Randomisierte Algorithmen

10. Schnell mischende Markov-Ketten

Thomas Worsch

Fakultät für Informatik
Karlsruher Institut für Technologie

Wintersemester 2017/2018

Überblick

Ein schönes Buch

David A. Levin, Yuval Peres, Elizabeth L. Wilmer:
Markov Chains and Mixing Times
AMS (2008)
http://darkwing.uoregon.edu/~dlevin/MARKOV/markovmixing.pdf

Überblick

10.11 Lemma

Es sei P eine zeilenstochastische Matrix.

- Dann ist 1 ein Eigenwert von P.
- Für jeden Eigenwert λ von P gilt: $|\lambda| \leq 1$.

10.12 Beweis

- Erster Teil klar wegen $P(1, \ldots, 1)^T = (1, \ldots, 1)^T$.
- sei $\lambda \in \mathbb{C}$ Eigenwert mit $P(x_1, \ldots, x_n)^T = \lambda(x_1, \ldots, x_n)^T$ und i_0 ein Index mit $|x_{i_0}| = \max_j |x_j|$.

10.12 Beweis

- Erster Teil klar wegen $P(1,\ldots,1)^T = (1,\ldots,1)^T$.
- sei $\lambda \in \mathbb{C}$ Eigenwert mit $P(x_1,\ldots,x_n)^T = \lambda(x_1,\ldots,x_n)^T$ und i_0 ein Index mit $|x_{i_0}| = \max_j |x_j|$.
- dann

$$
\begin{aligned}
|\lambda| \cdot |x_{i_0}| &= |\lambda \cdot x_{i_0}| = \left|\sum_j P_{i_0 j} x_j\right| \\
&\le \sum_j P_{i_0 j} |x_j| \le \sum_j P_{i_0 j} |x_{i_0}| \\
&= |x_{i_0}| \sum_j P_{i_0 j} = |x_{i_0}| \ .
\end{aligned}
$$

also $|\lambda| \le 1$.

10.13 Lemma

sei P stochastische Matrix

- einer reversiblen Markov-Kette
- mit $|S| = N$ Zuständen

dann besitzt P

- nur reelle Eigenwerte
- mit $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_N > -1$.

Beweis: lineare Algebra …

10.14 Abkürzung

- $\lambda_{max} = \max\{|\lambda| \mid \lambda$ ist Eigenwert von P und $\lambda \neq 1\}$.
- Eben bewiesen:
  - für ergodische Markov-Ketten: $\lambda_{max} \leq 1$.
  - für reversible Markov-Ketten: $\lambda_{max} = \max\{|\lambda_2|, |\lambda_N|\} < 1$

10.15 Bemerkung

Falls $\lambda_N < 0$:

- Gehe von $\mathbf{P}$ zu $\mathbf{P}' = \frac{1}{2}(\mathbf{P} + \mathbf{I})$ über.
- Eigenwerte $\lambda_i' = \frac{1}{2}(\lambda_i + 1) > 0$, also
- $\lambda_{max} = \lambda_2$

Überblick

10.16 Definition

- Für zwei Verteilungen $\mathbf{p}$ und $\mathbf{q}$ ist die *totale Variationsdistanz*

$$\|\mathbf{p}-\mathbf{q}\|_{tv} = \frac{1}{2}\sum_{j\in S} |\mathbf{p}_j - \mathbf{q}_j| \ .$$

- Übung:

$$\|\mathbf{p}-\mathbf{q}\|_{tv} = \max_{T\subseteq S} |\mathbf{p}(T) - \mathbf{q}(T)|$$

wobei $\mathbf{p}(T) = \sum_{j\in T} \mathbf{p}_j$

10.17 Definition

- ergodische Markov-Kette $\mathbf{P}$, stationäre Verteilung $\mathbf{w}$;
  für alle Verteilungen $\mathbf{p}$ sei

$$\delta_{\mathbf{p}}(t) = \|\mathbf{p}\mathbf{P}^t - \mathbf{w}\|_{tv}$$

- Die *Variationsdistanz zum Zeitpunkt t* ist

$$\Delta(t) = \max_{\mathbf{p}} \delta_{\mathbf{p}}(t) \ .$$

- Übung: Maximum für einen Einheitsvektor

$$\Delta(t) = \max_{i} \|\mathbf{P}_i^t - \mathbf{w}\|_{tv}$$

  wobei $\mathbf{P}_i^t$ die i-te Zeile von $\mathbf{P}^t$ ist.
- klar: $\lim_{t\to\infty} \Delta(t) = 0$
  Wie schnell geht das?

