeichnet. Sie ermutigen die Autoren von Webseiten, gute Inhalte zu pflegen, da sich die Rangfolge der Seiten aus der Bewertung des Inhalts und der Beliebtheit zusammensetzt. Sie warnen die Autoren auch davor, dass Suchmaschinen Seiten bestrafen, die sie als vorsätzliches

Spamming wahrnehmen. Ethische SEOs und Suchmaschinen betrachten sich als Partner, die durch den Austausch von Informationen und Tipps gemeinsam die Suchqualität verbessern. Unethische SEOs hingegen versuchen absichtlich, Suchmaschinen zu überlisten und Spamming-Techniken zu fördern. In der Randbemerkung auf Seite 52 finden Sie einen konkreten Fall von unethischen SEO-Praktiken. Seit den Anfängen des Internets haben Suchmaschinen

in einen ewigen Kampf mit unethischen SEOs verwickelt. Der Kampf tobt überall im Web, vom sichtbaren Webseiteninhalt bis zu versteckten Metatags, von Links bis zu Ankertexten und von internen Servern bis zu Linkfarmen (siehe auch die Anmerkung auf Seite 52).

SEOs hatten bei den ersten Suchmaschinen Erfolg, indem sie Term-Spamming und Verstecktechniken einsetzten [84]. Beim Term-Spamming werden Spam-Wörter in den Text der Seite eingefügt, oft wiederholt, im Titel, in den Metatags, im Ankertext und im URL-Text. Versteckte Techniken verwenden Farbschemata und Tarnungen, um Suchmaschinen zu täuschen. Die Verwendung von weißem Text auf weißem Hintergrund beispielsweise macht Spam für menschliche Leser unsichtbar, was bedeutet, dass Suchmaschinen weniger wahrscheinlich hilfreiche Beschwerden über Seiten mit verstecktem Spam erhalten. Cloaking bezieht sich auf die Technik, eine mit Spam beladene Webseite für normale Benutzeranfragen und eine andere, spamfreie Seite für Anfragen von Suchmaschinen-Crawlern zu liefern. Solange die Autoren Web-Crawling-Agenten eindeutig identifizieren können, können sie den Agenten mit einer sauberen, spamfreien Seite wegschicken. Da diese Techniken für Webseitenautoren so einfach zu handhaben sind, mussten die Suchmaschinen zurückschlagen. Sie taten dies, indem sie den IQ ihrer Spider und Indexer erhöhten. Viele Spider und Indexer sind darauf trainiert, Metatags zu ignorieren, da diese in den späten 1990er Jahren kaum noch genaue Seiteninformationen enthielten. Sie ignorieren auch wiederholte Schlüsselwörter. Das Cloaking ist jedoch schwieriger zu bekämpfen. Suchmaschinen bitten die Nutzer um Hilfe, um Cloaking zu unterbinden. So bittet Google die Surfer, als Schiedsrichter zu fungieren und zu melden, wenn sie eine verdächtige Seite finden, die sie sofort auf eine neue Seite umleitet.

1998 nahmen die Suchmaschinen die Linkanalyse in ihre Trickkiste auf. Infolgedessen konnten Content-Spam und Cloaking allein die Linkanalyse-Maschinen nicht mehr täuschen und den Spammern ungerechtfertigt hohe Platzierungen verschaffen. Spammer und SEOs passten sich an, indem sie lernten, wie die Linkanalyse funktioniert. Die SEO-Gemeinschaft war schon immer aktiv - ihre Mitglieder, damals wie heute, halten Konferenzen ab, schreiben Abhandlungen und Bücher, betreiben Weblogs und verkaufen ihre Geheimnisse. Die bekanntesten und informativsten SEO-Beiträge stammen von Chris Ridings, "PageRank explained: Everything you've always wanted to know about PageRank" [143] und "PageRank uncovered" [144]. Diese Abhandlungen bieten praktische Strategien, um PageRank zu horten und unerwünschte Dinge wie PageRank-Leaks zu vermeiden. Suchmaschinen passen ihre Algorithmen ständig an, um den SEO-Spielern einen Schritt voraus zu sein. Während Suchmaschinen unethische SEOs als Werbung betrachten, bezeichnen einige Webanalysten sie als wesentlichen Teil der Nahrungskette im Web, da sie Innovation, Forschung und Entwicklung vorantreiben.

ASIDE: Wie verdienen Suchmaschinen Geld?

Diese Frage wird uns oft gestellt. Es ist eine gute Frage. Suchmaschinen bieten kostenlosen und unbegrenzten Zugang zu ihren Diensten, woher kommen also die Milliarden von Dollar an Sucheinnahmen? Suchmaschinen haben mehrere Einnahmequellen. Da ist zum einen die Aufnahmegebühr, die einige Suchmaschinen von Website-Autoren verlangen. Einige ungeduldige Autoren wollen eine Garantie, dass ihre neue Website bald (in ein oder zwei Tagen) indexiert wird und nicht erst in ein oder zwei Monaten, wenn ein Spider sie endlich in der Liste der zu durchsuchenden URLs findet. Suchmaschinen bieten diese Garantie gegen eine geringe Gebühr an, und gegen eine etwas höhere Gebühr können Autoren garantieren, dass ihre Website in kürzeren Abständen, vielleicht monatlich, neu indexiert wird.

Die meisten Suchmaschinen erzielen auch Einnahmen durch den Verkauf von

Profildaten an Interessenten. Suchmaschinen sammeln täglich enorme Mengen von Nutzerdaten. Diese Daten werden verwendet, um die Qualität der Suche zu verbessern und die Bedürfnisse der Nutzer vorherzusagen, aber sie werden auch in aggregierter Form an verschiedene Unternehmen verkauft. So können beispielsweise Suchmaschinenoptimierungsunternehmen, die an beliebten Suchbegriffen oder dem Prozentsatz der Suchanfragen mit kommerziellem Charakter interessiert sind, diese Informationen direkt von einer Suchmaschine kaufen.

Suchmaschinen verkaufen zwar nicht den Zugang zu ihren Suchfunktionen an einzelne Nutzer, aber sie verkaufen Suchdienste an Unternehmen. Netscape beispielsweise bezahlt Google dafür, dass die Google-Suche als Standardsuche in seinem Browser verwendet wird. Früher verkaufte GoTo (das von Overture aufgekauft wurde, das jetzt zu Yahoo gehört) seine sieben besten Ergebnisse für jeden Suchbegriff an Yahoo und AltaVista, die ihrerseits diese sieben Ergebnisse als ihre besten Ergebnisse verwendeten.

Trotz dieser Einnahmequellen ist die bei weitem profitabelste und am schnellsten wachsende Einnahmequelle für Suchmaschinen die Werbung. Es wird geschätzt, dass im Jahr 2004 3 Milliarden Dollar an Sucheinnahmen durch Werbung generiert werden. Der Börsengang von Google am 21. Juni 2004 machte die Abhängigkeit des Unternehmens von der Werbung deutlich: Über 97 % der Einnahmen des Unternehmens im Jahr 2003 wurden durch Werbung erzielt. Viele Suchmaschinen verkaufen Werbefbanner, die auf ihren Homepages und Ergebnisseiten erscheinen. Andere verkaufen Pay-for-Placement-Anzeigen. Diese umstrittenen Anzeigen ermöglichen es einem Unternehmen, sich an die Spitze des Rankings zu kaufen. Viele Webanalysten argumentieren, dass diese kostenpflichtigen Anzeigen die Suchergebnisse verschmutzen. Suchmaschinen, die diese Technik anwenden (GoTo ist ein Paradebeispiel), entgegnen jedoch, dass diese Art der Platzierung hervorragend für kommerzielle Suchen geeignet ist. Da jüngsten Erhebungen zufolge 15-30 % aller Suchanfragen kommerzieller Natur sind, bieten Suchmaschinen wie Overture einen wertvollen Dienst für diese Art von Suchanfragen. Andererseits sind viele Suchanfragen forschungsorientiert, und die Ergebnisse der kostenpflichtigen Suchmaschinen frustrieren diese Nutzer.

Google verfolgt einen anderen Ansatz bei Anzeigen und Rankings. Es zeigt die unbezahlten Ergebnisse in einer Hauptliste an, während die bezahlten Websites als "gesponserte Links" separat an der Seite erscheinen. Google und jetzt auch Yahoo sind die einzigen verbleibenden Unternehmen, die bezahlte Links nicht mit reinen Links vermischen. Google verwendet ein Cost-per-Click-Werbesystem, um gesponserte Links zu präsentieren. Die Unternehmen wählen ein Schlüsselwort, das mit ihrem Produkt oder ihrer Dienstleistung in Verbindung steht, und bieten dann auf einen Preis, den sie bereit sind zu zahlen, wenn ein Suchender auf ihren Link klickt. Ein Fahrradgeschäft in Raleigh kann zum Beispiel 5 Cent für jede Suchanfrage nach "bike Raleigh" bieten. Das Fahrradgeschäft wird nur dann in Rechnung gestellt, wenn ein Suchender tatsächlich auf seine Anzeige klickt. Ein anderes Unternehmen kann jedoch 17 Cent für dieselbe Suchanfrage bieten. Die Anzeige des zweiten Unternehmens wird wahrscheinlich zuerst erscheinen, da gesponserte Anzeigen trotz einiger Feinabstimmungen und Optimierungen in der Regel in der Reihenfolge des höchsten Gebots bis zum niedrigsten Gebot aufgelistet werden.

Cost-per-Click-Werbung ist eine Innovation im Marketing. Kleine Unternehmen, die traditionell wenig für Werbung ausgaben, geben jetzt viel mehr für Internetwerbung aus, weil Cost-per-Click-Werbung so kosteneffektiv ist. Wenn ein Suchender auf den Link klickt, zeigt er oder sie eine Kaufabsicht an, etwas, das andere Werbemittel wie Plakatwände oder Postwurfsendungen nicht leisten können. Interessanterweise war es, wie bei vielen anderen Dingen im Web, nur eine Frage der Zeit, bis die Cost-per-Click-Werbung zu einem Schlachtfeld zwischen den Wettbewerbern wurde. Ohne Schutz (der in Form eines Softwareprogramms erworben werden kann) können naive Unternehmen, die Cost-per-Click-Werbung kaufen, leicht von Konkurrenten sabotiert werden. Die Kunden klicken immer wieder auf die Anzeigen des unbedarften Unternehmens, was die Rechnung in die

Höhe treibt und das Werbebudget des Unternehmens aufbraucht.

4.7 THEOREM UND BEWEIS FÜR DAS SPEKTRUM DER GOOGLE-MATRIX

In diesem Kapitel haben wir die Google-Matrix als $\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{I}/n \mathbf{e} \mathbf{e}^T$ definiert. Im Abschnitt 5.3 des nächsten Kapitels erweitern wir diese Definition jedoch auf

eine allgemeinere Google-Matrix, bei der die Fudge-Faktor-Matrix \mathbf{E} von der einheitlichen Matrix $1/n \mathbf{e}\mathbf{e}^T$ zu $\mathbf{e}\mathbf{v}^T$ wechselt, wobei $\mathbf{v}^T > \mathbf{0}$ ein Wahrscheinlichkeitsvektor ist. In diesem Abschnitt stellen wir das Theorem und den Beweis für den zweiten Eigenwert dieser allgemeineren Google-Matrix vor.

Theorem 4.7.1. Wenn das Spektrum der stochastischen Matrix $\mathbf{S} \{1, \lambda_2, \lambda_3, \dots, \lambda_n\}$ ist, dann ist das Spektrum der Google-Matrix $\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{e} \mathbf{v}^T \{1, \alpha \lambda_2, \alpha \lambda_3, \dots, \alpha \lambda_n\}$, wobei \mathbf{v}^T ein Wahrscheinlichkeitsvektor ist.

Proof. Since \mathbf{S} is stochastic, $(1, \mathbf{e})$ is an eigenpair of \mathbf{S} . Let $\mathbf{Q} = (\mathbf{e} \ \mathbf{X})$ be a non-singular matrix that has the eigenvector \mathbf{e} as its first column. Let $\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{y}^T \\ \mathbf{Y}^T \end{pmatrix}$. Then

$\mathbf{Q}^{-1} \mathbf{Q} = \begin{pmatrix} \mathbf{y}^T \mathbf{e} & \mathbf{y}^T \mathbf{X} \\ \mathbf{Y}^T \mathbf{e} & \mathbf{Y}^T \mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$, woraus sich zwei nützliche Identitäten ergeben: $\mathbf{y}^T \mathbf{e} = 1$ und

$\mathbf{Y}^T \mathbf{e} = \mathbf{0}$. Daraus folgt, dass die Ähnlichkeitstransformation

$$\mathbf{Q}^{-1} \mathbf{S} \mathbf{Q} = \begin{pmatrix} \mathbf{y}^T \mathbf{e} & \mathbf{y}^T \mathbf{S} \mathbf{X} \\ \mathbf{Y}^T \mathbf{e} & \mathbf{Y}^T \mathbf{S} \mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{y}^T \mathbf{S} \mathbf{X} \\ \mathbf{0} & \mathbf{Y}^T \mathbf{S} \mathbf{X} \end{pmatrix}$$

zeigt, dass $\mathbf{Y}^T \mathbf{S} \mathbf{X}$ die restlichen Eigenwerte von \mathbf{S} , $\lambda_2, \dots, \lambda_n$ enthält. Die Anwendung der Ähnlichkeitstransformation auf $\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{e} \mathbf{v}^T$ ergibt

$$\begin{aligned} \mathbf{Q}^{-1}(\alpha \mathbf{S} + (1 - \alpha) \mathbf{e} \mathbf{v}^T) \mathbf{Q} &= \alpha \mathbf{Q}^{-1} \mathbf{S} \mathbf{Q} + (1 - \alpha) \mathbf{Q}^{-1} \mathbf{e} \mathbf{v}^T \mathbf{Q} \\ &= \begin{pmatrix} \alpha & \alpha \mathbf{y}^T \mathbf{S} \mathbf{X} \\ 0 & \alpha \mathbf{Y}^T \mathbf{S} \mathbf{X} \end{pmatrix} + (1 - \alpha) \begin{pmatrix} \mathbf{y}^T \mathbf{e} & \mathbf{y}^T \mathbf{S} \mathbf{X} \\ \mathbf{Y}^T \mathbf{e} & \mathbf{Y}^T \mathbf{S} \mathbf{X} \end{pmatrix} (\mathbf{v}^T \mathbf{e} \ \mathbf{v}^T \mathbf{X}) \\ &= \begin{pmatrix} \alpha & \alpha \mathbf{y}^T \mathbf{S} \mathbf{X} & (1 - \alpha) & (1 - \alpha) \mathbf{y}^T \mathbf{X} \\ 0 & \alpha \mathbf{Y}^T \mathbf{S} \mathbf{X} & \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \alpha \mathbf{y}^T \mathbf{S} \mathbf{X} + (1 - \alpha) \mathbf{v}^T \mathbf{X} \\ \mathbf{0} & \alpha \mathbf{Y}^T \mathbf{S} \mathbf{X} \end{pmatrix}. \end{aligned}$$

Daher sind die Eigenwerte von $\mathbf{G} = \alpha \mathbf{S} + (1 - \alpha) \mathbf{e} \mathbf{v}^T \{1, \alpha \lambda_2, \alpha \lambda_3, \dots, \alpha \lambda_n\}$. ■

Fünfzehntes Kapitel

Der Mathematik-Leitfaden

Um die Feinheiten von PageRank, HITS und anderen Rankingsystemen zu verstehen, sind Kenntnisse einiger mathematischer Konzepte erforderlich. Insbesondere ist es notwendig, einige Aspekte der linearen Algebra, der diskreten Markov-Ketten und der Graphentheorie zu verstehen. Anstatt einen umfassenden Überblick über diese Gebiete zu geben, wollen wir hier nur die wichtigsten Themen ansprechen, die bei der mathematischen Analyse von Websuchkonzepten auftreten. Technische Beweise werden in der Regel weggelassen.

