Mathematik für InformatikerInnen 2

Frank-Olaf Schreyer

Universität des Saarlandes, SS 2020

Stochastische Matrizen, Normierte Räume

Die Themen heute sind

- Hauptsatz über stochastische Matrizen
- Satz von Perron
- Beweis des Ergodensatz
- Skalarprodukte auf $\mathbb{C}$-Vektorräumen
- Normierte Räume
- Orthonormalsysteme

Am Freitag hatten wir den Hauptsatz über stochastische Matrizen formuliert und mit dem Beweis begonnen. Er wird uns noch eine Weile beschäftigen.

Das zweite Thema heute sind Skalarprodukte auf $\mathbb{C}$-Vektorräumen und normierten Räumen

Hauptsatz über stochastische Matrizen

Satz. *Sei $P \in \mathbb{R}^{n \times n}$ eine stochastische Matrix.*

1. *P hat den Eigenwert $\lambda_1 = 1$.*
2. *Jede stationäre Verteilung w_∞ ist ein Eigenvektor zu $\lambda_1 = 1$.*
3. *Alle komplexen Eigenwerte λ_j von P haben Betrag $|\lambda_j| \leqslant 1$.*
4. *Sind für ein $k > 0$ alle Einträge von P^k nicht Null, dann ist $\lambda_1 = 1$ der einzige Eigenwert vom Betrag 1 und*
$$\dim \operatorname{Eig}(P, 1) = 1 = m(\chi_P(t), 1).$$
5. *(Ergodensatz) Unter der Voraussetzung von 4. gilt*
$$\lim_{k \to \infty} P^k = (w_\infty w_\infty \ldots w_\infty) \in \mathbb{R}^{n \times n}$$
Ohne weitere Voraussetzungen gilt:
$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P^k = (\tilde{w}_\infty \tilde{w}_\infty \ldots \tilde{w}_\infty)$$

Satz von Perron

Satz. *Sei* $A = (a_{ij}) \in \mathbb{R}^{n\times n} \subset \mathbb{C}^{n\times n}$ *eine quadratische Matrix mit nur positiven Einträgen:* $a_{ij} > 0\ \forall i, j$. *Dann hat das charakteristische Polynom* $\chi_A(t)$ *eine einfache reelle Nullstelle* λ, *welche*

$$\lambda > |\lambda_j|$$

für jede andere Nullstelle λ_j *von* $\chi_A(t)$ *erfüllt. Ferner existiert ein Eigenvektor* $x \in \mathbb{R}^n$ *zu dem Eigenwert* λ *mit ausschließlich positven Einträgen* $x_i > 0$.

Wir können diesen Satz auf stochastische Matrizen $P = (p_{ij})$ mit positiven Einträgen $p_{ij} > 0$ anwenden. Da nach dem Satz von Gerschgorin $\lambda = 1$ der betragsmäßig größte Eigenwert von P ist, stimmt er mit dem λ im Satz von Perron überein.

Korollar. *Sei $P = (p_{ij})$ eine stochastische Matrix mit positven Einträgen $p_{ij} > 0$. Dann ist $\lambda = 1$ der betragsmäßig größte Eigenwert, und $\chi_P(t)$ hat eine einfache Nullstelle in 1. Alle anderen Eigenwerte λ_j haben einen Betrag $|\lambda_j| < 1$. Der 1-dimensionale Eigenraum $\operatorname{Eig}(P, 1)$ wird von einem Vektor $x \in \mathbb{R}^n$ mit positiven Komponenten $x_i > 0$ erzeugt.*

$$w_\infty = \frac{1}{\sum_{i=1}^n x_i} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

ist die einzige stationäre Verteilung von P.

Beweis. Da $\lambda = 1$ der betragsmäßig größte Eigenwert ist, folgt mit dem Satz von Perron $m(\chi_P(t), 1) = 1$. Dass der zugehörige Eigenraum von einem Vektor x mit positiven Komponenten erzeugt wird, folgt ebenfalls. Schließlich ist w_∞ eine stationäre Verteilung, da $Pw_\infty = w_\infty$ gilt. Dies ist die einzige stationäre Verteilung, da stationäre Verteilungen Eigenvektoren von P zum Eigenwert 1 sind und $\dim \operatorname{Eig}(P, 1) = 1$ gilt. $\square$

Potenzen von Jordanblöcken

Lemma. *Sei*

$$J = J(\lambda, m) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \in \mathbb{C}^{m \times m}$$

ein Jordanblock. Dann gilt

$$J^k = \begin{pmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots & \binom{k}{m-1}\lambda^{k-m+1} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & & \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \binom{k}{1}\lambda^{k-1} \\ 0 & & & & \lambda^k \end{pmatrix}.$$