10.18 Satz

ergodische Markov-Kette, Matrix $\mathbf{P}$, stationäre Verteilung $\mathbf{w}$

Es existieren Konstanten C und $\alpha < 1$ mit

$$\Delta(t) = \max_i \|\mathbf{P}_i^t - \mathbf{w}\|_{tv} \leq C\alpha^t$$

10.19 Beweis (1)

- $\mathbf{P}^{t_0}$ enthalte nur positive Werte
- $\mathbf{W}$ wie bisher
- dann exisitiert $0 < \delta < 1$ so, dass für alle i und j gilt: $\mathbf{P}^{t_0}_{ij} \geq \delta \mathbf{w}_j$
- $\vartheta = 1 - \delta$
- $\mathbf{Q}$ Matrix mit

  $$\mathbf{P}^{t_0} = (1 - \vartheta)\mathbf{W} + \vartheta \mathbf{Q}$$

  $\mathbf{Q}$ ist zeilenstochastisch
- es gilt:
  - $\mathbf{Q}^k \mathbf{W} = \mathbf{W}$
  - $\mathbf{W}\mathbf{P}^{t_0} = \mathbf{W}$
- Induktion lehrt: für jedes $k \in \mathbb{N}_+$ ist

  $$\mathbf{P}^{t_0 k} = (1 - \vartheta^k)\mathbf{W} + \vartheta^k \mathbf{Q}^k$$

10.19 Beweis (2)

- $\mathbf{P}^{t_0 k} = (1-\vartheta^k)\mathbf{W} + \vartheta^k \mathbf{Q}^k$
- $\mathbf{P}^{t_0 k+j} - \mathbf{W} = \vartheta^k \left(\mathbf{Q}^k \mathbf{P}^j - \mathbf{W}\right)$
- Zeile i: $\|\mathbf{P}_i^{t_0 k+j} - \mathbf{w}\|_{tv} \leq \vartheta^k$
- für beliebiges $t = t_0 k + j$ mit $k = t \operatorname{div} t_0$ und $j = t \bmod t_0$

$$
\begin{aligned}
\|\mathbf{P}_i^{t} - \mathbf{w}\|_{tv} &= \|\mathbf{P}_i^{t_0 k+j} - \mathbf{w}\|_{tv} \\
&\leq \vartheta^k \\
&= \frac{1}{\vartheta}\left(\vartheta^{1/t_0}\right)^{t_0}\left(\vartheta^{1/t_0}\right)^{t_0 k} \\
&\leq \frac{1}{\vartheta}\left(\vartheta^{1/t_0}\right)^{j+t_0 k} && \text{da } j < t_0 \text{ und } \vartheta < 1 \\
&= \frac{1}{\vartheta}\left(\vartheta^{1/t_0}\right)^{t}
\end{aligned}
$$

Mitteilung (ohne Beweis)

für reversible Markov-Kette mit stationärer Verteilung $\mathbf{w}$ sei

$$w_{min} = \min_j \mathbf{w}_j$$

Mitteilung (ohne Beweis)

für reversible Markov-Kette mit stationärer Verteilung $\mathbf{w}$ sei

$$w_{min} = \min_j \mathbf{w}_j$$

Satz

Für reversible Markov-Kette mit stationärer Verteilung $\mathbf{w}$ gilt:

$$\Delta(t) \leq \frac{\lambda_{max}^t}{w_{min}} .$$

Überblick

Definition

ergodische Markov-Kette mit stationärer Verteilung $\mathbf{w}$
für $\varepsilon > 0$ definiere ihre *ε-Konvergenzzeit*:

$$\tau(\varepsilon) = \min\{t \mid \forall t' \geq t : \Delta(t') \leq \varepsilon\}$$

Von Probleminstanzen zu Markovketten

mitunter vorliegende Situation

- gegeben: unendliche Menge von «Probleminstanzen» I
- zu jedem I wird eine Markovkette $M(I)$ konstruiert

Beispiel:

- I: Graph (V, E)
- $M(I)$: Markovkette mit
  - Zustände: Matchings in I
  - Übergangswahrscheinlichkeiten für Matching m:
    wähle zufällig gleichverteilt eine Kante $e \in E$
    - mit W.keit $1/2$ kein Übergang
    - mit W.keit $1/2$ Übergang zu

$$m' = \begin{cases} m - e, & \text{falls } e \in m \\ m + e, & \text{falls } e \notin m \text{ und } m + e \text{ Matching} \\ m, & \text{sonst} \end{cases}$$

Definition

Familie $M(I)$ von Markovketten heißt *schnell mischend*, falls die ε-Konvergenzzeit polynomiell in $|I|$ und $\ln 1/\varepsilon$ ist.