Die gemeinsame Basis ist die lineare Algebra, und damit beginnen wir. Der Leser, der mehr Details wünscht oder einfach die elementare lineare Algebra in einem größeren Umfang als dem hier gegebenen wiederholen möchte, sollte [127] konsultieren.

15.1 LINEARE ALGEBRA

Im Kontext der Websuche sind die Matrizen fast immer reell, aber da reelle Matrizen komplexe Zahlen erzeugen können (z. B. Eigenwerte), ist es oft notwendig, komplexe Zahlen, Vektoren und Matrizen zu betrachten. In diesem Kapitel werden reelle Zahlen, reelle Vektoren und reelle Matrizen mit \mathbb{R} , \mathbb{R}^n und $\mathbb{R}^{m \times n}$ bezeichnet, während komplexe Zahlen, Vektoren und Matrizen mit \mathbb{C} , \mathbb{C}^n und $\mathbb{C}^{m \times n}$ bezeichnet werden. Die folgenden grundlegenden Konzepte tauchen bei der mathematischen Analyse von Websuchproblemen auf.

Normen

Die gebräuchlichste Methode zur Messung der Größe eines Zeilen- (oder Spalten-) Vektors $\mathbf{x} = (x_1, x_2, \dots, x_n)$ reeller oder komplexer Zahlen ist die euklidische Norm (manchmal auch 2-Norm genannt), die wie folgt definiert ist

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}.$$

In den Anwendungen, die PageRank und Markov-Ketten beinhalten, ist es jedoch natürlicher (und bequemer), die Vektor-1-Norm zu verwenden, die durch

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$$

denn wenn zum Beispiel \mathbf{p} ein PageRank-Vektor (oder Wahrscheinlichkeitsvektor) ist (d. h. ein nicht negativer Vektor mit Komponenten, die sich zu eins summieren), dann $\|\mathbf{p}\|_1 = 1$. Gelegentlich kann der Vektor ∞ -norm

$$\|\mathbf{x}\|_{\infty} = \max_i |x_i|$$



verwendet wird. Alle Normen erfüllen die folgenden drei Eigenschaften

$$\begin{aligned} \|\mathbf{x}\| &\geq 0 \text{ wobei } \|\mathbf{x}\| = 0 \text{ wenn und nur wenn } \mathbf{x} \\ &= \mathbf{0}, \|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\| \quad \text{für alle Skalare } \alpha, \\ \|\mathbf{x} + \mathbf{y}\| &\leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad (\text{die Dreiecksungleichung}). \end{aligned}$$

Jeder Vektornorm ist eine *induzierte Matrixnorm* zugeordnet. Wenn $\mathbf{A} \in \mathbb{R}^{m \times n}$ und $\|\cdot\|_\lambda$ eine beliebige Vektornorm ist, dann ist die entsprechende induzierte Matrixnorm wie folgt definiert

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

Die jeweiligen Matrixnormen, die durch die 1- 2- und ∞ -Vektornormen induziert werden, sind

$$\|\mathbf{A}\|_1 = \max_j \sum_i |a_{ij}| = \text{die größte absolute Spaltensumme,}$$

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}}, \quad \text{wobei } \lambda_{\max} = \text{größter Eigenwert von } \mathbf{A}^T \mathbf{A},$$

(ersetzen Sie die Transposition durch die konjugierte

$$\text{Transposition, wenn } \mathbf{A} \text{ komplex ist), } \|\mathbf{A}\|_\infty = \max_j \sum_i |a_{ij}| = \text{die größte absolute Zeilensumme.}$$

Die Einzelheiten zu diesen Eigenschaften sind in [127] zu finden.

Das Schöne an induzierten Matrixnormen ist, dass jede von ihnen mit der entsprechenden Vektornorm in dem Sinne *kompatibel* ist, dass

$$\|\mathbf{A}\mathbf{x}\|_\lambda \leq \|\mathbf{A}\|_\lambda \|\mathbf{x}\|_\lambda.$$

Diese Kompatibilitätsbedingung gilt jedoch nur für die rechtshändige Matrix-Vektor-Multiplikation. Bei der linksseitigen Vektor-Matrix-Multiplikation, die in Markov-Ketten-Anwendungen häufig vorkommt, ist eine Transposition erforderlich, um zur rechtsseitigen Matrix-Vektor-Multiplikation zurückzukehren, was zu anderen Kompatibilitätsregeln führt. Wenn \mathbf{x}^T gleich $1 \times n$ und \mathbf{A} gleich $m \times n$ ist, dann

$$\|\mathbf{x}^T \mathbf{A}\|_1 \leq \|\mathbf{x}^T\|_1 \|\mathbf{A}\|_\infty, \quad \|\mathbf{x}^T \mathbf{A}\|_\infty \leq \|\mathbf{x}^T\|_\infty \|\mathbf{A}\|_1.$$

Empfindlichkeit von linearen Systemen

Es wird vorausgesetzt, dass der Leser mit den Gaußschen Eliminationsmethoden zur Lösung eines Systems $\mathbf{A} \mathbf{x}_{m \times n \times 1} = \mathbf{b}_{m \times 1}$ von m linearen Gleichungen in n Unbekannten vertraut ist. Falls nicht, lesen Sie [127]. Algorithmen zur Lösung von $\mathbf{A}\mathbf{x} = \mathbf{b}$ sind wichtig, aber das allgemeine Verhalten einer Lösung bei kleinen Unsicherheiten oder Störungen in den Koeffizienten ist besonders relevant, vor allem in Anbetracht der Tatsache, dass der PageRank-Vektor die Lösung eines bestimmten linearen Systems ist.

Obwohl eine größere Allgemeinheit möglich ist, reicht es aus, ein quadratisches, nicht-singuläres System $\mathbf{A}\mathbf{x} = \mathbf{b}$ zu betrachten, bei dem sowohl \mathbf{A} als auch \mathbf{b} Unsicherheiten unterliegen, die das Ergebnis von Modellierungsfehlern, numerischen

Rundungsfehlern, Messfehlern oder kleinen Störungen jeglicher Art sein können. Wie viel Unsicherheit (oder Empfindlichkeit) kann die Lösung $\mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$ aufweisen?

Eine Antwort wird mit Hilfe der Infinitesimalrechnung gegeben. Betrachten Sie die

Einträge von $\mathbf{A} = \mathbf{A}(t)$ und

$\mathbf{b} = \mathbf{b}(t)$ mit einem differenzierbaren Parameter t variieren, und berechnen Sie die relative Größe der

Ableitung von $\mathbf{x} = \mathbf{x}(t)$ durch Differenzieren von $\mathbf{b} = \mathbf{A}\mathbf{x}$, um $\mathbf{b}' = (\mathbf{A}\mathbf{x})' = \mathbf{A}'\mathbf{x} + \mathbf{A}\mathbf{x}'$ zu erhalten (wobei für d/dt steht). Durch Normierung (die Wahl der Norm ist nicht wichtig) erhält man

$$\|\mathbf{x}'\| = \|\mathbf{A}\mathbf{b}^{-1'} - \mathbf{A}\mathbf{A}^{-1'}\mathbf{x}\| \leq \|\mathbf{A}\mathbf{b}^{-1'}\| + \|\mathbf{A}\mathbf{A}^{-1'}\mathbf{x}\|$$

$$\leq \|\mathbf{A}^{-1}\| \|\mathbf{b}'\| + \|\mathbf{A}^{-1}\| \|\mathbf{A}'\| \|\mathbf{x}\|$$

Folglich,

$$\frac{\|\mathbf{x}'\|}{\|\mathbf{x}\|} \leq \frac{\|\mathbf{A}^{-1}\| \|\mathbf{b}'\|}{\|\mathbf{x}\|} + \|\mathbf{A}^{-1}\| \|\mathbf{A}'\|$$

$$\leq \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \frac{\|\mathbf{b}'\|}{\|\mathbf{A}\| \|\mathbf{x}\|} + \|\mathbf{A}\| \|\mathbf{A}'\| \frac{\|\mathbf{x}\|}{\|\mathbf{A}\| \|\mathbf{x}\|}$$

ÜBERNEHMEN.

$$\leq \kappa \frac{\|\mathbf{b}'\|}{\|\mathbf{b}\|} + \kappa \frac{\|\mathbf{A}'\|}{\|\mathbf{A}\|} = \kappa \left(\frac{\|\mathbf{b}'\|}{\|\mathbf{b}\|} + \frac{\|\mathbf{A}'\|}{\|\mathbf{A}\|} \right),$$

wobei $\kappa = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$. Die Terme $\|\mathbf{x}\| / \|\mathbf{x}\|$, $\|\mathbf{b}\| / \|\mathbf{b}\|$ und $\|\mathbf{A}\| / \|\mathbf{A}\|$ stellen die jeweiligen relativen Empfindlichkeiten von \mathbf{x} , \mathbf{b} und \mathbf{A} gegenüber kleinen Änderungen dar. Da κ eine Vergrößerung der Summe der relativen Empfindlichkeiten von \mathbf{b} und \mathbf{A} darstellt, wird κ als *Zustandszahl* für \mathbf{A} bezeichnet.

Empfindlichkeit von linearen Systemen

Für ein nichtsinguläres System $\mathbf{A}\mathbf{x} = \mathbf{b}$ ist die relative Empfindlichkeit von \mathbf{x} gegenüber Unsicherheiten oder Störungen in \mathbf{A} und \mathbf{b} nie größer als die Summe der relativen Änderungen in \mathbf{A} und \mathbf{b} , verstärkt durch die Bedingungszahl $\kappa = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$.

Eine praktische Faustformel. Wenn die Gaußsche Eliminierung mit partieller Pivotisierung verwendet wird, um ein gut skaliertes (Zeilennormen in \mathbf{A} sind annähernd eins) nichtsinguläres System $\mathbf{A}\mathbf{x} = \mathbf{b}$ unter Verwendung von t -stelliger Fließkommaarithmetik zu lösen, und wenn κ von der Ordnung 10^p ist, dann kann man unter der Annahme, dass keine andere Fehlerquelle existiert, erwarten, dass die berechnete Lösung auf mindestens $t - p$ signifikante Stellen genau ist, mehr oder weniger. Mit anderen Worten, man erwartet einen Verlust von etwa p signifikanten Stellen. Das schließt nicht aus, dass man mit etwas Glück einen höheren Genauigkeitsgrad erreichen kann - es sagt nur, dass man darauf nicht wetten sollte.

Rang-Eins-Updates

Angenommen, $\mathbf{A} \in \mathbb{R}^{n \times n}$ ist die Koeffizientenmatrix eines nichtsingulären Systems $\mathbf{A}\mathbf{x} = \mathbf{b}$, das Informationen enthält, die periodisch aktualisiert werden müssen, und jedes Mal, wenn neue Informationen eingehen, muss das System neu gelöst werden. Anstatt jedes Mal bei Null anzufangen, ist es sinnvoll, die Lösung aus der vorangegangenen Periode auf einfache, aber vorhersehbare Weise zu stören. Theoretisch ist die Lösung immer $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, so dass das Problem der Aktualisierung der Lösung eines linearen Systems gleichbedeutend ist mit dem Problem der Aktualisierung der inversen Matrix \mathbf{A}^{-1} . Wenn die neue Information als eine Rang-Eins-Matrix $\mathbf{c}\mathbf{d}^T$ formatiert werden kann, wobei $\mathbf{c}, \mathbf{d} \in \mathbb{R}^{n \times 1}$, dann gibt es eine Formel für die Aktualisierung von \mathbf{A}^{-1} .

Sherman-Morrison-Formel zur Aktualisierung des ersten Ranges

Wenn $\mathbf{A}_{n \times n}$ nichtsingulär ist und wenn \mathbf{c} und \mathbf{d} solche Spalten sind, dass $1 + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{c} \neq 0$,
dann ist die Summe $\mathbf{A} + \mathbf{c}\mathbf{d}^T$ nichtsingulär, und

$$(\mathbf{A} + \mathbf{c}\mathbf{d}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{c} \mathbf{d}^T \mathbf{A}^{-1}}{1 + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{c}}. \quad (15.1.1)$$

Die Sherman-Morrison-Formel macht deutlich, dass, wenn ein nichtsinguläres System $\mathbf{A}\mathbf{x} = \mathbf{b}$ aktualisiert wird, um ein anderes nichtsinguläres System $(\mathbf{A} + \mathbf{c}\mathbf{d}^T)\mathbf{z} = \mathbf{b}$ zu erzeugen, wobei $\mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^{n \times 1}$, die Lösung des aktualisierten Systems ist

$$\begin{aligned} \mathbf{z} &= (\mathbf{A} + \mathbf{c}\mathbf{d}^T)^{-1} \mathbf{b} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{c} \mathbf{d}^T \mathbf{A}^{-1}}{1 + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{c}} \mathbf{b} \\ &= \mathbf{A}^{-1} \mathbf{b} - \frac{\mathbf{A}^{-1} \mathbf{c} \mathbf{d}^T \mathbf{A}^{-1} \mathbf{b}}{1 + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{c}} = \mathbf{x} - \frac{\mathbf{A}^{-1} \mathbf{c} \mathbf{d}^T \mathbf{x}}{1 + \mathbf{d}^T \mathbf{A}^{-1} \mathbf{c}} \end{aligned}$$

Die Sherman-Morrison-Formel ist besonders nützlich, wenn eine Aktualisierung nur eine Zeile oder Spalte von \mathbf{A} betrifft. Nehmen wir zum Beispiel an, dass nur die Zeile i^{th} von \mathbf{A} betroffen ist - sagen wir, die Zeile \mathbf{A}_{i*} wird zu \mathbf{B}_{i*} aktualisiert, und $\mathbf{s}^T = \mathbf{B}_{i*} - \mathbf{A}_{i*}$. Wenn \mathbf{e}_i die Einheitsspalte i^{th} bezeichnet (die Spalte i^{th} der Identitätsmatrix \mathbf{I}), dann kann die aktualisierte Matrix wie folgt geschrieben werden

$$\mathbf{B} = \mathbf{A} + \mathbf{e}_i \mathbf{s}_i^T,$$

so dass \mathbf{e}_i die Rolle von \mathbf{c} in (15.1.1) spielt, und $\mathbf{A}^{-1} \mathbf{c} = \mathbf{A}^{-1} \mathbf{e}_i = [\mathbf{A}^{-1}]_{*i}$, die i^{th} Spalte von \mathbf{A}^{-1} . Folglich kann \mathbf{B}^{-1} direkt aus den Einträgen in \mathbf{A}^{-1} und dem Störungsvektor \mathbf{s}^T konstruiert werden, indem man schreibt.