Ende Teil 1

Beweis. Schreiben $J = \lambda E + N$. Dann gilt

$$J^k = (\lambda E + N)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} N^j,$$

da λE und N kommutieren. Für $j < m$ hat N^j Einsen auf der j-ten Nebendiagonale und sonst Nullen. Für $j \geqslant m$ gilt $N^j = 0$. Die Behauptung folgt. $\square$

Beweis des Ergodensatz

Wir behandeln zunächst den Fall, dass die stochastische Matrix $P = (p_{ij})$ selbst nur positive Einträge hat.
Sei

$$SPS^{-1} = J = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & J(\lambda_2, m_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & J(\lambda_r, m_r) \end{pmatrix},$$

Dann gilt

$$\lim_{k\to\infty} J^k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix},$$

In der Tat, da alle $|\lambda_j| < 1$, gilt $\lim_{k\to\infty}(J(\lambda_s, m_s))^k = 0$.

$$\left|\binom{k}{j}\lambda_s^{k-j}\right| \leqslant \frac{k^j}{j!}|\lambda_s|^{k-j}$$

und

$$\lim_{x\to\infty} \frac{x^j}{e^{-\ln|\lambda_s|x}} = 0,$$

da $|\lambda_s| < 1 \Rightarrow \ln|\lambda_s| < 0$ gilt und die Exponentialfunktion schneller als jedes Polynom wächst.

Es folgt

$$\lim_{k\to\infty} P^k = S^{-1} \lim_{k\to\infty} J^k S = (w_\infty \; 0 \ldots 0)S = (w_\infty \; w_\infty \; \ldots w_\infty),$$

da $(S^{-1})^t P^t S^t = J^t$ impliziert, dass die erste Spalte von S^t ein Eigenvektor von P^t zum Eigenwert 1 ist. Da auch $\dim \mathrm{Eig}(P^t, 1) = 1$ gilt, ist die erste Zeile z von S also ein Vielfaches von dem Zeilenvektor $(1\; 1 \ldots 1)$. Schließlich impliziert

$$S \cdot S^{-1} = E \text{ und } \sum_{j=1}^{n} (w_\infty)_i = 1$$

dass $z = (1\; 1 \ldots 1)$ gilt.

Damit ist der Ergodensatz für den Fall, dass alle Einträge $p_{ij} > 0$ sind, bewiesen. Wissen wir dies lediglich für eine Potenz P^k, so argumentieren wir wie folgt:

Lemma. *Sei $A \in \mathbb{C}^{n \times n}$. Die Eigenwerte von A^k sind die k-ten Potenzen der Eigenwerte von A.*

Beweis. Klar, da die Jordansche Normalform von A eine obere Dreiecksmatrix ist. □

Nach den Sätzen von Gerschgorin und Perron, angewendet auf P^k, ist 1 eine einfache Nullstelle von $\chi_{P^k}(t)$, und alle andere Eigenwerte λ_j^k haben einen Betrag $|\lambda_j^k| < 1$. Für die Eigenwerte von P gilt dann $|\lambda_j| < 1$ ebenfalls, und das obige Argument greift. □

Ende Teil 2

Skalarprodukte auf $\mathbb{C}$-Vektorräumen

Definition. Sei V ein $\mathbb{C}$-Vektorraum. Ein Skalarprodukt auf V ist eine Abbildung

$$b\colon V \times V \to \mathbb{C}, (v, w) \mapsto b(v, w)$$

die folgendes erfüllt:

S1a. $b(v_1 + v_2, w) = b(v_1, w) + b(v_1, w)\ \forall v_1, v_2$ und $w \in V$

S1b. $b(v, w_1 + w_2) = b(v, w_1) + b(v, w_2)\ \forall v$ und $w_1, w_2 \in V$

S2a. $b(\lambda v, w) = \overline{\lambda} b(v, w)\ \forall v, w \in V$ and $\forall \lambda \in \mathbb{C}$

S2b. $b(v, \lambda w) = \lambda b(v, w)\ \forall v, w \in V$ and $\forall \lambda \in \mathbb{C}$

S3. $\overline{b(v, w)} = b(w, v)\ \forall v, w \in V$

S4. $b(v, v) \geqslant 0\ \forall v \in V$ und $b(v, v) = 0 \Leftrightarrow v = 0$

