Seminararbeit

MCMC-Methoden - Der Swapping Algorithmus

Westfälische-Wilhelms-Universität Münster
Mathematisches Institut
Dozent: Prof. Dr. Löwe
Betreuung: Andrea Winkler
Verfasst von: Maximilian Mümken
Wintersemester 2012/13

Inhaltsverzeichnis

1 Einleitung

In dieser Arbeit und dem zugehörigen Vortrag geht es um eine spezielle Markov Ketten Monte-Carlo Methode für das schnelle Finden der stationären Verteilung. Angewandt wird diese auf zwei konkrete Beispiele zur Vereinfachung zum einen des Beweises und auch der Anwendung. Der betrachtete Algorithmus wird angewendet auf die zwei folgenden Beispie.

Beispiel 1. *Dieses Beispiel beschreibt eine* **Verteilung auf einem Intervall mit exponentiellem Tal**. *Sei $C > 1$ eine Konstante und J groß und $J \in \mathbb{N}$. Der Zustandsraum seien alle ganzen Zahlen im Intervall $[-J, +J]$ mit der bimodalen Verteilung*

$$\pi(x) = \frac{C^{|x|}}{Z} \quad (x = -J, -J+1, \ldots, J),$$

wobei Z die normalisierungskonstante ist. Man sieht leicht, dass für ein großes J (mit festem C) $\pi(o)$ exponentiell kleiner wird als $\pi(J)$.
Die Betrachtung wird einfacher, wenn man den Zustandsraum so einschränkt, dass nur negative Zahlen aus dem Intervall $[-(2M+1), +(2M+1)]$ verwendet werden, und er in zwei Hälften unterteilt wird.

Weiter geht es mit dem zweiten wichtigen Beispiel.

Beispiel 2. *In diesem Beispiel wird das* **Mean field Ising model** *behandelt. Sei $\beta > 0$ eine reelle Konstante und $M \in \mathbb{N}$ groß. Der Zustandsraum $\mathcal{A}$ ist wie folgt definiert: $\mathcal{A} = \mathcal{A}^{Ising} = \{-1, +1\}^M$. Das Mean field Ising model hat die Verteilung*

$$\pi(x) = \pi^{Ising}(x|\beta) = \frac{e^{\beta(\Sigma_{j=1}^{M} x_j)^2/2M}}{Z(\beta)} \quad \textit{für } x = (x_1, \ldots, x_M) \in \mathcal{A}.$$

Dabei ist $Z(\beta)$ die Normalisierungskonstante.

Beide Beispiele können auch in Form einer Gibbs Verteilung angegeben werden, sodass

$$\pi(x) = \frac{e^{-\beta H(x)}}{Z} \tag{1}$$

ist, wobei H eine bekannte Funktion und β eine nicht-negativer Parameter. In den Beispielen 1 und 2 ergibt sich dann für H und β:

$$H(x) = -|x| \text{und } \beta = ln(c) \text{ für Beispiel 1 und}$$

$$H(x) = -((\Sigma_{j=1}^{M} x_j)^2/2M) \text{ für Beispiel 2}$$

Eine sehr allgemeine MCMC-Methode ist der Metropolis Algorithmus. Zunächst hierzu eine kurze Erläuterung:
Sei π eine Wahrscheinlichkeitsverteilung auf einer endlichen Menge $\mathcal{S}$. Weiter sei K eine irreduzible, symmetrische Markov Kette auf $\mathcal{S}$. K ist die sogenannte Basis-Kette. Die neue Übergangsmatrix T der Metropolis Kette auf $\mathcal{S}$ ist dann wie folgt definiert:

$$T(x,y) = \begin{cases} K(x,y), & \text{falls } \pi(y) \geq \pi(x), y \neq x \\ K(x,y)\frac{\pi(y)}{\pi(x)}, & \text{falls} \pi(y) < \pi(x) \\ 1 - \Sigma_{z \neq x} T(x,z), & \text{falls } x = y \end{cases} \tag{2}$$