10.25 Rechnung

reversible Markov-Kette jedenfalls dann schnell mischend, wenn für ein t, das polynomiell in $|I|$ und $\log 1/\varepsilon$ ist, gilt

$$\frac{\lambda_{max}^t}{w_{min}} \leq \varepsilon$$

10.25 Rechnung (2)

- $$\begin{aligned} \left(\frac{1}{\lambda_{max}}\right)^t &\geq \frac{1}{\varepsilon w_{min}} \\ t &\geq \frac{\ln \varepsilon^{-1} + \ln w_{min}^{-1}}{\ln \lambda_{max}^{-1}} \end{aligned}$$
- Wegen $1 - x \leq \ln x^{-1}$ für $0 < x < 1$ ist dafür hinreichend:
  $$t \geq \frac{\ln \varepsilon^{-1} + \ln w_{min}^{-1}}{1 - \lambda_{max}}$$
- schnelles Mischen, falls $\ln w_{min}^{-1}$ und $1/(1 - \lambda_{max})$ polynomiell in $|I|$

10.26 Satz

Für reversible Markov-Ketten gilt:

$$\tau(\varepsilon) \leq \frac{1}{1-\lambda_{max}} \log \frac{1}{w_{min}\varepsilon}$$
$$\tau(\varepsilon) \geq \frac{1}{2(1-\lambda_{max})} \log \frac{1}{2\varepsilon}$$

ohne Beweis

Woher bekommt man λ_{max}?

10.27 Definition

- reversible Markov-Kette $M = (S, \mathbf{P})$, stationärer Verteilung $\mathbf{w}$
- sei F_M der gerichtete gewichtete Graph mit
  - Knotenmenge S und
  - Kantenmenge $E_F = \{(i,j) \mid i \neq j \wedge P_{ij} > 0\}$.
- jede Kante (i,j) gewichtet mit Zahl $c(i,j) = \mathbf{w}_i P_{ij}$.

10.28 Definition

- Für reversible Markov-Kette mit Zustandsmenge S und stationärer Verteilung $\mathbf{w}$
- definiere für jede Teilmenge $T \subseteq S$

$$\begin{array}{rrcl} \text{die } \textit{Kapazität} & C(T) & = & \sum_{i \in T} \mathbf{w}_i \\ \text{den } \textit{Fluß} & F(T) & = & \sum_{i \in T, j \notin T} \mathbf{w}_i \mathbf{P}_{ij} \\ \text{und} & \Phi(T) & = & F(T)/C(T) \end{array}$$

- Der *Leitwert* Φ der Markov-Kette ist dann

$$\Phi = \min_{T \subseteq S} \max\{\Phi(T), \Phi(S \setminus T)\} = \min_{T \subseteq S, 0 < C(T) \leq 1/2} \Phi(T) .$$

10.29 Satz

Für jede reversible Markov-Kette gilt:

$$1 - 2\Phi^2 \leq \lambda_2 \leq 1 - \frac{\Phi^2}{2} \ .$$

10.30 Korollar

Für reversible Markov-Ketten (mit $\lambda_2 = \lambda_{max}$) ist

$$\tau(\varepsilon) \leq \frac{2}{\Phi^2}(\ln \varepsilon^{-1} + \ln w_{min}^{-1})$$

10.32 Beobachtung

Für die Markov-Ketten $M_{G,\beta}$ aus Definition 9.2 gilt:

$$\Phi = \beta\mu/d \ .$$

Woher bekommt man im allgemeinen Φ?