Eine Abbildung b, die S1a-S2b erfüllt ,nennt man auch eine **Sesquilinearform**. Häufig verwendet man auch die Notation

$$\langle v, w \rangle := b(v, w)$$

Die zugehörige **Norm** ist

$$\|v\| = \sqrt{b(v, v)} = \sqrt{\langle v, v \rangle}.$$

Beispiele

1. Das **Standardskalarprodukt** auf $\mathbb{C}^n$ ist durch
$$\langle z, w\rangle = \sum_{j=1}^{n} \overline{z}_j w_j$$
definiert. Schreiben wir $z_j = x_j + iy_j$ mit $x_j, y_j \in \mathbb{R}$, so ist
$$\|z\| = \sqrt{x_1^2 + y_1^2 + \ldots + x_n^2 + y_n^2}$$
die eulkidische Norm auf dem zugrundeliegenden $\mathbb{R}$-Vektorraum $\mathbb{C}^n = \mathbb{R}^{2n}$.
2. Sei $V = \{f : [0,1] \to \mathbb{C} \mid f \text{ ist stetig}\}$. Dann ist
$$\langle f, g\rangle = \int_0^1 \overline{f(t)} g(t) dt$$
ein Skalarprodukt auf V. $\|f\|^2 = \int_0^1 |f(t)|^2 dt \geqslant 0$ und Gleichheit gilt nur, wenn $f = 0$, da das Integral über eine nicht-negative stetige Funktion nur dann Null ist, wenn es die Nullfunktion ist.

3. Etwas allgemeiner: Für $\varphi : [0,1] \to \mathbb{R}$ eine positive stetige Funktion (eine Dichte) definiert

$$\langle f, g \rangle_\varphi = \int_0^1 \overline{f(t)} g(t) \varphi(t) dt$$

ein Skalarprodukt auf $V = \{f : [0,1] \to \mathbb{C} \mid f \text{ ist stetig}\}$.

Sei V ein endlich-dimensionaler $\mathbb{C}$-Vektorraum mit Skalarprodukt b und $\mathcal{B} = \{v_1, \ldots, v_n\}$ eine Basis. Dann erfüllt die Matrix $B = (b_{kj}) = (b(v_k, v_j))$ wegen S3

$$b_{jk} = \overline{b}_{kj}, \text{ also } B^t = \overline{B}.$$

Definition. Eine Matrix $b \in \mathbb{C}^{n \times n}$, die $B^t = \overline{B}$ erfüllt, nennt man **hermitisch**.

Man kann b aus der Matrix B zurückgewinnen. Für Vektoren $v = z_1 v_1 + \ldots z_n v_n$ und $w = w_1 v_1 + \ldots + w_n v_n$ in V gilt:

$$b(v,w) \overset{S1\&S2}{=} (\overline{z}_1 \dots \overline{z}_n) B \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \left\langle \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, B \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle$$

$$= \left\langle B \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \right\rangle, \quad \text{da } B^t \overset{S3}{=} \overline{B} \text{ gilt.}$$

Hierbei bezeichnet $\langle\ ,\ \rangle$ das Standardskalarprodukt auf $\mathbb{C}^n$.

Bemerkung. Hermitische Matrizen erfüllen obige Gleichungen, und b, definiert durch die Formel

$$b(z,w) = (\overline{z}_1 \dots \overline{z}_n) B \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix},$$

ist eine Sesquilinearform auf $\mathbb{C}^n$, die S3 erfüllt. Allerdings ist b im allgemeinen kein Skalarprodukt, da S4 $b(z,z) \geqslant 0$ für beliebige hermitische Matrizen B nicht erfüllt ist. Dazu später mehr.

Definition. Ein Skalarprodukt auf einem $\mathbb{R}$-Vektorraum V ist eine Abbildung

$$\langle\ ,\ \rangle\colon V \times V \to \mathbb{R}, (v, w) \mapsto \langle v, w\rangle,$$

welche die folgenden Bedingungen erfüllt:

S1a. $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_1, w\rangle\ \forall v_1, v_2$ und $w \in V$,

S1b. $\langle v, w_1 + w_2\rangle = \langle v, w_1\rangle + \langle v, w_2\rangle\ \forall v$ und $w_1, w_2 \in V$,

S2a. $\langle \lambda v, w\rangle = \lambda\langle v, w\rangle \forall v, w \in V$ and $\forall \lambda \in \mathbb{R}$,

S2b. $\langle v, \lambda w\rangle = \lambda\langle v, w\rangle \forall v, w \in V$ and $\forall \lambda \in \mathbb{R}$,

S3. $\langle v, w\rangle = \langle w, v\rangle \forall v, w \in V$,

S4. $\langle v, v\rangle \geqslant 0\ \forall v \in V$ und $\langle v, v\rangle = 0 \Leftrightarrow v = 0$.

Cauchy-Schwarzsche Ungleichung

Satz. *Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum mit einem Skalarprodukt $\langle\ ,\ \rangle$. Dann gilt*

$$|\langle v, w\rangle| \leqslant \|v\|\|w\| \ \forall v, w \in V$$

und Gleicheit gilt für $v \neq 0$ genau dann, wenn $w = \lambda v$ für ein Skalar λ gilt.

Beweis. Der Beweis aus der zweiten Vorlesung überträgt sich (fast) wortwörtlich. □

Korollar. *Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum mit einem Skalarprodukt $\langle\ ,\ \rangle$. Dann ist die Abbildung*

$$V \to \mathbb{R}, v \mapsto \|v\| = \sqrt{\langle v, v\rangle}$$

eine Norm auf V im Sinne folgender Definition.