Das bedeutet ein neuer Zustand y wird mit der Wahrscheinlichkeit $\min\{1, \frac{\pi(y)}{\pi(x)}\}$ und falls y nicht der nächste Zustand ist, bleibt die Markov Kette im Zustand x. Die Metropolis Kette T ist irreduzibel zu π und π ist sogar ihre Gleichgewichts-Verteilung.
Nun kann man die Basis Ketten für die beiden Beispiele bilden und betrachten. Im ersten Fall ist die Basis Kette K die symmetrische Irrfahrt auf $\mathcal{A}$:

$$K(i,j) = \begin{cases} \frac{1}{2}, & \text{falls } |i-j| = 2 \text{ oder } i = j = \pm(2M+1) \\ 0, & \text{sonst} \end{cases} \tag{3}$$

Diese Metropolis Kette ist also eine Irrfahrt mit Verschiebung weg vom Ursprung und daher benötigt sie für ein großes M auch exponentiell lange um in die negative Hälfte zu gelangen, falls sie vom Zustand $2M+1$ ausgeht.
Im Beispiel 2 sieht wählt die Basis Kette zufällig ein $i \in \{1, \ldots, M\}$ aus und dreht dann das Vorzeichen von x_i um. Also folgt

$$K(x,y) = \begin{cases} \frac{1}{M}, & \text{falls } x, y \in \mathcal{A}^{Ising} \text{ und } ||x-y||_1 = 2 \\ 0, & \text{sonst} \end{cases} \tag{4}$$

wobei $||z|| = |z_1| + \ldots |z_M|$ für $z \in \mathbb{R}^M$ ist. Auch diese Metropolis Kette erreicht ihr Gleichgewicht exponentiel langsam in M, für feste $\beta > 1$.
Für vorigen Teil beschrieben Gibbs Verteilungen würde eine Metropolis Kette mit den gewünschten Wert β nur sehr langsam konvergieren. Andersherum hingegen würde sie für kleine Werte von β, insbesondere für $\beta = 0$ sehr schnell konvergieren. Wenn man β also als Zufallsvariable setzen würde und eine Irrfahrt auf den Werten von 0 bis β machen lassen würde und danach umgekehrt, könnte man annehmen es würde also N^2 Schritte (wenn ein "Weg" je N Schritte besitzt) geben und die "Zeit" bis zur Konvergenz könnte sich als Polynom darstellen lassen. Das ist ein Ansatz, der der Swapping Methode ähnelt.
Die Swapping Methode interpoliert eine Folge von β's, sodass sie für

$$0 = \beta_0 < \beta_1 < \cdots < \beta_N = \beta^*$$

je eine Variante des System hat. Daher besteht der Zustandsraum aus $(N+1)$-Tupeln $((x_0, \dots, x_N) \in \mathcal{A}^{N+1})$. Der Swapping Algorithmus wechselt immer zwischen 2 verschiedenen Schritten. Der eine ist das übliche Update von x_i bei β_i und der "swapping" Schritt, bei dem ein i zufällig ausgewählt und dann versucht wird x_i mit x_{i+1} zu vertauschen. Die genaue formale Beschreibung folgt im nächsten Kapitel.

2 Konvergenzrate und Hauptaussage

Eine Markov Kette mit Übergangsmatrix P und Gleichgewichtsverteilung π muss für die Konvergenz folgende Bedingung erfüllen

$$P^t(x,y) \to \pi(y) \text{ für } t \to \infty \text{ für alle } x \text{ und } y$$

Von Interesse ist allerdings eine genauere Information. Man möchte wissen wie groß dieses t sein muss, damit P^t nah genug an der Gleichgewichtsverteilung ist. Um diese "Entfernung" zu messen benötigt man zuerst ein Maß, die Total Variation Distance:

$$\Delta_x(t) := ||P_x^t - \pi|| = \tfrac{1}{2}\Sigma_y |P^t(x,y) - \pi(y)|$$

Da es, wie in Rosenthals Paper gezeigt und berichtet, viele Möglichkeiten gibt, die Konvergenz einer Markov Kette zu untersuchen. Hier wird der Vergleich mit Markov Ketten, deren Konvergenz bekannt ist oder abgeschätzt werden kann, benutzt mithilfe des *Spectral Gap*. Dafür vorher ein paar Definitionen:
Angenommen P sei die Übergangsmatrix einer irreduziblen, aperiodischen Markov Kette. Diese Kette habe den endlichen Zustandsraum $\mathcal{S}$ und man kann P also auch als Operator für Funktionen auf $\mathcal{S}$ betrachten. Dann gilt: $(Pf)(x) = \Sigma_y P(x,y)f(y)$. Wenn P ebenfalls reversibel ist zur stationären Verteilung π ($P(x,y)\pi(x) = P(y,x)\pi(y)$ für alle $x, y \in \mathcal{S}$), dann ist die Reversiblität äquivalent dazu, dass P selbstadjungiert ist auf dem Hilbertraum $\mathfrak{l}^2(\pi)$. Dazu sei $(f,g)_\pi = \Sigma f(x)g(x)\pi(x)$ das innere Produkt für reellwertige Funktionen. Die Selbstadjungiertheit von P bedeutet, dass $(Pf,g)_\pi = (f,Pg)_\pi$ für alle $f,g \in \mathfrak{l}^2(\pi)$. Damit folgt, dass P reelle Eigenwerte

$$1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{|\mathcal{S}|-1} > -1$$

und eine Orthonormalbasis aus Eigenvektoren besitzt. Im Folgenden geht es also darum eine Schranke für die total Variation Distance zwischen $P^t(x,\cdot)$ und $\pi(\cdot)$ zu finden. Diaconis und Saloff-Coste haben gezeigt, dass

$$\Delta_x(t) \leq \frac{1}{2\sqrt{\pi(x)}}\lambda_*^t \tag{5}$$

gilt, wobei $\lambda_* = \max(\lambda_1, \lambda_{|\mathcal{S}|-1})$ ist. Anhand dessen kann man $\Delta_x(t)$ über die Eigenwerte abschätzen. Dadurch dass $\tilde{P}$ wie später gezeigt wird ein positives Spektrum hat, gilt, dass $lambda_* = \lambda_1$ ist. Es ist also sinnvoll zu versuchen Grenzen für die Eigenwerte zu finden.
Nun betrachtet man die *Dirichlet-Form* von P:

$$\mathcal{E}(f,f) = \tfrac{1}{2}\Sigma_{x,y}|f(x) - f(y)|^2 P(x,y)\pi(x)$$

für reelle Funktionen f auf $\mathcal{S}$ und weiter

$$Var(f) = E_\pi(f^2) - (E_\pi(f))^2 = \tfrac{1}{2}\Sigma_{x,y}|f(x) - f(y)|^2\pi(x)\pi(y).$$

Definition 1. *Das* **Spectral Gap** *einer Übergangsmatrix P einer Markov Kette, ist*

$$Gap(P) = 1 - \lambda_1 = \inf\left\{\tfrac{\mathcal{E}(f,f)}{Var(f)} \mid f \in \mathfrak{l}^2(\pi), Var(f) \neq 0\right\}$$

Per Definition gilt, dass ein großer Wert des Spectral Gap bedeutet, dass der Eigenwert $\lambda - 1$ sehr klein ist und dadurch die rechte Seite der obigen Ungleichung schnell klein wird. Umgekehrt hat ein kleines Spectral Gap langsame Konvergenz zur Folge. Setzt man f als Indikator Funktion für die positive Hälfte, sieht man, dass in Beispiel 1 für die Metropolis Kette $Gap(T) \leq 4C/Z$ mit $Z > C^{2M+1}$ und $C > 1$ gilt. Also ist das Spectral Gap exponentiell klein in M. Dies gilt, da für den beschriebenen Fall $\mathcal{E}(f,f) = C/Z$ und $Var(f)$=1/4. Es folgt, dass die Markov Kette ihr Gleichgewicht nur sehr langsam erreicht.
Im folgenden Abschnitt wird eine genauere Beschreibung des *Swapping Algorithmus* gegeben. Man betrachtet $N+1$ Verteilungen, die jeweils einem β_i zugeordnet sind. Dies bedeutet für Beispiel 1:

$$h_i(x) = \pi^{Int}(x|C^{\beta_i}) = \tfrac{C^{\beta_i|x|}}{Z_i}, \text{ für } x \in \mathcal{A}^{Int}$$