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} + \mathbf{e}_i \mathbf{s}_i^T \mathbf{A}^{-1} = \mathbf{A}^{-1} + \frac{[\mathbf{A}^{-1}]_{*i} \mathbf{s}_i^T \mathbf{A}^{-1}}{1 + \mathbf{s}_i^T [\mathbf{A}^{-1}]_{*i}}.$$

Eigenwerte und Eigenvektoren

Für eine Matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ sind die Skalare λ und die Vektoren $\mathbf{x}_{n \times 1} \neq \mathbf{0}$, die $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ erfüllen, die jeweiligen *Eigenwerte* und *Eigenvektoren* von \mathbf{A} . Ein Zeilenvektor \mathbf{y}^T ist ein *linker Eigenvektor*, wenn $\mathbf{y}^T \mathbf{A} = \lambda \mathbf{y}^T$.

Die Menge $\sigma(\mathbf{A})$ der *verschiedenen* Eigenwerte wird als *Spektrum* von \mathbf{A} bezeichnet, und der *Spektralradius* von \mathbf{A} ist die nichtnegative Zahl

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|.$$

Der Kreis in der komplexen Ebene, der im Ursprung zentriert ist und den Radius $\rho(\mathbf{A})$ hat, wird als *Spektralkreis* bezeichnet, und es ist eine einfache Übung, zu überprüfen, dass

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\| \quad (15.1.2)$$

für alle Matrixnormen.

Die Eigenwerte von $\mathbf{A}_{n \times n}$ sind die Wurzeln des charakteristischen Polynoms $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$, wobei $\det(\cdot)$ die Determinante bezeichnet. Der Grad von $p(\lambda)$ ist n , also alles in allem,

\mathbf{A} hat n Eigenwerte, aber einige können komplexe Zahlen sein (auch wenn die Einträge von \mathbf{A} reelle Zahlen sind), und einige Eigenwerte können sich wiederholen. Wenn \mathbf{A} nur reelle Zahlen enthält, dann müssen seine komplexen Eigenwerte in konjugierten Paaren auftreten, d. h. wenn $\lambda \in \sigma(\mathbf{A})$, dann $\bar{\lambda} \in \sigma(\mathbf{A})$.

Die *algebraische Mannigfaltigkeit* eines Eigenwerts λ von \mathbf{A} ist die Anzahl der Wiederholungen von λ als Wurzel der charakteristischen Gleichung. Wenn $\text{alg mult}_{\mathbf{A}}(\lambda) = 1$ ist, dann ist λ ein *einfacher Eigenwert*.

Die *geometrische Multiplizität* eines Eigenwerts λ von \mathbf{A} ist die Anzahl der linear unabhängigen Eigenvektoren, die mit λ assoziiert sind. Formaler ausgedrückt: $\text{geo mult}_{\mathbf{A}}(\lambda) = \dim N(\mathbf{A} - \lambda \mathbf{I})$, wobei $N(\cdot)$ den Nullraum oder Kern einer Matrix bezeichnet. Es ist immer der Fall, dass $\text{geo mult}_{\mathbf{A}}(\lambda) \leq \text{alg mult}_{\mathbf{A}}(\lambda)$ ist. Wenn $\text{geo mult}_{\mathbf{A}}(\lambda) = \text{alg mult}_{\mathbf{A}}(\lambda)$ ist, dann heißt es, dass λ ein *semisimpler Eigenwert* ist.

The *index of an eigenvalue* $\lambda \in \sigma(\mathbf{A})$ is defined to be the smallest positive integer k such that $\text{rank}(\mathbf{A} - \lambda \mathbf{I})^k = \text{rank}(\mathbf{A} - \lambda \mathbf{I})^{k+1}$. It is understood that $\text{index}(\lambda) = 0$ when $\lambda \notin \sigma(\mathbf{A})$.

There are several different ways to characterize index. For $\lambda \in \sigma(\mathbf{A}_{n \times n})$, saying that $k = \text{index}(\lambda)$ is equivalent to saying that k is the smallest positive integer such that any of the following statements hold.

- $R(\mathbf{A} - \lambda \mathbf{I})^k = R(\mathbf{A} - \lambda \mathbf{I})^{k+1}$, wobei $R(\cdot)$ den Bereich bezeichnet.
- $N(\mathbf{A} - \lambda \mathbf{I})^k = N(\mathbf{A} - \lambda \mathbf{I})^{k+1}$, wobei $N(\cdot)$ den Nullraum (oder Kern) bezeichnet.
- $R(\mathbf{A} - \lambda \mathbf{I})^k \cap N(\mathbf{A} - \lambda \mathbf{I})^k = \mathbf{0}$.
- $\mathbb{C}^n = R(\mathbf{A} - \lambda \mathbf{I})^k \oplus N(\mathbf{A} - \lambda \mathbf{I})^k$, wobei \oplus die direkte Summe bezeichnet.

Die Jordanienform

Eigenwerte und Eigenvektoren sind für Matrizen das, was die DNA für biologische Einheiten ist, und die Jordan-Form für eine quadratische Matrix \mathbf{A} charakterisiert vollständig die Eigenstruktur von \mathbf{A} . Die theoretische Grundlage dafür, warum die Jordan-Form so aussieht, wie sie aussieht, ist etwas kompliziert, aber die "Form" selbst ist leicht zu verstehen, und das ist alles, was Sie brauchen, um mit den Problemen umzugehen, die beim Verständnis von Websuchkonzepten auftreten.

Bei einer Matrix $\mathbf{A}_{n \times n}$ ist ein *Jordan-Block*, der mit einem Eigenwert $\lambda \in \sigma(\mathbf{A})$ assoziiert ist, definiert als eine Matrix der Form

$$\mathbf{J}_{\lambda}(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \quad (15.1.3)$$

Ein *Jordan-Segment* $\mathbf{J}(\lambda)$, das mit $\lambda \in \sigma(\mathbf{A})$ assoziiert ist, ist definiert als eine Block-Diagonal-Matrix, die einen oder mehrere Jordan-Blöcke enthält. Mit anderen Worten, ein Jordan-Segment sieht wie folgt aus

$$\mathbf{J}(\lambda) = \begin{pmatrix} \mathbf{J}_1(\lambda) & \mathbf{0} & \cdots & \mathbf{0} \\ & \mathbf{J}_2(\lambda) & \cdots & \mathbf{0} \\ & & \ddots & \\ & & & \mathbf{J}_k(\lambda) \end{pmatrix} \quad \text{wobei jeder } \mathbf{J}_i(\lambda) \text{ ein Jordan-Block ist.}$$

$$\mathbf{0} \quad \mathbf{0} \quad \text{---} \mathbf{J_t}(\lambda)$$

Die *kanonische Jordan-Form* (oder einfach die *Jordan-Form*) für \mathbf{A} ist eine blockdiagonale Matrix, die aus den Jordan-Segmenten für jeden einzelnen Eigenwert besteht. Mit anderen Worten: Wenn $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$ dann ist die Jordan-Form für \mathbf{A}

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}(\lambda_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}(\lambda_2) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{J}(\lambda_s) \end{pmatrix}. \quad (15.1.4)$$

Es gibt nur ein Jordan-Segment für jeden Eigenwert, aber jedes Segment kann mehrere Jordan-Blöcke von unterschiedlicher Größe enthalten. Die Formel, die die Größe und Anzahl der Jordan-Blöcke bestimmt, ist in der folgenden vollständigen Aussage zur Jordan-Form enthalten.

Jordans Theorem

Für jedes $\mathbf{A} \in \mathbb{C}^{n \times n}$ gibt es eine nichtsinguläre Matrix \mathbf{P} , für die gilt

(15.1.5)

$\mathbf{P}^{-1} \mathbf{A} \mathbf{P} = \mathbf{J}$ ist die Jordanform (15.1.4), die durch folgende Merkmale gekennzeichnet ist.

- \mathbf{J} enthält ein Jordan-Segment $\mathbf{J}(\lambda)$ für jeden eindeutigen Eigenwert $\lambda \in \sigma(\mathbf{A})$.
- Jedes Segment $\mathbf{J}(\lambda)$ enthält $t = \dim N(\mathbf{A} - \lambda \mathbf{I})$ Jordanblöcke.
- Die Anzahl der $i \times i$ Jordan-Blöcke in $\mathbf{J}(\lambda)$ ist gegeben durch $v_i(\lambda) = r_{i-1}(\lambda) - 2r_i(\lambda) + r_{i+1}(\lambda)$, wobei $r_i(\lambda) = \text{Rang}(\mathbf{A} - \lambda \mathbf{I})^i$.
- Der größte Jordan-Block in jedem Segment $\mathbf{J}(\lambda)$ ist $k \times k$, wobei $k = \text{Index}(\lambda)$.

Die Struktur von \mathbf{J} ist einzigartig in dem Sinne, dass die Anzahl und Größe der Jordan-Blöcke in jedem Segment eindeutig durch die Einträge in \mathbf{A} bestimmt wird. Zwei $n \times n$ -Matrizen \mathbf{A} und \mathbf{B} sind *ähnlich* (d. h. $\mathbf{B} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$ für ein nichtsinguläres \mathbf{Q}), wenn und nur wenn \mathbf{A} und \mathbf{B} die gleiche Jordan-Form haben.

The matrix \mathbf{P} in (15.1.5) is not unique, but its columns always form *Jordan chains* (or *generalized eigenvectors*) in the following sense. For each Jordan block $\mathbf{J}_k(\lambda)$, there is a set of columns \mathbf{P}_λ of corresponding size and position in $\mathbf{P} = \dots | \mathbf{P}_\lambda | \dots$ such that

$$\mathbf{P}_\lambda = (\mathbf{A} - \lambda \mathbf{I})^i \mathbf{x}_\lambda \quad (\mathbf{A} - \lambda \mathbf{I})^{i-1} \mathbf{x}_\lambda \dots (\mathbf{A} - \lambda \mathbf{I}) \mathbf{x}_\lambda \quad \mathbf{x}_\lambda \quad (i+1) \times n$$

für ein bestimmtes i und ein bestimmtes \mathbf{x}_λ , wobei $(\mathbf{A} - \lambda \mathbf{I})^i \mathbf{x}_\lambda$ ein bestimmter, mit λ verbundener Eigenvektor ist. Es gibt Formeln zur Bestimmung von i und \mathbf{x}_λ [127, S. 594], aber die Berechnungen können kompliziert sein. Glücklicherweise müssen wir \mathbf{P} nur selten berechnen.

Eine wichtige Folge des Satzes von Jordan (15.1.5) ist die folgende Aussage über die Diagonalisierbarkeit einer quadratischen Matrix.

Diagonalisierbarkeit

Jede der folgenden Aussagen ist gleichbedeutend mit der Aussage, dass $A \in \mathbb{C}^{n \times n}$ einer Diagonalmatrix ähnlich ist - d. h., J ist diagonal (alle Jordanblöcke sind 1×1).

- $\text{Index}(\lambda) = 1$ für jedes $\lambda \in \sigma(A)$ (d.h. jeder Eigenwert ist semisimpel).
- $\text{alg mult}_A(\lambda) = \text{geo mult}_A(\lambda)$ für jedes $\lambda \in \sigma(A)$.
- A hat eine vollständige Menge von n linear unabhängigen Eigenvektoren (d. h. jede Spalte von P ist ein Eigenvektor für A).

Funktionen einer Matrix

Eine wichtige Anwendung der Jordan-Form ist die Definition von Funktionen von $A \in \mathbb{C}^{n \times n}$. Das heißt, wenn eine Funktion $f: \mathbb{C} \rightarrow \mathbb{C}$ gegeben ist, was soll $f(A)$ bedeuten? Die Antwort ist leicht zu finden. Angenommen,

dass $A = PJP^{-1}$, wobei $J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_s \end{pmatrix}$ hat die Jordanform, wobei die J_λ 's für

die in (15.1.3) beschriebenen Jordanblöcke. Es ist natürlich, den Wert von f bei A zu definieren als

$$f(A) = Pf(J)P^{-1} = P \begin{pmatrix} f(J_1) & & 0 \\ & \ddots & \\ 0 & & f(J_s) \end{pmatrix} P^{-1}, \quad (15.1.6)$$

aber der Trick besteht darin, $f(J_\lambda)$ richtig zu definieren. Es stellt sich heraus, dass der richtige Weg, dies zu tun, darin besteht, Folgendes festzulegen

$$f(J_\lambda) = f \left(\begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \right) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(k-1)}(\lambda)}{(k-1)!} \\ 0 & f(\lambda) & f'(\lambda) & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \frac{f''(\lambda)}{2!} \\ 0 & \vdots & 0 & f(\lambda) & f'(\lambda) \\ 0 & 0 & 0 & 0 & f(\lambda) \end{pmatrix}. \quad (15.1.7)$$

Matrix-Funktionen

Sei $A \in \mathbb{C}^{n \times n}$ mit $\sigma(A) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$ und $f: \mathbb{C} \rightarrow \mathbb{C}$ sei so, dass $f(\lambda_i), f'(\lambda_i), \dots, f^{(k_i-1)}(\lambda_i)$ existieren für jedes i , wobei $k_i = \text{Index}(\lambda_i)$. Definieren Sie

$$f(A) = Pf(J)P^{-1} = P \begin{pmatrix} f(J_1) & & 0 \\ & \ddots & \\ 0 & & f(J_s) \end{pmatrix} P^{-1}, \quad (15.1.8)$$

wobei J die Jordanform für A ist und $f(J_\lambda)$ durch (15.1.7) gegeben ist.

Es gibt mindestens zwei weitere gleichwertige und nützliche Möglichkeiten, Funktionen von Matrizen zu betrachten. Die erste davon wird als *Spektralsatz für Matrixfunktionen* bezeichnet und ergibt sich, wenn man das Produkt auf der rechten Seite von (15.1.8) erweitert, um Folgendes zu erhalten.

Spektraltheorem für allgemeine Matrizen

Wenn $\mathbf{A} \in \mathbb{C}^{n \times n}$ mit $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$, dann

$$f(\mathbf{A}) = \sum_{i=1}^s \sum_{j=0}^{k_i-1} \frac{f^{(j)}(\lambda_i)}{j!} (\mathbf{A} - \lambda_i \mathbf{I})^j \mathbf{G}_i \quad (15.1.9)$$

wobei jedes \mathbf{G}_i die folgenden Eigenschaften

• \mathbf{G}_i ist ein Projektor (d. h. $\mathbf{G}_i^2 = \mathbf{G}_i$) auf $N(\mathbf{A} - \lambda_i \mathbf{I})^{k_i}$ entlang $R(\mathbf{A} - \lambda_i \mathbf{I})^{k_i}$.

• $\mathbf{G}_1 + \mathbf{G}_2 + \dots + \mathbf{G}_s = \mathbf{I}$.

• $\mathbf{G}_i \mathbf{G}_j = \mathbf{0}$ wenn $i \neq j$. $(\mathbf{A} - \lambda_i \mathbf{I}) \mathbf{G}_i$ ist nilpotent vom Index k_i .

Die \mathbf{G}_i werden als *Spektralprojektoren* bezeichnet, die der Matrix \mathbf{A} zugeordnet sind.