Normierte Vektorräume

Definition. Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum. Eine **Norm** auf V ist eine Abbildung

$$\| \ \| \colon V \to \mathbb{R}, \ v \mapsto \|v\|,$$

die folgendes erfüllt:

N1 $\|v\| \geqslant 0 \ \forall v \in V$ und $\|v\| = 0 \iff v = 0$,

N2 $\|\lambda v\| = |\lambda| \|v\| \quad \forall v \in V$ und $\forall \lambda \in \mathbb{R}$ bzw. $\forall \lambda \in \mathbb{C}$,

N3 $\|v + w\| \leqslant \|v\| + \|w\| \ \forall v, w \in V$ (Dreiecksungleichung).

Ein Vektorraum zusammen mit einer Norm nennt man auch einen **normierten Vektorraum**.

Beweis des Korollars.

1. Wegen S3 gilt $\langle v, v \rangle = \overline{\langle v, v \rangle}$, also $\langle v, v \rangle \in \mathbb{R}$ auch bei $\mathbb{C}$-Vektorräumen mit Skalarprodukt, und $\langle v, v \rangle \geqslant 0$ gilt nach S4. $\|v\| = \sqrt{\langle v, v \rangle}$ erfüllt also N1 wegen S4.
2. N2 gilt wegen S2a & S2b: $\overline{\lambda}\lambda = |\lambda|^2$.
3. N3 beweist man genauso wie in der zweiten Vorlesung.

Beispiele für Normen

1. Auf $V = \mathbb{R}^n$ oder $\mathbb{C}^n$ haben wir folgende Normen

$$\|z\|_p := (\sum_{k=1}^{n} |z_k|^p)^{\frac{1}{p}} \text{ für } 1 \leqslant p < \infty$$

und

$$\|z\|_\infty := \max\{|z_k| \mid k = 1, \ldots n\}$$

2. Für ein abgeschlossenes beschränktes Intervall $[a, b] \subset \mathbb{R}$ haben wir auf

$$V = C^0[a, b] := \{f : [a, b] \to \mathbb{R} \mid f \text{ ist stetig } \}$$

die Normen

$$\|f\|_p := (\int_a^b |f(t)|^p dt)^{\frac{1}{p}} \text{ für } 1 \leqslant p < \infty$$

und

$$\|f\|_\infty := \sup\{|f(t)| \mid t \in [a, b]\}.$$

Dabei ist jeweils $\| \ \|_2$ die Norm, die von einem Skalarprodukt kommt.

Orthogonale Vektoren

Definition. Sei V ein $\mathbb{R}$- oder $\mathbb{C}$-Vektorraum mit einem Skalarprodukt $\langle\ ,\ \rangle$. Zwei Vektoren v, w sind **orthogonal** oder **senkrecht** zueinander, in Zeichen $v \perp w$, wenn

$$\langle v, w \rangle = 0$$

gilt. Zwei Untervektorräume $U, W \subset V$ sind orthogonal zueinander, in Zeichen $U \perp W$, wenn

$$\langle u, w \rangle = 0 \quad \forall u \in U \text{ und } \forall w \in W$$

gilt. Ist dies der Fall, dann gilt $U \cap W = 0$. Man schreibt auch

$$U \ominus W := U \oplus W = U + W \subset V$$

in diesem Fall.

Orthonormalsysteme von Vektoren

Definition. Ein **Orthogonalsystem** von Vektoren $\{v_k \mid k \in J\}$ ist eine Familie von Vektoren $\neq 0$ in einem Vektorraum mit Skalarprodukt für die

$$v_k \perp v_\ell \quad \forall k \neq \ell$$

gilt. Ein **Orthonormalsystem** von Vektoren $\{v_k \mid k \in J\}$ ist ein System für das

$$\langle v_k, v_\ell \rangle = \delta_{k\ell} \quad \forall k, \ell \in J$$

gilt, also zusätzlich die Vektoren normiert sind, d.h., Norm 1 haben.

Gram-Schmidt Verfahren

Algorithmus.
Input. Eine Familie $\{v_1, \dots, v_n\}$ von linear unabhängigen Vektoren in einem Vektorraum V mit Skalarprodukt.
Output. Eine orthonormale Familie $\{w_1, \dots, w_n\}$ von Vektoren aus V, so dass

$$\mathrm{Spann}(v_1, \dots, v_k) = \mathrm{Spann}(w_1, \dots, w_k)$$

für alle $k = 1, \dots, n$ gilt.

Bemerkung. Orthonormalsysteme sind nützlich, da man mit ihnen orthogonale Projektionen leicht berechnen kann.

Ende Teil 4