Dabei sind die Z_i Normalisierungskonstanten und für die β_i gilt

$$\beta_i = \tfrac{i}{N}, \text{ für } i = 0, \dots, N$$

Für Beispiel 2 gilt analog

$$h_i(x) = \pi^{Ising}(x|\beta_i) \text{ für } x \in \mathcal{A}^{Ising}$$

wobei wieder für die β_i gilt

$$\beta_i = \tfrac{i}{N}\beta_*, \text{ für } i = 0, \dots, N$$

Dementsprechend ist in beiden Beispielen h_N die gewünschte Verteilung π und h_0 die Gleichverteilung auf dem Zustandsraum.
Weiter geht es mit der Übergangsmatrix T_i:

$$T_i(j,k) = \begin{cases} K(j,k)\min\{1, \frac{h_i(k)}{h_i(j)}\}, & \text{falls } j \neq k \\ 1 - \Sigma_{l \neq j} T_i(j,l), & \text{falls } j = k \end{cases}$$

Der Zustandsraum der Swapping-Kette mit $N+1$ Komponenten sei das $N+1$ fache karthesische Produkt von $\mathcal{A}$, also $\Omega := \mathcal{A}^{N+1}$.
Die Elemente aus Ω seien $\vec{x} = (x_0, \ldots, x_N)$ und für $i = 1, \ldots, N-1$ und $\vec{x} \in \Omega$ sei $(i, i+1)\vec{x}$ das Element aus Ω, bei dem die i'te und und $(i+1)$'te Komponente von $\vec{x}$ vertauscht wurden. Also

$$(i, i+1)\vec{x} = (x_o, x_1, \ldots, x_{i-1}, x_{i+1}, x_i, x_{x+2} \ldots, x_N)$$

Die Produktverteilung ψ auf Ω ist dann:

$$\psi(\vec{x}) = h_0(x_0)h_1(x_1)\ldots h_N(x_N) \tag{6}$$

Die einzelnen Komponenten von $\vec{x}$ sind also unter ψ unabhängig. ψ soll die Gleichgewichtsverteilung der Swapping Algorithmus sein. Der Algorithmus besteht aus zwei verschiedenen Markov Ketten, der "reinen Swapping Kette" Q und der Update Kette $\tilde{P}$.
Bei der Kette $\{\mathcal{Q}_k\}_{k\in\mathbb{N}}$ mit der Übergangsmatrix Q wird ein zufälliges $i \in \{1, \ldots, N-1\}$ ausgewählt und $(i, i+1)\vec{x}$ als nächster Zustand "vorgeschlagen". Dieser wird mit Wahrscheinlichkeit

$$\rho_{i,i+1}(\vec{x}) = \min\left\{1, \frac{\psi((i,i+1)\vec{x})}{\psi(\vec{x})}\right\} = \min\left\{1, \frac{h_i(x_{i+1})h_{i+1}(x_i)}{h_i(x_i)h_{i+1}(x_{i+1})}\right\} \tag{7}$$

angenommen. Für ein kleine Vereinfachung wird der Fall hinzugefügt, dass mit Wahrscheinlichkeit $\frac{1}{2}$ garnichts passiert. Genauer ist es also eine Markov Kette, die die Übergangswahrscheinlichkeiten

$$Q(\vec{x},\vec{y}) = \begin{cases} \frac{1}{2N}\rho_{i,i+1}(\vec{x}), & \text{falls } \vec{x} \neq \vec{y} \text{ und } \vec{y} = (i,i+1)\vec{x} \\ 0, & \text{falls } \vec{x} \neq \vec{y} \text{ und } \vec{y} \neq (i,i+1)\vec{x} \\ 1 - \Sigma_{\vec{z}\neq\vec{x}} Q(\vec{x},\vec{z}), & \text{falls } \vec{x} = \vec{y} \end{cases} \quad \text{hat.}$$