Eine weitere nützliche Methode, mit Funktionen einer Matrix umzugehen, sind unendliche Reihen.

Unendliche Reihendarstellungen

Wird $\sum_{j=0}^{\infty} c_j (z - z_0)^j$ in jedem Punkt eines Kreises $|z - z_0| = r$ gegen $f(z)$ gleichmäßig konvergiert und wenn $\sum_{j=0}^{\infty} c_j (\mathbf{A} - z_0 \mathbf{I})^j$ konvergiert, dann

$$f(\mathbf{A}) = \sum_{j=0}^{\infty} c_j (\mathbf{A} - z_0 \mathbf{I})^j$$

Wenn \mathbf{A} diagonalisierbar ist, d. h. wenn es einer Diagonalmatrix ähnelt

$$\mathbf{A} = \mathbf{P} \begin{pmatrix} \lambda_1 \mathbf{I}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \lambda_2 \mathbf{I}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \lambda_s \mathbf{I}_s \end{pmatrix} \mathbf{P}^{-1},$$

dann

$$f(\mathbf{A}) = \mathbf{P} \begin{pmatrix} f(\lambda_1) \mathbf{I}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & f(\lambda_2) \mathbf{I}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & f(\lambda_s) \mathbf{I}_s \end{pmatrix} \mathbf{P}^{-1},$$

und der Formel (15.1.9) ergibt sich der folgende Spektralsatz für diagonalisierbare Matrizen

Spektraltheorem für diagonalisierbare Matrizen

Wenn \mathbf{A} diagonalisierbar ist mit $\sigma(\mathbf{A}) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$, dann

$$\mathbf{A} = \lambda_1 \mathbf{G}_1 + \lambda_2 \mathbf{G}_2 + \dots + \lambda_s \mathbf{G}_s, \quad (15.1.10)$$

und

$$f(\mathbf{A}) = f(\lambda_1) \mathbf{G}_1 + f(\lambda_2) \mathbf{G}_2 + \dots + f(\lambda_s) \mathbf{G}_s, \quad (15.1.11)$$

wobei die Spektralprojektoren \mathbf{G}_i die folgenden Eigenschaften haben.

- $\mathbf{G}_1 + \mathbf{G}_2 + \dots + \mathbf{G}_s = \mathbf{I}$

- $\mathbf{G}_i \mathbf{G}_j = \mathbf{0}$ wenn $i \neq j$

- $\mathbf{G}_i = \prod_{j \neq i} \frac{(\mathbf{A} - \lambda_j \mathbf{I})}{(\lambda_i - \lambda_j)}$ für $i = 1, 2, \dots, s$.

- Wenn λ_i zufällig ein einfacher Eigenwert ist, dann

$$\mathbf{G}_i = \mathbf{xy}^* / \mathbf{y}^* \mathbf{x}^* \quad (15.1.12)$$

wobei \mathbf{x} und \mathbf{y}^* die jeweiligen rechten und linken Eigenvektoren von λ_i sind.

Potenzen von Matrizen und Konvergenz

Ein grundlegendes Problem bei der Analyse des PageRank betrifft die Konvergenz von Potenzen von Matrizen. Aus (15.1.8) folgt, dass jede Potenz von $\mathbf{A} \in \mathbb{C}^{n \times n}$ gegeben ist durch

$$\mathbf{A}^k = \mathbf{P} \mathbf{J} \mathbf{P}^{k-1} = \mathbf{P} \mathbf{J}_s^k \mathbf{P}^{-1}, \quad \text{wobei } \mathbf{J}_s = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_s \end{pmatrix}$$

und

$$\mathbf{J}_s^k = \begin{pmatrix} \lambda_1^k & \lambda_1^{k-1} & \dots & \lambda_1^{k-m+1} \\ & \lambda_1^k & & \\ & & \ddots & \\ & & & \lambda_s^k & \lambda_s^{k-1} & \dots & \lambda_s^{k-m+1} \end{pmatrix} \quad (15.1.13)$$

Diese Beobachtung führt zu den folgenden einschränkenden Eigenschaften.

Konvergenz gegen Null und die Neumann-Reihe

Für $A \in \mathbb{C}^{n \times n}$ sind die folgenden Aussagen gleichwertig.

$$\bullet \quad \rho(A) < 1. \quad (15.1.14)$$

$$\bullet \quad \lim_{k \rightarrow \infty} A^k = 0. \quad \sum_{k=0}^{\infty} A^k \text{ konvergiert zu } (I - A)^{-1}. \quad (15.1.15)$$

$$\bullet \quad \text{Die Neumann-Reihe} \quad (15.1.16)$$

Es kann sein, dass die Potenzen A^k konvergieren, aber nicht gegen die Nullmatrix. Die vollständige Geschichte bezüglich $\lim_{k \rightarrow \infty} A^k$ ist wie folgt.

Grenzen der Befugnisse

Für $A \in \mathbb{C}^{n \times n}$ existiert $\lim_{k \rightarrow \infty} A^k$ nur dann, wenn $\rho(A) < 1$ ist. In diesem Fall ist $\lim_{k \rightarrow \infty} A^k = 0$, andernfalls ist $\rho(A) = 1$, wobei $\lambda = 1$ semisimpel ist und der einzige Eigenwert auf dem Einheitskreis. Wenn er existiert,

$$\lim_{k \rightarrow \infty} A^k = G = \text{der Projektor auf } N(I - A) \text{ entlang } R(I - A). \quad (15.1.17)$$

Durchschnittswerte und Summierbarkeit

Zu jeder skalaren Folge $\{a_1, a_2, a_3, \dots\}$ gibt es eine zugehörige Folge von Durchschnittswerten

$\{\mu_1, \mu_2, \mu_3, \dots\}$, wobei

$$\mu_1 = a_1, \quad \mu_2 = \frac{a_1 + a_2}{2}, \quad \dots, \quad \mu_n = \frac{a_1 + a_2 + \dots + a_n}{n}.$$

Diese Folge von Mittelwerten wird *Cesàro-Folge* genannt, und wenn $\lim_{n \rightarrow \infty} \mu_n = \alpha$ ist, sagen wir, dass $\{a_n\}$ *Cesàro-summierbar* (oder einfach *summierbar*) zu α ist. Es kann bewiesen werden, dass, wenn $\{a_n\}$ zu α konvergiert, dann $\{\mu_n\}$ zu α konvergiert, aber nicht umgekehrt. Mit anderen Worten: Konvergenz impliziert Summierbarkeit, aber Summierbarkeit ist keine Garantie für Konvergenz. Um zu sehen, dass eine Folge summierbar sein kann, ohne konvergent zu sein, ist zu beachten, dass die oszillierende Folge $\{0, 1, 0, 1, \dots\}$ nicht konvergiert, aber sie ist summierbar zu $1/2$, dem Mittelwert von $\{0, 1\}$. Die Mittelwertbildung hat eine glättende Wirkung, so dass Schwingungen, die die Konvergenz der ursprünglichen Sequenz verhindern, in der Cesàro-Sequenz geglättet oder gemittelt werden.

Ähnliche Aussagen gelten für Folgen von Vektoren und Matrizen, aber Cesàro summability ist besonders interessant, wenn sie auf die Folge $P = \{A^k\}_{k=0}^{\infty}$ von Potenzen angewendet wird

einer quadratischen Matrix A .

Summierbarkeit

$\mathbf{A} \in \mathbb{C}^{n \times n}$ ist nur dann Cesàro-summierbar, wenn $\rho(\mathbf{A}) < 1$ oder $\rho(\mathbf{A}) = 1$ ist und jeder Eigenwert auf dem Einheitskreis semisimpel ist. Wenn sie existiert, ist der Grenzwert

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{A} + \dots + \mathbf{A}^{k-1}}{k} = \mathbf{G} \quad (15.1.18)$$

ist der Projektor auf $N(\mathbf{I} - \mathbf{A})$ entlang $R(\mathbf{I} - \mathbf{A})$.

Man beachte, dass $\mathbf{G} \neq \mathbf{0}$ nur dann ist, wenn $1 \in \sigma(\mathbf{A})$, in diesem Fall ist \mathbf{G} der mit $\lambda = 1$ verbundene Spektralprojektor. Wenn $\lim_{k \rightarrow \infty} \mathbf{A}^k = \mathbf{G}$ ist, dann ist \mathbf{A} zu \mathbf{G} summierbar, aber nicht umgekehrt.

Die Power-Methode

Googles ursprüngliche Methode zur Berechnung des PageRank-Vektors war die *Power-Methode*, eine iterative Technik zur Berechnung eines dominanten Eigenpaares (λ_1, \mathbf{x}) einer diagonalisierbaren Matrix $\mathbf{A} \in \mathbb{C}^{m \times m}$ mit Eigenwerten

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_k|.$$

Für die Google-Matrix ist der dominante Eigenwert $\lambda_1 = 1$, aber da die Analyse der Potenzmethode nicht von dieser Tatsache abhängt, lassen wir λ_1 allgemeiner sein. Man beachte jedoch, dass die Hypothese $|\lambda_1| > |\lambda_2|$ impliziert, dass λ_1 reell ist – andernfalls ist λ_1 (die komplex konjugierte) ein weiterer Eigenwert mit der gleichen Größe wie λ_1 . Betrachten Sie die Funktion $f(z) = (z/\lambda_1)^n$, und verwenden Sie die Spektraldarstellung (15.1.11) zusammen mit $|\lambda_i/\lambda_1| < 1$ für $i = 2, 3, \dots, k$, um zu schließen, dass

$$\begin{aligned} \frac{\mathbf{A}^n}{\lambda_1^n} &= f(\mathbf{A}) = f(\lambda_1) \mathbf{G}_1 + f(\lambda_2) \mathbf{G}_2 + \dots + f(\lambda_k) \mathbf{G}_k \\ &= \mathbf{G}_1 + \frac{\lambda_2^n}{\lambda_1^n} \mathbf{G}_2 + \dots + \frac{\lambda_k^n}{\lambda_1^n} \mathbf{G}_k \rightarrow \mathbf{G}_1 \text{ mit } n \rightarrow \infty. \end{aligned} \quad (15.1.19)$$

Für jedes \mathbf{x}_0 gilt $(\mathbf{A} \mathbf{x} / \lambda_1^n) \rightarrow \mathbf{G} \mathbf{x}_0 \in N(\mathbf{A} - \lambda_1 \mathbf{I})$, d. h. wenn $\mathbf{G} \mathbf{x}_0 \neq \mathbf{0}$, dann konvergiert $\mathbf{A} \mathbf{x} / \lambda_1^n$ zu einem mit λ_1 assoziierten Eigenvektor. Das bedeutet, dass die Richtung von $\mathbf{A} \mathbf{x}_0$ in Richtung eines Eigenvektors tendiert, da λ^n nur als Skalierungsfaktor wirkt, um die Länge von $\mathbf{A} \mathbf{x}_0$ unter Kontrolle zu halten. Anstatt λ^n zu verwenden, können wir die Skalierung von $\mathbf{A} \mathbf{x}_0$ mit einem geeigneteren Faktor vornehmen. Zum Beispiel ist $\|\mathbf{A} \mathbf{x}_0\|$ (für jede Vektornorm) ein vernünftiger Skalierungsfaktor, aber es gibt bessere Möglichkeiten. Für Vektoren \mathbf{v} soll $m(\mathbf{v})$ die Komponente mit dem maximalen Betrag bezeichnen, und wenn es mehr als eine maximale Komponente gibt, soll $m(\mathbf{v})$ die erste maximale Komponente sein – z. B. $m(1, 3, -2) = 3$ und $m(-3, 3, -2) = -3$. Die Potenzmethode lässt sich wie folgt zusammenfassen.

Leistung Methode

Beginnen Sie mit einer beliebigen Schätzung \mathbf{x}_0 . (Eigentlich kann sie nicht völlig willkürlich sein, denn Sie brauchen $\mathbf{x}_0 \in R(\mathbf{A} - \lambda_1 \mathbf{I})$, um $\mathbf{G} \mathbf{x}_{10} \neq \mathbf{0}$ zu gewährleisten, aber es ist höchst unwahrscheinlich, dass ein zufällig gewählter Vektor \mathbf{x}_0 $\mathbf{G} \mathbf{x}_{10} = \mathbf{0}$ erfüllt.) Es kann gezeigt werden [127, S. 534], dass, wenn

$$\mathbf{w}_n^T \mathbf{A} \mathbf{x}_n, \quad \mathbf{v}_n^T = m(\mathbf{y}_n), \quad \mathbf{x}_{n+1} = \frac{\mathbf{y}_n}{\mathbf{v}_n}, \quad \text{für } n = 0, 1, 2, \dots, \quad (15.1.20)$$

dann $\mathbf{x}_n \rightarrow \mathbf{x}$ und $\mathbf{v}_n \rightarrow \lambda_1$, wobei $\mathbf{A} \mathbf{x} = \lambda_1 \mathbf{x}$.

Es gibt mehrere Gründe, warum die Power-Methode für die Berechnung des PageRank-Vektors von Google attraktiv sein könnte.

- Für jede Iteration ist nur ein Matrix-Vektor-Produkt erforderlich. Dies kann ausgenutzt werden, um den Rechenaufwand zu verringern, wenn \mathbf{A} groß und spärlich ist (überwiegend Nullen), wie es bei der Google-Anwendung der Fall ist.
- Die Berechnungen können parallel durchgeführt werden, indem gleichzeitig die inneren Produkte der Zeilen von \mathbf{A} mit \mathbf{x}_n berechnet werden.
- Aus (15.1.19) geht hervor, dass für eine diagonalisierbare Matrix die Konvergenzrate von (15.1.20) davon abhängt, wie schnell $(\lambda / \lambda_{21})^n \rightarrow 0$ ist. Wie in Abschnitt 4.7 besprochen, kann Google $|\lambda_2|$ durch die Wahl des Google-Parameters α regulieren und somit die Konvergenzrate steuern (es wird einfach angenommen, dass die Google-Matrix diagonalisierbar ist).
- Da $\lambda_1 = 1$ für das PageRank-Problem von Google ist, ist der Skalierungsfaktor nicht erforderlich
 \mathbf{v}_n . Mit anderen Worten, die Iterationen sind einfach $\mathbf{x}_{n+1} = \mathbf{A} \mathbf{x}_n$.

Lineare stationäre Iterationen

Das Lösen von linearen Gleichungssystemen $\mathbf{A}_{n \times n} \mathbf{x} = \mathbf{b}$ ist eine häufige Notwendigkeit für Web-Suchanwendungen, aber die Größe von n ist in der Regel zu groß, als dass direkte Lösungsmethoden auf der Grundlage der Gauß'schen Eliminierung effektiv wären. Folglich sind iterative Techniken oft die einzige Wahl, und aufgrund von Größen-, Sparsamkeits- und Speichererwägungen sind die bevorzugten Algorithmen die einfacheren Methoden, die auf Matrix-Vektor-Produkten basieren und keinen zusätzlichen Speicherplatz über den der Originaldaten hinaus benötigen. Lineare stationäre Iterationsverfahren sind am weitesten verbreitet.