10.33 Definition

- für Paar (i, j) von Knoten in F_M soll
- von einem „Gut" g_{ij} die Menge $\mathbf{w}_i \mathbf{w}_j$ von i nach j transportiert werden.
- gesucht: Flüsse $f_{ij} : E_F \longrightarrow \mathbb{R}_+$, so dass gilt:

$$\begin{aligned} \sum_k f_{ij}(i,k) &= \mathbf{w}_i \mathbf{w}_j \\ \text{für alle } \ell \neq i,j: \sum_k f_{ij}(k,\ell) &= \sum_m f_{ij}(\ell,m) \\ \sum_k f_{ij}(k,j) &= \mathbf{w}_i \mathbf{w}_j \end{aligned}$$

10.33 Definition (2)

- Gesamtfluss durch eine Kante e sei

$$f(e) = \sum_{i \neq j} f_{ij}(e)$$

und

- die *relative Kantenauslastung*

$$\rho(f) = \max_{e \in E_F} f(e)/c(e)$$

mit $c((i,j)) = \mathbf{w}_i P_{ij}$

10.34 Lemma

Für jede Markov-Kette mit Flüssen f_{ij} gilt:

$$\Phi \geq \frac{1}{2\rho(f)} \ .$$

10.35 Beispiel: Hyperwürfel

- reversible Markov-Kette M gemäß Definition 9.2 für Hyperwürfel H_n der Dimension n mit $\beta = 1/2$.
- $P_{ii} = 1/2$ und $P_{ij} = 1/2n$ für $i \neq j$.
- stationäre Verteilung
  - Gleichverteilung mit $w_{min} = 1/2^n$
  - also $\log 1/w_{min} = n$ polynomiell in n
- Schnelles Mischen: Suche Flüsse f_{ij}
  - die $1/2^{2n}$ transportieren und
  - kleine Kantenauslastung erzeugen.

10.35 Beispiel: Hyperwürfel (2)

- Wähle: gleichmäßige Verteilung auf alle $d!$ kürzeste Pfade zwischen i und j (d = Hammingdistanz)
- Aus Symmetriegründen gleicher Gesamtfluss auf jeder Kante.
- Für jedes d gibt es $2^n \cdot \binom{n}{d}$ Paare (i, j) mit Abstand d.
- Für ein *festes* Paar (i, j) haben alle für den Fluss f_{ij} verwendeten Pfade Länge d. Also:

$$\sum_{e \in E_F} f_{ij}(e) = d\mathbf{w}_i\mathbf{w}_j \ .$$

10.35 Beispiel (3)

$$
\begin{aligned}
f(e) &= \frac{1}{|E_F|} \sum_{d=1}^{n} 2^n \binom{n}{d} \cdot d \cdot \mathbf{w}_i \mathbf{w}_j \\
&= \frac{1}{n2^n} \sum_{d=1}^{n} 2^n \binom{n}{d} \cdot d \cdot \frac{1}{2^{2n}} \\
&= \frac{1}{2^{2n}} \sum_{d=1}^{n} \binom{n}{d} \cdot \frac{d}{n} \\
&= \frac{1}{2^{2n}} \sum_{d=1}^{n} \binom{n-1}{d-1} = \frac{1}{2^{2n}} \sum_{d=0}^{n-1} \binom{n-1}{d} \\
&= \frac{2^{n-1}}{2^{2n}} = \frac{1}{2 \cdot 2^n}
\end{aligned}
$$

10.35 Beispiel (4)

- eben: $f(e) = 1/(2 \cdot 2^n)$.
- andererseits $c(e) = \mathbf{w}_i P_{ij} = 1/(2n \cdot 2^n)$
- also $\rho(f) = \frac{1/(2 \cdot 2^n)}{1/(2n \cdot 2^n)} = n$
- wegen 10.32 daher $\Phi \geq \frac{1}{2\rho(f)} = \frac{1}{2n}$
- wegen 10.27 daher $1 - \frac{1}{2n^2} \leq \lambda_2 \leq 1 - \frac{1}{8n^2}$
- also $1/(1 - \lambda_2) \in \Theta(n^2)$
- zusammen mit $\ln w_{min}^{-1} = n$
- Korollar 10.28: Markov-Kette schnell mischend

Zusammenfassung

- bei schnellem Mischen geht es um ganze Familien von Markov-Ketten (passend zu von Probleminstanzen)
  - und nicht um eine einzelne Markov-Kette
- Kriterien, um festzustellen, ob sie schnell mischend sind