Damit ist die Übergangsmatrix Q reversibel zu ψ und es gilt, dass $Q(\vec{x},\vec{x}) \geq \frac{1}{2}$ für jedes $\vec{x} \in \Omega$.
Als nächstes wird die Update Kette beschrieben. Das Update wird durch die Produktkette $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ durchgeführt. Diese Markov Kette wählt zufällig eine Komponente von $\vec{x}$ aus und auf diesen wird dann das Update ausgeführt. Hier soll ebenfalls mit Wahrscheinlicheit $\frac{1}{2}$ nichts passieren. Für die Übergangsmatrix gilt dann

$$\tilde{P}(\vec{x},\vec{y}) =$$

$$\tfrac{1}{2}\mathcal{E}(\vec{x},\vec{y}) + \tfrac{1}{2(N+1)}\Sigma_{i=0}^{N}\mathcal{E}(x_0,y_0)\cdots\mathcal{E}(x_{i-1},y_{i-1})T_i(x_i,y_i)\mathcal{E}(x_{i+1},y_{i+1})\cdots\mathcal{E}(x_N,y_N),$$

wobei $\vec{x}$ und $\vec{y}$ aus Ω sind. Weiter gilt $\mathcal{E}(u,v) = 1$, falls $u = v$ und 0 sonst, und T_i ist die Metropolis Kette wie vorher definiert.
Der Swapping Algorithmus läuft nun so ab, dass die beiden Schritte (Vertauschung und Update) abwechselnd ausgeführt werden. Dafür gibt es natürlich verschiedene Möglichkeiten. Hier wird die Variante gewählt $\mathcal{QPQ}$, denn sie hat den Vorteil, dass diese Kette reversibel ist.
Die Hauptaussage dieser Arbeit besteht darin eine untere Schranke für den Spectral Gap der Swapping Kette $\mathcal{QPQ}$ zu bestimmen. Noch spezieller geht es um die beiden Anfangs eingeführten Beispiele dabei. Diese untere Schranke wird in Abhängigkeit einer Konstanten $\Theta \in (0,1)$ angegeben. Für die beiden Beispiele ist Θ dann

$$\Theta = \begin{cases} C^{-2}, & \text{in Beispiel 1} \\ e^{-\beta^*/2}, & \text{in Beispiel 2} \end{cases} \tag{8}$$

Über Θ lässt sich dann eine Aussage treffen.

Lemma 1. Θ *sei wie definiert wie vorher. Dann gilt*

$$\rho_{i,i+1}(\vec{x}) \geq \Theta^{M/N}$$

für alle $i \in \{0,\ldots,N-1\}$ *und* $\vec{x} \in \Omega$.

Beweis. Der Beweis wird hier nur für Beispiel 2 gezeigt. Für Beispiel 1 verfahre ähnlich.
Es gilt $\beta_i = i\beta^*/N$.
Weiter sei $G(x) = \frac{(\Sigma_{j=1}^M x_j)^2}{2M}$ für $x \in \mathcal{A}$.
Dann gilt $0 \leq G(x) \leq M/2$. Damit folgt

$$\begin{aligned} \rho_{i,i+1}(\vec{x}) = & \min\left\{1, \tfrac{h_{i+1}(x_i)h_i(x_{i+1}}{h_i(x_i)h_{i+1}(x_{i+1})}\right\} & (9) \\ = & \min\left\{1, \tfrac{exp((i+1)\beta^* G(x_i)/N)exp(i\beta^* G(x_{i+1})/N)}{exp(i\beta^* G(x_i)/N)exp((i+1)\beta^* G(x_{i+1})/N)}\right\} & (10) \\ = & \min\{1, exp(\beta^* G(x_i)/N - \beta^* G(x_{i+1})/N)\} & (11) \\ \geq & exp(-\beta^* M/2N) = \Theta. & (12) \end{aligned}$$

Daraus folgt die Aussage. □

Nun folgt die Hauptaussage, die es später zu beweisen gilt.