Lineare stationäre Iterationen

$\mathbf{Ax} = \mathbf{b}$ sei ein lineares System, das quadratisch, aber ansonsten beliebig ist. Schreibt man \mathbf{A} als $\mathbf{A} = \mathbf{M} - \mathbf{N}$, wobei \mathbf{M}^{-1} existiert, so nennt man dies eine *Aufspaltung* von \mathbf{A} , und das Produkt $\mathbf{H} = \mathbf{M}^{-1} \mathbf{N}$ heißt die zugehörige *Iterationsmatrix*. Für $\mathbf{d} = \mathbf{M}^{-1} \mathbf{b}$ und für einen Anfangsvektor $\mathbf{x}(0)$ ist die Folge definiert durch

$$\mathbf{x}(k) = \mathbf{H}\mathbf{x}(k-1) + \mathbf{d} \quad k = 1, 2, 3, \dots \quad (15.1.21)$$

wird eine *lineare stationäre Iteration* genannt. Das wichtigste Ergebnis für die Konvergenz von (15.1.21) ist die Tatsache, dass, wenn $\rho(\mathbf{H}) < 1$ ist, \mathbf{A} nichtsingulär ist, und

$$\lim_{k \rightarrow \infty} \mathbf{x}(k) = \mathbf{x} = \mathbf{A}^{-1} \mathbf{b} \text{ (die Lösung von } \mathbf{Ax} = \mathbf{b} \text{) für jedes } \mathbf{x}(0). \quad (15.1.22)$$

Theoretisch wird die Konvergenzrate von (15.1.21) durch die Größe von $\rho(\mathbf{H})$ zusammen mit dem Index des zugehörigen Eigenwerts bestimmt - siehe (15.1.13). Für die praktische Arbeit wird jedoch eine Angabe darüber benötigt, wie viele Stellen der Genauigkeit pro Iteration erwartet werden können. Nehmen wir an, dass $\mathbf{H}_{n \times n}$ diagonalisierbar ist mit

$$\sigma(\mathbf{H}) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}, \text{ wobei } 1 > |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_s|$$

(was in Anwendungen häufig der Fall ist), und lassen Sie $\mathbf{s}(k) = \mathbf{x}(k) - \mathbf{x}$ den Fehler nach der k^{th} Iteration bezeichnen. Subtrahiert man $\mathbf{x} = \mathbf{H}\mathbf{x} + \mathbf{d}$ (den Grenzwert in (15.1.21)) von $\mathbf{x}(k) = \mathbf{H}\mathbf{x}(k-1) + \mathbf{d}$, so erhält man (für große k)

$$\mathbf{s}(k) = \mathbf{H}\mathbf{s}(k-1) = \mathbf{H}^k \mathbf{s}(0) = (\lambda_1 G_1^k + \lambda_2 G_2^k + \dots + \lambda_s G_s^k) \mathbf{s}(0) \approx \lambda_1 G_1^k \mathbf{s}(0),$$

wobei die \mathbf{G}_i die Spektralprojektoren sind, die in der spektralen Zerlegung (15.1.11) von \mathbf{H}^k vorkommen. In ähnlicher Weise ist $\mathbf{s}(k-1) \approx \lambda_1 G_1^{k-1} \mathbf{s}(0)$, so dass der Vergleich der i^{th} Komponenten von $\mathbf{s}(k-1)$ und $\mathbf{s}(k)$ nach mehreren Iterationen zeigt, dass,

$$\frac{s_i(k-1)}{s_i(k)} \approx \frac{1}{|\lambda_1|} = \frac{1}{\rho(\mathbf{H})} \quad \text{für jedes } i = 1, 2, \dots, n.$$

Um zu verstehen, was das bedeutet, nehmen wir zum Beispiel an, dass

$$|s_i(k-1)| = 10^{-q} \quad \text{und} \quad |s_i(k)| = 10^{-p} \quad \text{mit} \quad p \geq q > 0,$$

so dass der Fehler in jedem Eintrag um $p - q$ Ziffern pro Iteration reduziert wird, und wir haben

$$p - q = \log \frac{1}{10} \frac{s_i(k-1)}{s_i(k)} \approx -\log_{10} \rho(\mathbf{H}).$$

Nachstehend finden
Sie eine
Zusammenfassung.

Asymptotische Konvergenzrate

Die Zahl $R = -\log_{10} \rho(\mathbf{H})$, die als *asymptotische Konvergenzrate* für (15.1.21) bezeichnet wird, wird zum Vergleich verschiedener linearer stationärer Iterationsalgorithmen verwendet, da sie die Anzahl der Stellen der Genauigkeit angibt, die bei jeder Iteration von $\mathbf{x}(k) = \mathbf{H}\mathbf{x}(k-1) + \mathbf{d}$ zu erwarten ist.

Jede unterschiedliche Aufteilung $\mathbf{A} = \mathbf{M} - \mathbf{N}$ führt zu einem anderen Iterationsalgorithmus, aber es gibt drei besondere Aufteilungen, die weit verbreitet sind.

Die drei klassischen Splittings

- **Die Jacobi-Methode** ist das Ergebnis der Aufteilung von $\mathbf{A} = \mathbf{D} - \mathbf{N}$, wobei \mathbf{D} der diagonale Teil von \mathbf{A} ist (unter der Annahme, dass jedes $a_{ii} \neq 0$ ist) und $(-\mathbf{N})$ die Matrix ist, die die nicht diagonalen Einträge von \mathbf{A} enthält. Die Jacobi-Iteration ist $\mathbf{x}(k) = \mathbf{D}^{-1} \mathbf{N} \mathbf{x}(k-1) + \mathbf{D}^{-1} \mathbf{b}$.
- **Die Gauß-Seidel-Methode** ergibt sich aus der Aufteilung von $\mathbf{A} = (\mathbf{D} - \mathbf{L}) - \mathbf{U}$, wobei \mathbf{D} der Diagonaleil von \mathbf{A} ist (unter der Annahme, dass jedes $a_{ii} \neq 0$) und $(-\mathbf{L})$ und $(-\mathbf{U})$ die Einträge unterhalb bzw. oberhalb der Diagonalen von \mathbf{A} enthalten. Die Iterationsmatrix ist $\mathbf{H} = (\mathbf{D} - \mathbf{L})^{-1} \mathbf{U}$, und $\mathbf{d} = (\mathbf{D} - \mathbf{L})^{-1} \mathbf{b}$. Die Gauss-Seidel-Iteration ist $\mathbf{x}(k) = (\mathbf{D} - \mathbf{L})^{-1} \mathbf{U} \mathbf{x}(k-1) + (\mathbf{D} - \mathbf{L})^{-1} \mathbf{b}$.
- **Bei der Methode der sukzessiven Überrelaxation (SOR)** wird ein *Relaxationsparameter* $\omega \neq 0$ in das Gauß-Seidel-Verfahren integriert, um eine Aufspaltung $\mathbf{A} = \mathbf{M} - \mathbf{N}$ zu bilden, wobei $\mathbf{M} = \omega^{-1} \mathbf{D} - \mathbf{L}$ und $\mathbf{N} = (\omega^{-1} - 1) \mathbf{D} + \mathbf{U}$.

Es kann gezeigt werden, dass sowohl die Jacobi-Methode als auch die Gauß-Seidel-Methode konvergieren, wenn \mathbf{A} *diagonal dominant* ist (d.h. wenn $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ für jedes $i = 1, 2, \dots, n$.) Dies und weitere Details zur Konvergenz können in [127] nachgelesen werden.

M-Matrizen

Da der PageRank-Vektor als Lösung einer Markov-Kette betrachtet werden kann und $\mathbf{I} - \mathbf{P}$ eine M-Matrix ist, wenn \mathbf{P} eine Wahrscheinlichkeitsübergangsmatrix ist, ist es praktisch, ein paar Fakten über M-Matrizen (benannt zu Ehren von Hermann Minkowski) zu kennen.

M-Matrizen

Eine quadratische (reelle) Matrix \mathbf{A} wird als M-Matrix bezeichnet, wenn es eine Matrix gibt

$\mathbf{B} \geq \mathbf{0}$ (d. h. $b_{ij} \geq 0$) und eine reelle Zahl $r \geq \rho(\mathbf{B})$, so dass $\mathbf{A} = r\mathbf{I} - \mathbf{B}$.

Wenn $r > \rho(\mathbf{B})$ in der obigen Definition, dann ist \mathbf{A} eine *nichtsinguläre* M-Matrix. Im Folgenden werden einige wichtige Eigenschaften nichtsingulärer M-Matrizen beschrieben.

- \mathbf{A} ist eine nichtsinguläre M-Matrix, wenn und nur wenn $a_{ij} \leq 0$ für alle $i \neq j$ und $\mathbf{A}^{-1} \geq \mathbf{0}$.
- Wenn \mathbf{A} eine nichtsinguläre M-Matrix ist, dann ist $\operatorname{Re}(\lambda) > 0$ für alle $\lambda \in \sigma(\mathbf{A})$. Umgekehrt sind alle Matrizen mit nichtpositiven Off-Diagonal-Einträgen, deren Spektren in der rechten Halbebene liegen, nichtsinguläre M-Matrizen.

- Hauptuntermatrizen von nichtsingulären M-Matrizen sind ebenfalls nichtsinguläre M-Matrizen.
- Wenn A eine M-Matrix ist, dann sind alle ihre Hauptminoren nicht negativ. Wenn A eine nichtsinguläre M-Matrix ist, dann sind alle Hauptminoren positiv.

- Alle Matrizen mit nichtpositiven Off-Diagonal-Einträgen, deren Hauptminoren nicht negativ sind, sind M-Matrizen. Alle Matrizen mit nicht-positiven Off-Diagonal-Einträgen, deren Hauptminoren positiv sind, sind nicht-singuläre M-Matrizen.
- Wenn $A = M - N$ eine Aufspaltung einer nichtsingulären M-Matrix ist, für die $M^{-1} \geq 0$ ist, dann konvergiert die lineare stationäre Iteration (15.1.21) für alle Anfangsvektoren $x(0)$ und für alle rechten Seiten b . Insbesondere konvergiert das Jacobi-Verfahren.

15.2 PERRON-FROBENIUS-THEORIE

Auf einer Mathematikkonferenz, die vor einigen Jahren stattfand, hielt unser Freund Hans Schneider einen denkwürdigen Vortrag mit dem Titel "Warum ich Perron-Frobenius liebe", in dem er darlegte, dass die Perron-Frobenius-Theorie der nichtnegativen Matrizen nicht nur zu den elegantesten Theorien der Mathematik gehört, sondern auch zu den nützlichsten. Man könnte Hans' Standpunkt mit den Worten zusammenfassen, dass Perron-Frobenius ein Beweis dafür ist, dass schöne Mathematik letztendlich dazu neigt, nützlich zu sein, und nützliche Mathematik letztendlich dazu neigt, schön zu sein. Die Anwendungen von PageRank, HITS und anderen Rankingsystemen [103] helfen, dieses Prinzip zu verstehen.

Eine Matrix A gilt als *nichtnegativ*, wenn jeder Eintrag eine nichtnegative Zahl ist (schreiben Sie $A \geq 0$). Analog dazu ist A eine *positive Matrix*, wenn jedes $a_{ij} > 0$ ist (schreiben Sie $A > 0$). So sind beispielsweise die Hyperlink-Matrix H und die stochastische Matrix S (aus Kapitel 4), die die Grundlage des PageRank bilden, nichtnegative Matrizen, und die Google-Matrix G ist eine positive Matrix. Folglich bestimmen Eigenschaften von positiven und nichtnegativen Matrizen das Verhalten von PageRank, und die Perron-Frobenius-Theorie zeigt diese Eigenschaften auf, indem sie die Art der dominanten Eigenwerte und Eigenvektoren von positiven und nichtnegativen Matrizen beschreibt.

Perron

Ein Großteil der Mathematik von PageRank, HITS und verwandten Konzepten beinhaltet nicht-negative Matrizen und Graphen. Dieser Abschnitt liefert Ihnen das nötige Rüstzeug, um mit diesen Konzepten umzugehen. Das Theorem von Perron aus dem Jahr 1907 liefert die Grundlage für das Verständnis der Eigenstruktur von positiven Matrizen. Perrons Theorem für positive Matrizen wird im Folgenden dargelegt, und der Beweis findet sich in [127].

Frobenius-Form

Für jede imprimitive Matrix A mit einem Imprimivitätsindex $h > 1$ gibt es eine Permutationsmatrix P , für die gilt

$$P^T A P = \begin{pmatrix} 0 & A_{12} & 0 & \dots & 0 \\ 0 & 0 & A_{23} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & A_{h-1,h} \\ A_{h1} & \dots & \dots & 0 & 0 \end{pmatrix}, \quad (15.2.5)$$

wobei die Nullblöcke auf der Hauptdiagonale quadratisch sind.

15.3 MARKOV-KETTEN

Die mathematische Komponente des PageRank-Vektors von Google ist die stationäre Verteilung einer zeitdiskreten, zustandsbegrenzten Markov-Kette. Um also die Mathematik des PageRank zu verstehen und zu analysieren, ist es notwendig, die Konzepte der Markov-Kette zu verstehen, und das ist der Zweck dieses Abschnitts. Beginnen wir mit einigen Definitionen.

- Eine *stochastische Matrix* ist eine nichtnegative Matrix $P_{n \times n}$, in der jede Zeilensumme gleich 1 ist. Manche Autoren sagen "zeilenstochastisch", um dies von dem Fall zu unterscheiden, in dem jede Spaltensumme 1 ist.
- Ein *stochastischer Prozess* ist eine Menge von X_t $t=0$ die einen gemeinsamen Bereich haben $\{S_1, S_2, \dots, S_n\}$, der als *Zustandsraum* für den Prozess bezeichnet wird. Der Parameter t wird im Allgemeinen als Zeit betrachtet, und X_t steht für den Zustand des Prozesses zum Zeitpunkt t . Betrachten wir zum Beispiel den Prozess des Surfers im Internet, bei dem man nacheinander auf Links klickt, um von einer Webseite zur nächsten zu gelangen. Der Zustandsraum ist die Menge aller Webseiten, und die Zufallsvariable X_t ist die zum Zeitpunkt t betrachtete Webseite.
 - Um zu betonen, dass die Zeit diskret und nicht kontinuierlich betrachtet wird, wird häufig der Begriff "zeitdiskreter Prozess" verwendet, und der Begriff "endlicher Prozess" kann verwendet werden, um zu betonen, dass der Zustandsraum endlich und nicht unendlich ist. Unsere Diskussion beschränkt sich auf zeitdiskrete Prozesse mit endlichen Zuständen.
- Eine *Markov-Kette* ist ein stochastischer Prozess, der die *Markov-Eigenschaft* erfüllt

$$P(X_{t+1} = S_j \mid X_t = S_i, X_{t-1} = S_{i_{t-1}}, \dots, X_0 = S_i) = P(X_{t+1} = S_j \mid X_t = S_i)$$

für jedes $t = 0, 1, 2, \dots$. Die Schreibweise $P(E \mid F)$ bezeichnet die bedingte Wahrscheinlichkeit

dass das Ereignis E eintritt, wenn das Ereignis F eintritt - eine Wiederholung der elementaren Wahrscheinlichkeitsrechnung ist angebracht, falls dieses Konzept nicht bereits bekannt ist.

- Die Markov-Eigenschaft besagt, dass der Prozess in dem Sinne gedächtnislos ist, dass der Zustand der Kette im nächsten Zeitabschnitt nur vom aktuellen Zustand und nicht von der Vergangenheit der Kette abhängt. Zum Beispiel ist der Prozess des Surfens im Web eine Markov-Kette, vorausgesetzt, dass die nächste Seite, die der Websurfer besucht, nicht von den Seiten abhängt, die er in der Vergangenheit besucht hat - die Wahl hängt ab von

nur auf der aktuellen Seite. Mit anderen Worten, wenn der Surfer zufällig einen Link auf der aktuellen Seite auswählt, um zur nächsten Webseite zu gelangen, dann ist der Prozess eine Markov-Kette. Diese Art von Kette wird als "random walk" auf der Linkstruktur des Webs bezeichnet.

- Die *Übergangswahrscheinlichkeit* $p_{ij}(t) = P(X_t = S_j \mid X_{t-1} = S_i)$ ist die Wahrscheinlichkeit, sich zum Zeitpunkt t im Zustand S_j zu befinden, wenn sich die Kette zum Zeitpunkt $t - 1$ im Zustand S_i befindet. Man kann sich dies also einfach als die Wahrscheinlichkeit vorstellen, zum Zeitpunkt t von S_i nach S_j zu wechseln.
- Die *Übergangswahrscheinlichkeitsmatrix* $\mathbf{P}_{n \times n}(t) = [p_{ij}(t)]$ ist eindeutig eine nichtnegative Matrix, und ein wenig Nachdenken sollte Sie davon überzeugen, dass die Summe jeder Zeile 1 sein muss. Mit anderen Worten: $\mathbf{P}(t)$ ist eine stochastische Matrix für jedes t .
- Eine *stationäre Markov-Kette* ist eine Kette, bei der sich die Übergangswahrscheinlichkeiten nicht mit der Zeit ändern, d. h. $p_{ij}(t) = p_{ij}$ für alle t . Stationäre Ketten werden auch als *homogene Ketten* bezeichnet.
 - In diesem Fall ist die Übergangswahrscheinlichkeitsmatrix eine konstante stochastische Matrix
 $\mathbf{P} = [p_{ij}]$. In der Folge wird Stationarität angenommen.
 - Auf diese Weise definiert jede Markov-Kette eine stochastische Matrix, aber auch der Umkehrschluss gilt: Jede stochastische Matrix $\mathbf{P}_{n \times n}$ definiert eine Markov-Kette mit n Zuständen, da die Einträge p_{ij} eine Menge von Übergangswahrscheinlichkeiten definieren, die als stationäre Markov-Kette mit n Zuständen interpretiert werden können.
- Eine *irreduzible Markov-Kette* ist eine Kette, für die die Übergangswahrscheinlichkeitsmatrix \mathbf{P} eine irreduzible Matrix ist. Eine Kette gilt als *reduzierbar*, wenn \mathbf{P} eine reduzierbare Matrix ist.
 - Eine *periodische Markov-Kette* ist eine irreduzible Kette, deren Übergangswahrscheinlichkeitsmatrix \mathbf{P} eine imprimitive Matrix ist. Diese Ketten werden periodisch genannt, weil jeder Zustand nur zu periodischen Zeitpunkten eingenommen werden kann, wobei die Periode der Index der Imprimitivität ist. Betrachten wir zum Beispiel eine irreduzible Kette, deren Imprimitivitätsindex $h = 3$ ist. Die Frobenius-Form (15.2.5) bedeutet, dass die Zustände neu geordnet (umbenannt) werden können, um drei Gruppen von Zuständen zu bilden, für die die Übergangsmatrix und ihre Potenzen die folgende Form haben

$$\mathbf{P} = \begin{pmatrix} \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & 0 \\ 0 & 0 & 0 & \bullet \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet \end{pmatrix}, \quad \mathbf{P}^4 = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet \end{pmatrix} \dots,$$

wobei sich dieses Muster unbegrenzt fortsetzt. Wenn die Kette in einem Zustand in Cluster i beginnt, dann stellt dieses periodische Muster sicher, dass die Kette nur am Ende jedes dritten Schrittes einen Zustand in Cluster i einnehmen kann - siehe transiente Eigenschaften auf Seite 179.

- Eine *aperiodische Markov-Kette* ist eine irreduzible Kette, deren

Übergangswahrscheinlichkeitsmatrix \mathbf{P} eine primitive Matrix ist.

- Ein *Wahrscheinlichkeitsverteilungsvektor* (oder kurz "Wahrscheinlichkeitsvektor") ist definiert als ein nichtnegativer Zeilenvektor $\mathbf{p}^T = (p_1, p_2, \dots, p_n)$, so dass $\sum_k p_k = 1$. (Jede Zeile in einer stochastischen Matrix ist ein Wahrscheinlichkeitsvektor).
- Ein *stationärer Wahrscheinlichkeitsverteilungsvektor* für eine Markov-Kette, deren Übergangswahrscheinlichkeitsmatrix \mathbf{P} ist, ist ein Wahrscheinlichkeitsvektor $\boldsymbol{\pi}^T$, so dass $\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T$.

- The k^{th} step probability distribution vector for an n -state chain is defined to be

$$\mathbf{p}^T(k) = p_1(k), p_2(k), \dots, p_n(k), \text{ wobei } p_j(k) = P(X_k = S_j).$$

Mit anderen Worten: $p_j(k)$ ist die Wahrscheinlichkeit, sich nach dem k^{th} Schritt, aber vor dem $(k+1)^{\text{ten}}$ Schritt im Zustand j^{th} zu befinden.

- The initial distribution vector is

$$\mathbf{p}^T(0) = p_1(0), p_2(0), \dots, p_n(0), \text{ wobei } p_j(0) = P(X_0 = S_j).$$

Mit anderen Worten: $p_j(0)$ ist die Wahrscheinlichkeit, dass die Kette in S_j beginnt.

Um diese Konzepte zu veranschaulichen, betrachten Sie das winzige dreiseitige Netz in Abbildung 15.2,

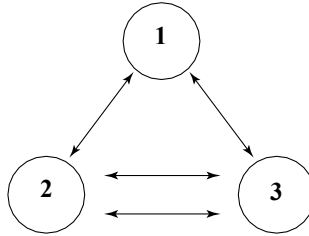


Abbildung 15.2

wobei die Pfeile Links anzeigen - z. B. enthält Seite 2 zwei Links zu Seite 3 und umgekehrt. Die Markov-Kette, die durch einen Random Walk auf dieser Link-Struktur definiert ist, entwickelt sich, wenn ein Web-Surfer auf einen zufällig ausgewählten Link auf der gerade betrachteten Seite klickt, und die Übergangs-Wahrscheinlichkeitsmatrix für diese Kette ist die irreduzible stochastische Matrix

$$\mathbf{H} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 0 & 2/3 \\ 1/3 & 2/3 & 0 \end{pmatrix}.$$

In diesem Beispiel ist \mathbf{H} (die *Hyperlink-Matrix*) stochastisch, aber wenn es einen Dangling-Knoten gegeben hätte (eine Seite, die keine Links zum Anklicken enthält), dann hätte \mathbf{H} eine Nullzeile, in diesem Fall wäre \mathbf{H} nicht stochastisch und der Prozess wäre keine Markov-Kette. ¹

Wenn unser Web-Surfer auf Seite 2 in Abbildung 15.2 beginnt, dann ist der Anfangsverteilungsvektor für die Kette $\mathbf{p}^T(0) = (0, 1, 0) = \mathbf{e}^T$. Wählt der Surfer jedoch einfach eine zufällige Anfangsseite aus, dann ist $\mathbf{p}^T(0) = (1/3, 1/3, 1/3) = \mathbf{e}^T/3$ der *Vektor der Gleichverteilung*. Eine Standard-Eigenwertberechnung zeigt, dass $\sigma(\mathbf{H}) = \{1, -1/3, -2/3\}$, so dass es offensichtlich ist, dass \mathbf{H} eine nichtnegative Matrix mit dem Spektralradius $\rho(\mathbf{H}) = 1$ ist.

Die Tatsache, dass $\rho(\mathbf{H}) = 1$ ist, ist ein Merkmal aller stochastischen Matrizen $\mathbf{P}_{n \times n}$, denn Zeilensummen gleich 1 zu haben, bedeutet, dass $\|\mathbf{P}\|_\infty = 1$ oder, äquivalent, $\mathbf{P}\mathbf{e} = \mathbf{e}$ ist, wobei \mathbf{e} die Spalte mit allen 1en ist. Da $(1, \mathbf{e})$ ein Eigenpaar für jede stochastische Matrix ist, und da $\rho(\cdot) \leq \|\cdot\|$ für jede Matrixnorm ist, folgt daraus, dass

$$1 \leq \rho(\mathbf{P}) \leq \|\mathbf{P}\|_\infty = 1 \Rightarrow \rho(\mathbf{P}) = 1. \quad (15.3.1)$$

¹ Wie bereits erläutert, ändert Google aus diesem Grund die rohe Hyperlink-Matrix, bevor es den PageRank berechnet.

Außerdem ist \mathbf{e} ein positiver Eigenvektor, der mit $\rho(\mathbf{P}) = 1$ verbunden ist. Aber Vorsicht! Das bedeutet nicht, dass man \mathbf{e} zwangsläufig als Perron-Vektor für \mathbf{P} bezeichnen kann, da \mathbf{P} möglicherweise nicht irreduzibel sein kann. z.B.

$$\begin{pmatrix} 2 & .5 & .5 \\ 0 & 1 & \end{pmatrix}$$

Bei fast allen Markov-Analysen geht es um Fragen des Übergangsverhaltens der Kette und des Grenzwertens, und die Standardziele sind wie folgt.

- Beschreiben Sie die k^{th} Stufenverteilung $\mathbf{p}^T(k)$ für einen beliebigen Ausgangsverteilungsvektor $\mathbf{p}^T(0)$.
- Bestimmen Sie, ob $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$ existiert, und wenn ja, finden Sie den Wert von $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$.
- Wenn $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$ nicht existiert, bestimmen Sie, ob der Cesàro-Grenzwert

$$\lim_{k \rightarrow \infty} \frac{\mathbf{p}^T(0) + \mathbf{p}^T(1) + \dots + \mathbf{p}^T(k-1)}{k}$$

existiert, und, wenn ja, seinen Wert zu ermitteln und seine Bedeutung zu interpretieren.

Transientes Verhalten

Given an initial distribution vector $\mathbf{p}^T(0) = (p_1(0), p_2(0), \dots, p_n(0))$, the first aim is to calculate the probability of being in any given state after the first transition (but before the second)—i.e., determine $\mathbf{p}^T(1) = (p_1(1), p_2(1), \dots, p_n(1))$. Let \wedge and \vee respectively denote *AND* and *OR*. It follows from elementary probability theory that for each j ,

$$\begin{aligned} p_j(1) &= P(X_1 = S_j) = P(X_1 = S_j \wedge (X_0 = S_1 \vee X_0 = S_2 \vee \dots \vee X_0 = S_n)) \\ &= P(X_1 = S_j \wedge X_0 = S_1) \vee P(X_1 = S_j \wedge X_0 = S_2) \vee \dots \vee P(X_1 = S_j \wedge X_0 = S_n) \\ &= \sum_{i=1}^n P(X_1 = S_j \wedge X_0 = S_i) = \sum_{i=1}^n P(X_0 = S_i) P(X_1 = S_j | X_0 = S_i) \\ &= \sum_{i=1}^n p_i(0) p_{ij} \end{aligned}$$

Mit anderen Worten: $\mathbf{p}^T(1) = \mathbf{p}^T(0)\mathbf{P}$, was die Entwicklung von den anfänglichen Verteilungen zur Verteilung nach einem Schritt beschreibt. Die Markov-Eigenschaft "kein Gedächtnis" liefert den Zustand am Ende von zwei Schritten - sie besagt, dass man einfach neu beginnen muss, aber mit $\mathbf{p}^T(1)$ als Anfangsverteilung. Folglich ist $\mathbf{p}^T(2) = \mathbf{p}^T(1)\mathbf{P}$, und $\mathbf{p}^T(3) = \mathbf{p}^T(2)\mathbf{P}$, usw., und eine sukzessive Substitution ergibt

$$\mathbf{p}^T(k) = \mathbf{p}^T(0)\mathbf{P}^k, \quad (15.3.2)$$

was einfach ein Spezialfall der Potenzmethode (15.1.20) ist, mit der Ausnahme, dass die linkshändige Vektor-Matrix-Multiplikation verwendet wird. Wenn außerdem $\mathbf{P}^k = \mathbf{P}^{(k)}$ (15.3.2) gilt, dann ist $\mathbf{p}^T(k) = \mathbf{p}^T(0)\mathbf{P}^{(k)}$ für jedes $i = 1, 2, \dots, n$. Es folgt eine Zusammenfassung.

ij

²Die Notwendigkeit, Irreduzibilität zu erzwingen, ist ein weiterer Grund, warum Google die rohe Hyperlink-Matrix verändert.

Transiente Eigenschaften

Wenn \mathbf{P} die Übergangswahrscheinlichkeitsmatrix für eine Markov-Kette mit Zuständen ist

$\{S_1, S_2, \dots, S_n\}$, dann ist jede der folgenden Aussagen wahr.

- Die Matrix \mathbf{P}^k stellt die Übergangswahrscheinlichkeitsmatrix für k Schritte im Sinne von \mathbf{P} dar. Das heißt, der (i, j) -Eintrag $[\mathbf{P}^k]_{ij}$ ist die Wahrscheinlichkeit, von S_i nach S_j zu gelangen, wenn k Schritte gemacht werden.
- Der Vektor der Schrittverteilung k^{th} ist gegeben durch $\mathbf{p}^T(k) = \mathbf{p}^T(0)\mathbf{P}^k$.

Verhalten einschränken

Die Analyse der begrenzenden Eigenschaften von Markov-Ketten erfordert, dass die Klasse der stochastischen Matrizes (und damit die Klasse der stationären Markov-Ketten) in vier sich gegenseitig ausschließende Kategorien unterteilt wird.

- (1) \mathbf{P} ist irreduzibel mit $\lim_{k \rightarrow \infty} \mathbf{P}^k$ vorhanden (d.h. \mathbf{P} ist primitiv).
- (2) \mathbf{P} ist irreduzibel, wobei $\lim_{k \rightarrow \infty} \mathbf{P}^k$ nicht existiert (d. h., \mathbf{P} ist imprimitiv).
- (3) \mathbf{P} ist reduzierbar mit $\lim_{k \rightarrow \infty} \mathbf{P}^k$ vorhanden.
- (4) \mathbf{P} ist reduzierbar wobei $\lim_{k \rightarrow \infty} \mathbf{P}^k$ nicht existiert.

Im Fall (1) (einer aperiodischen Kette) kann $\lim_{k \rightarrow \infty} \mathbf{P}^k$ leicht ausgewertet werden. Der Perron-Vektor für \mathbf{P} ist \mathbf{e}/n (der Vektor der Gleichverteilung). Wenn also $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n)^T$ der Perron-Vektor für \mathbf{P}^T ist (d. h. $\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T$), dann gilt nach (15.2.4),

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \frac{(\mathbf{e}/n)\boldsymbol{\pi}^T}{\boldsymbol{\pi}^T(\mathbf{e}/n)} = \frac{\mathbf{e}\boldsymbol{\pi}^T}{\boldsymbol{\pi}^T \mathbf{e}} = \mathbf{e}\boldsymbol{\pi}^T = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix} > \mathbf{0}. \quad (15.3.3)$$

Wenn also \mathbf{P} primitiv ist, dann existiert eine begrenzte Wahrscheinlichkeitsverteilung, die gegeben ist durch

$$\lim_{k \rightarrow \infty} \mathbf{p}^T(k) = \lim_{k \rightarrow \infty} \mathbf{p}^T(0)\mathbf{P}^k = \mathbf{p}^T(0)\mathbf{e}\boldsymbol{\pi}^T = \boldsymbol{\pi}^T. \quad (15.3.4)$$

Beachten Sie, dass, da $\sum_k p_k(0) = 1$ ist, fällt der Term $\mathbf{p}^T(0)\mathbf{e}$ weg, so dass der Wert des Grenzwertes *unabhängig* vom Wert der Anfangsverteilung $\mathbf{p}^T(0)$ ist, was nicht allzu überraschend ist.

Im Fall (2), in dem \mathbf{P} irreduzibel, aber imprimitiv ist, stellt (15.2.4) sicher, dass $\lim_{k \rightarrow \infty} \mathbf{P}^k$ nicht existieren kann, und daher kann $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$ nicht existieren (andernfalls würde die Annahme $\mathbf{p}^T(0) = \mathbf{e}^T$ für jedes i sicherstellen, dass \mathbf{P}^k einen Grenzwert hat). Die Ergebnisse auf Seite 173 stellen jedoch sicher, dass die Eigenwerte von \mathbf{P} , die auf dem Einheitskreis liegen, alle einfach sind, so dass \mathbf{P} nach (15.1.18) für den Spektralprojektor \mathbf{G} , der mit dem Eigenwert $\lambda = 1$ verbunden ist, Cesa'ro-summierbar ist. Aus (15.1.12) und der Tatsache, dass \mathbf{e}/n der Perron-Vektor für \mathbf{P} ist, folgt, dass, wenn

$\pi^T = (\pi_1, \pi_2, \dots, \pi_n)$ ist der linke Perron-Vektor, dann

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}}{k} = \frac{(e/n)\pi^T}{\pi^T(e/n)} = \frac{e\pi^T}{\pi^T e} = e\pi^T = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_2 & \pi_2 & \dots & \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix},$$

was genau die gleiche Form hat wie der Grenzwert (15.3.3) für den primitiven Fall.

Folglich haben die k^{th} Stufenverteilungen einen Cesa'ro-Grenzwert, der gegeben ist durch

$$\lim_{k \rightarrow \infty} \frac{\# \frac{\mathbf{p}^T(0) + \mathbf{p}^T(1) + \dots + \mathbf{p}^T(k-1)}{k}}{\$} = \lim_{k \rightarrow \infty} \mathbf{p}^T(0) \frac{\# \frac{\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}}{k}}{\$} \\ = \mathbf{p}^T(0)e\pi^T = \pi^T,$$

und genau wie im primitiven Fall (15.3.4) ist dieser Cesa'ro-Grenzwert unabhängig von der Anfangsverteilung. Um die Bedeutung dieser Cesa'ro-Grenze zu verstehen,

konzentrieren wir uns auf einen Zustand, sagen wir S_j , und

lassen $\{Z_k\}_{k=0}^\infty$ sind Zufallsvariablen, die die Anzahl der Besuche auf S_j zählen, indem sie

$$Z_0 = \begin{cases} 1, & \text{wenn die Kette in } S_j \text{ beginnt,} \\ 0 & \text{sonst,} \end{cases}$$

and for $i > 1$,

$$Z_i = \begin{cases} 1, & \text{wenn sich die Kette nach dem Zug } i^{\text{th}} \text{ in } S_j \text{ befindet,} \\ 0 & \text{sonst.} \end{cases}$$

Beachten Sie, dass $Z_0 + Z_1 + \dots + Z_{k-1}$ die Anzahl der Besuche auf S_j vor dem Zug k^{th} zählt, so dass $(Z_0 + Z_1 + \dots + Z_{k-1})/k$ den Anteil der Besuche auf S_j vor dem Zug k^{th} darstellt. Der erwartete (oder mittlere) Wert von jedem Z_i ist

$$E[Z_i] = 1 - P(Z_i = 1) + 0 - P(Z_i = 0) = P(Z_i = 1) = p_j(i).$$

Da die Erwartung linear ist, ist der erwartete Anteil der Treffer von S_j vor dem Zug k

$$\begin{aligned} E \frac{\# Z_0 + Z_1 + \dots + Z_{k-1}}{k} &= \frac{E[Z_0] + E[Z_1] + \dots + E[Z_{k-1}]}{k} \\ &= \frac{p_j(0) + p_j(1) + \dots + p_j(k-1)}{k} = \frac{\# \mathbf{p}^T(0) + \mathbf{p}^T(1) + \dots + \mathbf{p}^T(k-1)}{k} \\ &\rightarrow \pi_j. \end{aligned}$$

Mit anderen Worten, der langfristige Zeitanteil, den die Kette in S_j verbringt, ist π_j , das ist die j^{th} Komponente des Cesa'ro-Grenzwerts oder, äquivalent, die j^{th} Komponente des linken Perron-Vektors für \mathbf{P} . Wenn $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$ existiert, ist es leicht zu argumentieren, dass

$$\lim_{k \rightarrow \infty} \mathbf{p}^T(k) = \lim_{k \rightarrow \infty} \frac{\# \mathbf{p}^T(0) + \mathbf{p}^T(1) + \dots + \mathbf{p}^T(k-1)}{k} \$$$

Die Interpretation der Grenzwertverteilung $\lim_{k \rightarrow \infty} \mathbf{p}^T(k)$ für den primitiven Fall ist also genau dieselbe wie die Interpretation des Cesa'ro-Grenzwerts für den imprimitiven Fall. Im Folgenden finden Sie eine Zusammenfassung der irreduziblen Ketten.

Unreduzierbare Markov-Ketten

Sei \mathbf{P} die Übergangswahrscheinlichkeitsmatrix für eine irreduzible Markov-Kette auf den Zuständen $\{S_1, S_2, \dots, S_n\}$, und sei $\boldsymbol{\pi}^T$ der linke Perron-Vektor für \mathbf{P} ($\|\boldsymbol{\pi}\|_1 = 1$, $\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T$, $\pi_1 = 1$). Für jede Anfangsverteilung $\mathbf{p}^T(0)$ gilt Folgendes.

- Die Übergangsmatrix für k^{th} Schritte ist \mathbf{P}^k . Mit anderen Worten: Der (i, j) -Eintrag in \mathbf{P}^k ist die Wahrscheinlichkeit, in genau k Schritten von S_i nach S_j zu gelangen.
- Der Vektor der Schrittverteilung k^{th} ist gegeben durch $\mathbf{p}^T(k) = \mathbf{p}^T(0) \mathbf{P}^k$.
- Wenn \mathbf{P} primitiv ist (die Kette ist also aperiodisch), und wenn \mathbf{e} die Spalte aller 1en ist, dann

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \mathbf{e} \boldsymbol{\pi}^T \quad \lim_{k \rightarrow \infty} \mathbf{p}^T(k) = \boldsymbol{\pi}^T.$$

- Wenn \mathbf{P} imprimitiv ist (also die Kette periodisch ist), dann

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}}{k} = \mathbf{e} \boldsymbol{\pi}^T$$

und

$$\lim_{k \rightarrow \infty} \frac{\mathbf{p}^T(0) + \mathbf{p}^T(1) + \dots + \mathbf{p}^T(k-1)}{k} = \boldsymbol{\pi}^T.$$

- Unabhängig davon, ob \mathbf{P} primitiv oder imprimitiv ist, ist die j^{th} Komponente π_j von $\boldsymbol{\pi}^T$ ist der langfristige Anteil der Zeit, in der sich die Kette in S_j befindet.
- Der Vektor $\boldsymbol{\pi}^T$ ist der einzige *stationäre Verteilungsvektor* für die Kette, da er der einzige Wahrscheinlichkeitsverteilungsvektor ist, der $\boldsymbol{\pi}^T \mathbf{P} = \boldsymbol{\pi}^T$ erfüllt.

Reduzierbare Markov-Ketten

Das Perron-Frobenius-Theorem ist nicht direkt auf reduzierbare Ketten anwendbar (Ketten, für die \mathbf{P} eine reduzierbare Matrix ist), so dass die Strategie zur Analyse reduzierbarer Ketten darin besteht, die Situation so weit wie möglich auf den irreduziblen Fall zu reduzieren. Wenn \mathbf{P} reduzierbar ist, dann gibt es per Definition eine Permutationsmatrix \mathbf{Q} und quadratische Matrizen \mathbf{X} und \mathbf{Z} , so dass

$$\mathbf{Q}^T \mathbf{P} \mathbf{Q} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}. \text{ } \text{Für Einfachheit halber wird dies mit } \mathbf{P} \sim \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix} \text{ ausgedrückt.}$$

Wenn \mathbf{X} oder \mathbf{Z} reduzierbar ist, kann eine weitere symmetrische Permutation durchgeführt werden, die Folgendes ergibt

$$\begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix} \sim \begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{0} & \mathbf{U} \end{pmatrix} \begin{pmatrix} \mathbf{V} & \mathbf{W} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}^T, \text{ wobei } \mathbf{R}, \mathbf{U} \text{ und } \mathbf{W} \text{ quadratisch sind.}$$

Die Wiederholung dieses Prozesses führt schließlich zu

$$\mathbf{P} \sim \begin{pmatrix} \mathbf{0} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1k} \\ \vdots & \mathbf{X}_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_{kk} \end{pmatrix}, \quad \text{wobei jedes } \mathbf{X}_{ii} \text{ irreduzibel ist oder } \mathbf{X}_{ii} = [0]_{1 \times 1}.$$

Wenn es Zeilen gibt, die nur in diagonalen Blöcken Einträge ungleich Null haben, dann werden alle diese Zeilen symmetrisch nach unten permutiert, um Folgendes zu erhalten

$$\mathbf{P} \sim \left(\begin{array}{cccc|cccc} \mathbf{P}_{11} & \mathbf{P}_{12} & \cdots & \mathbf{P}_{1r} & \mathbf{P}_{1,r+1} & \mathbf{P}_{1,r+2} & \cdots & \mathbf{P}_{1m} \\ \mathbf{0} & \mathbf{P}_{22} & \cdots & \mathbf{P}_{2r} & \mathbf{P}_{2,r+1} & \mathbf{P}_{2,r+2} & \cdots & \mathbf{P}_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P}_{rr} & \mathbf{P}_{r,r+1} & \mathbf{P}_{r,r+2} & \cdots & \mathbf{P}_{rm} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{P}_{r+1,r+1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{P}_{r+2,r+2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P}_{mm} \end{array} \right), \quad (15.3.5)$$

wobei jedes $\mathbf{P}_{11}, \dots, \mathbf{P}_{rr}$ entweder irreduzibel oder $[0]_{1 \times 1}$ ist und $\mathbf{P}_{r+1,r+1}, \dots, \mathbf{P}_{mm}$ irreduzierbar sind (sie können nicht Null sein, da jede Zeilensumme gleich 1 ist). Wie auf Seite 171 erwähnt, besteht die Wirkung einer symmetrischen Permutation einfach darin, die Knoten in $\mathbf{G}(\mathbf{P})$ neu zu benennen oder, äquivalent dazu, die Zustände in der Kette neu zu ordnen. Wenn die Zustände einer Kette so umgeordnet wurden, dass \mathbf{P} die Form auf der rechten Seite von (15.3.5) annimmt, sagen wir, dass \mathbf{P} die *kanonische Form für reduzierbare Matrizen* hat.

Die Ergebnisse auf Seite 173 garantieren, dass wenn eine irreduzible stochastische Matrix \mathbf{P} h Eigenwerte auf dem Einheitskreis hat, dann sind diese h Eigenwerte die h^{th} Wurzeln der Einheit, und jeder ist ein einfacher Eigenwert für \mathbf{P} . Dasselbe kann nicht für reduzierbare stochastische Matrizen gesagt werden, aber (15.3.5) führt zum nächstbesten Ergebnis (der Beweis dafür ist in [127]).

Einheitseigenwerte

Die *Einheitseigenwerte* sind diejenigen Eigenwerte, die auf dem Einheitskreis liegen. Für jede stochastische Matrix $\mathbf{P}_{n \times n}$ sind die folgenden Aussagen wahr.

- Jeder Einheits-Eigenwert von \mathbf{P} ist semisimpel.
- Jeder Einheitseigenwert hat die Form $\lambda = e^{2k\pi i/h}$ für einige $k < h \leq n$.
- Insbesondere ist $\rho(\mathbf{P}) = 1$ immer ein semisimpler Eigenwert von \mathbf{P} .

Die Diskussion auf Seite 163 besagt, dass eine Matrix $\mathbf{A}_{n \times n}$ dann und nur dann Cesàro-summierbar ist, wenn $\rho(\mathbf{A}) < 1$ oder $\rho(\mathbf{A}) = 1$ ist und jeder Eigenwert auf dem Einheitskreis semisimpel ist. Da das obige Ergebnis besagt, dass Letzteres für alle stochastischen Matrizen \mathbf{P} gilt, haben wir die folgende mächtige Erkenntnis für alle stochastischen Matrizen.

Alle stochastischen Matrizen sind summierbar

Jede stochastische Matrix \mathbf{P} ist Cesàro-summierbar in dem Sinne, dass

$$\lim_{k \rightarrow \infty} \frac{1}{k} (\mathbf{I} + \mathbf{P} + \mathbf{P}^2 + \cdots + \mathbf{P}^{k-1})$$
 existiert immer, und wenn \mathbf{P} auf Seite 163 erläutert, ist \mathbf{G} der Grenzwert der Spektralprojektor \mathbf{G} auf $N(\mathbf{I} - \mathbf{P})$ entlang $R(\mathbf{I} - \mathbf{P})$.

Die Struktur und Interpretation des Cesa`ro-Grenzwerts für den Fall, dass \mathbf{P} eine irreduzible stochastische Matrix ist, wurde auf Seite 181 entwickelt. Um das Bild zu vervollständigen, bleibt nur noch, die Natur von $\lim_{k \rightarrow \infty} (\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1})/k$ für den reduzierbaren Fall zu analysieren.

Angenommen, dass \mathbf{P} in der Kanon-Blockform (15.3.5) ist, wobei

$$\mathbf{T}_{11} = \begin{pmatrix} p_{11} & \dots & p_{1r} \\ \vdots & & \vdots \\ p_{rr} & \dots & p_{rm} \end{pmatrix}, \quad \mathbf{T}_{12} = \begin{pmatrix} p_{1,r+1} & \dots & p_{1m} \\ \vdots & & \vdots \\ p_{r,r+1} & \dots & p_{rm} \end{pmatrix}, \quad \mathbf{T}_{22} = \begin{pmatrix} p_{r+1,r+1} & \dots & p_{r+1,m} \\ \vdots & & \vdots \\ p_{mm} & \dots & p_{mm} \end{pmatrix}$$

Da jede Zeile in \mathbf{T}_{11} einen von Null verschiedenen diagonalen Block hat, folgt daraus, dass $\rho(\mathbf{P}_{kk}) < 1$ für jedes $k = 1, 2, \dots, r$. Folglich ist $\rho(\mathbf{T}_{11}) < 1$, und

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{T}_{11} + \dots + \mathbf{T}_{11}^{k-1}}{k} = \lim_{k \rightarrow \infty} \frac{\mathbf{T}_k}{11} = \mathbf{0}$$

Außerdem sind $\mathbf{P}_{r+1,r+1}, \dots, \mathbf{P}_{mm}$ jeweils irreduzible stochastische Matrizen. Wenn also π^T der linke Perron-Vektor für \mathbf{P}_{jj} ist, $r+1 \leq j \leq m$, dann ergibt (15.1.12) in Verbindung mit (15.1.18)

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{T}_{22} + \dots + \mathbf{T}_{22}^{k-1}}{k} = \begin{pmatrix} \pi^T & & \\ & \ddots & \\ & & \pi^T \end{pmatrix} = \mathbf{E}.$$

Aus (15.2.4) ergibt sich, dass $\lim_{k \rightarrow \infty} \mathbf{T}_{22}^k$ existiert nur dann, wenn $\mathbf{P}_{r+1,r+1}, \dots, \mathbf{P}_{mm}$ jedes Primitiv, in diesem Fall $\lim_{k \rightarrow \infty} \mathbf{T}_{22}^k = \mathbf{E}$. Daher sind die Grenzen, seien sie Cesa`ro oder

ordinary (falls vorhanden), haben alle die Form

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}}{k} = \mathbf{0} \quad \mathbf{Z} = \mathbf{G} = \lim_{k \rightarrow \infty} \mathbf{P}^k \text{ (wenn es existiert).}$$

Um die genaue Beschaffenheit von \mathbf{Z} zu bestimmen, muss man die Tatsache nutzen, dass $R(\mathbf{G}) = N(\mathbf{I} - \mathbf{P})$ ist (weil \mathbf{G} der Projektor auf $N(\mathbf{I} - \mathbf{P})$ entlang $R(\mathbf{I} - \mathbf{P})$ ist), um zu schreiben

$$(\mathbf{I} - \mathbf{P})\mathbf{G} = \mathbf{0} \Rightarrow \begin{pmatrix} \mathbf{I} - \mathbf{T}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{T}_{22} \end{pmatrix} \mathbf{Z} = \mathbf{0} \Rightarrow (\mathbf{I} - \mathbf{T}_{11})\mathbf{Z} = \mathbf{T}_{12} \mathbf{E}.$$

Da $\mathbf{I} - \mathbf{T}_{11}$ nichtsingulär ist (weil $\rho(\mathbf{T}_{11}) < 1$), folgt daraus, dass

$$\mathbf{Z} = (\mathbf{I} - \mathbf{T}) \mathbf{T}_{11}^{-1} \mathbf{T}_{12} \mathbf{E},$$

und daraus ergeben sich die folgenden Ergebnisse bezüglich der Grenzen der reduzierbaren Ketten.

Reduzierbare Markov-Ketten

Wenn die Zustände in einer reduzierbaren Markov-Kette so geordnet wurden, dass die Übergangsmatrix die kanonische Form annimmt

$$\mathbf{P} = \begin{pmatrix} T_{11} & T_{12} \\ \mathbf{0} & T_{22} \end{pmatrix}$$

die in (15.3.5) beschrieben ist, und wenn π^T der linke Perron-Vektor für \mathbf{P} ($\pi_j = 1 \leq j \leq m$), dann ist $\mathbf{I} - T_{11}$ nichtsingulär, und

$$\lim_{k \rightarrow \infty} \frac{\mathbf{I} + \mathbf{P} + \dots + \mathbf{P}^{k-1}}{\mathbf{P}^{k-1}} = \begin{pmatrix} \mathbf{0} & (\mathbf{I} - T_{11})^{-1} T_{12} \\ \mathbf{0} & \mathbf{E} \end{pmatrix},$$

wobei

$$\mathbf{E} = \begin{pmatrix} \mathbf{e}_{r+1}^T \\ \pi \\ \vdots \\ \pi^T m \end{pmatrix}$$

Außerdem existiert $\lim_{k \rightarrow \infty} \mathbf{P}^k$ wenn und nur wenn die stochastischen Matrizen $\mathbf{P}_{r+1, r+1}, \dots, \mathbf{P}_{mm}$ in (15.3.5) jeweils primitiv sind, wobei in diesem Fall

$$\lim_{k \rightarrow \infty} \mathbf{P}^k = \begin{pmatrix} \mathbf{0} & (\mathbf{I} - T_{11})^{-1} T_{12} \\ \mathbf{0} & \mathbf{E} \end{pmatrix}. \quad (15.3.6)$$

Transiente und ergodische Klassen

Wenn die Zustände einer Kette so umgeordnet sind, dass \mathbf{P} die kanonische Form (15.3.5) hat, wird die Teilmenge der Zustände, die \mathbf{P}_{kk} für $1 \leq k \leq r$ entspricht, als k^{th} *transiente Klasse* bezeichnet, da eine einmal verlassene transiente Klasse nicht wieder betreten werden kann. Die Teilmenge der Zustände, die $\mathbf{P}_{r+j, r+j}$ für $j \geq 1$ entspricht, wird die j^{th} *ergodische Klasse* genannt. Jede ergodische Klasse ist eine irreduzible Markov-Kette für sich, die in die größere reduzierbare Kette eingebettet ist. Von nun an nehmen wir an, dass die Zustände in reduzierbaren Ketten so geordnet sind, dass \mathbf{P} die kanonische Form (15.3.5) hat.

Jede reduzierbare Kette tritt schließlich in eine der ergodischen Klassen ein, aber was dann geschieht, hängt davon ab, ob die ergodische Klasse primitiv ist oder nicht. Wenn $\mathbf{P}_{r+j, r+j}$ primitiv ist, dann pendelt sich die Kette auf einen stationären Zustand ein, der durch den linken Perron-Vektor von $\mathbf{P}_{r+j, r+j}$ definiert ist, aber wenn $\mathbf{P}_{r+j, r+j}$ imprimitiv ist, dann wird der Prozess für immer in der j^{th} ergodischen Klasse oszillieren. Viel mehr lässt sich über den Grenzwert nicht sagen, aber es gibt immer noch wichtige Fragen dazu, in welcher ergodischen Klasse die Kette am Ende landet und wie lange sie braucht, um dorthin zu gelangen. Diesmal hängt die Antwort davon ab, wo die Kette beginnt, d. h. von der Anfangsverteilung.

Der Einfachheit halber bezeichnen wir mit T_i die i^{th} transiente Klasse und mit E_j die j^{th} ergodische Klasse. Nehmen wir an, dass die Kette in einem bestimmten Übergangszustand beginnt - sagen wir, wir beginnen im Zustand p^{th} von T_i . Da die vorliegende Frage nur betrifft, welche ergodische Klasse getroffen wird, aber nicht, was

nach dem Eintritt in diese Klasse geschieht, können wir auch jeden Zustand in jeder ergodischen Klasse in eine Falle umwandeln, indem wir $p_{r+j,r+j} = 1$ für jedes $j \geq 1$ in (15.3.5) setzen. Die Übergangsmatrix für diese

modifizierte Kette ist $\tilde{P} = \begin{pmatrix} T_{11} & T_{12} \\ 0 & I \end{pmatrix}$, und aus (15.3.6) folgt, dass $\lim_{k \rightarrow \infty} \tilde{P}^k$ existiert und hat die Form

$$\lim_{k \rightarrow \infty} \tilde{P}^k = \begin{pmatrix} 0 & (I - T_{11})^{-1} T_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} \begin{matrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{matrix} & \begin{matrix} L_{1,1} & L_{1,2} & \dots & L_{1,s} \\ L_{2,1} & L_{2,2} & \dots & L_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ L_{r,1} & L_{r,2} & \dots & L_{r,s} \end{matrix} \\ \begin{matrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{matrix} & \begin{matrix} I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I \end{matrix} \end{pmatrix}.$$

Folglich stellt der (p, q) -Eintrag im Block L_{ij} die Wahrscheinlichkeit dar, den Zustand q^{th} in E_j zu erreichen, wenn man vom Zustand p^{th} in T_i ausgeht. Wenn also e der Vektor aller 1en ist, dann ist die Wahrscheinlichkeit, irgendwo in E_j zu landen, gegeben durch

$$P(\text{Absorption in } E_j \mid \text{Start im Zustand } p^{\text{th}} \text{ von } T_i) = \sum_k L_{ij}^{pk} = L_{ij} e.$$

Wenn $p^T(0)$ eine Anfangsverteilung für den Start in den verschiedenen Zuständen von T_i ist, dann

$$P(\text{Absorption in } E_j \mid p_i^T(0)) = p_i^T(0) L_{ij} e.$$

Die erwartete Anzahl der Schritte, die erforderlich sind, um erstmals einen ergodischen Zustand zu erreichen, wird wie folgt bestimmt. Zählen Sie, wie oft sich die Kette im Übergangszustand S_j befindet, wenn sie im Übergangszustand S_i beginnt, indem Sie das auf Seite 180 angeführte Argument anwenden. Das heißt, wenn die Kette in S_i beginnt, sei

$$Z_0 = 1, \text{ wenn } S_i = S_j, \quad Z_k = 1, \text{ wenn sich die Kette nach Schritt } k \text{ in } S_j \text{ befindet,} \\ 0 \text{ sonst,} \quad 0 \text{ sonst.}$$

Seit

$$E[Z_k] = 1 - P(Z_k = 1) + 0 - P(Z_k = 0) = P(Z_k = 1) = T_{ij}^k, \\ \text{und da } \sum_{k=0}^{\infty} Z_k \text{ die Gesamtzahl der Male ist, die die Kette in } S_j \text{ ist, haben wir} \\ E[\# \text{ Mal in } S_j \mid \text{Start in } S_i] = E \sum_{k=0}^{\infty} Z_k = \sum_{k=0}^{\infty} E[Z_k] = \sum_{k=0}^{\infty} T_{ij}^k \\ = (I - T)_{11}^{-1} \quad \text{(weil } \rho(T_{11}) < 1 \text{).}$$

Summiert man dies über alle Übergangszustände, so erhält man die erwartete Anzahl von Malen, die sich die Kette in *einem* Übergangszustand befindet, was der erwarteten Anzahl von Malen entspricht, bevor ein ergodischer Zustand erreicht wird. Mit anderen Worten,

$$E[\# \text{ Schritte bis zum Beginn der Absorption in } i^{\text{th}} \text{ transient state}] = (I - T)_{11}^{-1} e.$$

In praktischen Anwendungen kommt es häufig vor, dass es nur eine Übergangsklasse gibt und die ergodischen Klassen nur einzelne absorbierende Zustände sind (Zustände, die, wenn sie einmal eingetreten sind, nie wieder verlassen werden). Wenn die einzige transiente Klasse r Zustände enthält und es s absorbierende Zustände gibt

Die kanonische Form für die Übergangsmatrix lautet dann

$$\mathbf{P} = \begin{pmatrix} \begin{array}{ccc|ccc} p_{11} & \dots & p_{1r} & p_{1,r+1} & \dots & p_{1s} \\ \vdots & & \vdots & \vdots & & \vdots \\ p_{r1} & \dots & p_{rr} & p_{r,r+1} & \dots & p_{rs} \\ \hline 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{array} \end{pmatrix} \quad (15.3.7)$$

In diesem Fall ist $\mathbf{L}_{ij} = (\mathbf{I} - \mathbf{T}_{11})^{-1} \mathbf{T}_{1j}$, und die frühere Entwicklung spezialisiert sich auf die Aussage, dass

jede Absorptionskette muss schließlich einen ihrer Absorptionszustände erreichen. Die Absorptionswahrscheinlichkeiten und Absorptionszeiten sind in der folgenden Zusammenfassung enthalten.

Absorption Probabilities and Absorption Times

Für eine reduzierbare Kette, deren Übergangsmatrix $\mathbf{P} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{T}_{21} & \mathbf{T}_{22} \end{pmatrix}$ im Kanon liegt (15.3.5), seien \mathbf{T}_i und \mathbf{E}_j die i^{th} und j^{th} transienten bzw. ergodischen Klassen, und sei $\mathbf{p}^T(0)$ eine Anfangsverteilung für den Start in den verschiedenen Zuständen von \mathbf{T}_i . Wenn $(\mathbf{I} - \mathbf{T}_{11})^{-1} \mathbf{T}_{12}$ partitioniert ist als

$$(\mathbf{I} - \mathbf{T}) \mathbf{T}_{11}^{-1} \mathbf{T}_{12} = \begin{pmatrix} L_{1,1} & L_{1,2} & \dots & L_{1s} \\ \vdots & \vdots & & \vdots \\ L_{r,1} & L_{r,2} & \dots & L_{rs} \end{pmatrix},$$

dann

- $P(\text{Absorption in } \mathbf{E}_i \mid \mathbf{p}_i^T(0) = \mathbf{p}_i^T(0) \mathbf{L}_{ij}) = \mathbf{p}_i^T(0) \mathbf{L}_{ij} \mathbf{e}$,
- $P(\text{Absorption in } \mathbf{E}_j \mid \text{Start im Zustand } p^{\text{th}} \text{ von } \mathbf{T}_i) = \sum_k \mathbf{L}_{ij}^{\text{pk}} = \mathbf{L}_{ij} \mathbf{e}$,
- $E[\text{\# Schritte bis zum Beginn der Absorption in } i^{\text{th}} \text{ transientem Zustand}] = \mathbf{e}_i^{-1}$

Wenn es nur eine transiente Klasse gibt und jede ergodische Klasse ein einzelner absorbierender Zustand ist ($\mathbf{E}_j = \mathbf{S}_{r+j}$), hat \mathbf{P} die Form (15.3.7). Wenn \mathbf{S}_i und \mathbf{S}_j transiente Zustände sind, dann

- $P(\text{Aufnahme in } \mathbf{S}_{r+j} \mid \text{Beginn in } \mathbf{S}_i) = (\mathbf{I} - \mathbf{T}_{11})^{-1} \mathbf{T}_{12}$,
- $E[\text{\# steps until absorption} \mid \text{start in } \mathbf{S}_i] = (\mathbf{I} - \mathbf{T}_{11})^{-1} \mathbf{1} \mathbf{e}_i$,
- $E[\text{\# times in } \mathbf{S}_j \mid \text{start in } \mathbf{S}_i] = (\mathbf{I} - \mathbf{T}_{11})^{-1}$.

15.4 PERRONKOMPLEMENTIERUNG

Die Theorie der stochastischen Komplementierung in Abschnitt 15.5 befasst sich mit der Entwicklung von Methoden, die es ermöglichen, die stationäre Verteilung einer großen irreduziblen Markov-Kette durch Zusammenfügen stationärer Verteilungen kleinerer Ketten zu erhalten. Die Konzepte beruhen auf der Theorie der Perron-Komplementierung, die beschreibt, wie der Perron-Vektor einer großen irreduziblen Matrix durch Perron-Vektoren kleinerer Matrizen ausgedrückt werden kann.