Theorem 1 (Hauptaussage). *Der Spectral Gap der Swapping Kette* $\mathcal{QPQ}$ *hat folgende untere Schranke:*

$$Gap(Q\tilde{P}Q) \geq \begin{cases} \frac{\Theta^{2M/N}}{384M^2(N+1)^6}, & \textit{in Beispiel 1} \\ \frac{\Theta^{2M/N+8}}{768M^3(N+1)^6}, & \textit{in Beispiel 2} \end{cases} \tag{13}$$

3 Nützliche Nebenaussagen

Um im nächsten Kapitel den Beweis von Theorem 1 genau zu führen, benötigt man vorher ein paar Aussagen, die im Beweis Anwendung finden. In diesem Abschnitt geht es darum diese Nebenaussagen zu formulieren und zu beweisen.

Lemma 2. *Sei P eine reversible Markov Kette mit $P(x,x) \geq \frac{1}{2}$ für alle x aus dem endlichen Zustandsraum $\mathcal{S}$. Dann ist P ein positiver Operator.*

Beweis. Man kann P umschreiben als $(I+P^{\#})/2$, wobei $P^{\#}$ eine reversible Übergangsmatrix ist. Dann folgt die Behauptung $\square$

Das nächste Lemma behandelt Produktketten, wie $\tilde{P}$ zum Beispiel.

Lemma 3. *Für jedes $i = 1, \ldots, m$ sei K_i eine reversible Markov Kette auf endlichem Zustandsraum $\mathcal{T}_i$. $\mathcal{K}$ sei dann die Produktkette auf $\Pi_{i=1}^m \mathcal{T}_i$ mit Übergangsmatrix K der Form*

$$K = \frac{1}{m} \Sigma_{i=1}^m I \otimes \cdots \otimes I \otimes K_i \otimes I \otimes \cdots \otimes I.$$

Dann gilt:

$$Gap(K) = \frac{1}{m} \min\{Gap(K_i), i = 1, \ldots, m\}$$

Das nächste Lemma ist von wichtiger Bedeutung im Verlauf des Beweises von Theorem 2. Es wird eine Aussage getroffen über den Zusammenhang zwischen dem Vergleich von Dirichlet-Formen zweier Ketten und den Spectral Gaps.

Lemma 4. *Seien (K, π) und(K', π') die Übergangsmatrizen zweier reversibler Markov Ketten auf dem gleichen endlichen Zustandsraum. Ihre Dirichlet-Formen seien $\mathcal{E}$ und $\mathcal{E}'$. Weiter gebe es Konstanten $A, a > 0$, sodass*

$$\mathcal{E}' \leq A\mathcal{E} \text{ und } a\pi \leq \pi'$$

Dann gilt:

$$Gap(K') \leq \frac{A}{a} Gap(K)$$

Der Beweis, ist nachzulesen "Logarithmic Sobolev inequalities for finite Markov chains" von Diaconis und Saloff-Coste.

Das nächste Lemma stellt eine Beziehung her zwischen einer Markov Kette P und der positiven ganzzahligen Potenz P^m.

Lemma 5. *Für eine reversible, endliche Markov Kette P gilt*

$$Gap(P) \geq \frac{1}{m} Gap(P^m) \text{ für alle } m \in \mathbb{N}$$

Beweis. Sei λ_1 der zweitgrößte Eigenwert von P und v_1 der zugehörige Eigenvektor, so dass $Pv_1 = \lambda_1 v_1$ ist. Dann ist $P^m v_1 = \lambda_1^m v_1$. Also kann der zweitgrößte Eigenwert von P^m nicht kleiner sein als λ_1^m.
Daher gilt nach Induktion $Gap(P^m) \leq 1 - \lambda_1^m \leq m(1-\lambda_1) = mGap(P)$
Das ist äquivalent zu $Gap(P) \geq \frac{1}{m} Gap(P^m)$. □

Für das nächste Lemma müssen einige Voraussetzungen erfüllt sein. Zunächst sei wieder $(\cdot,\cdot)_\pi$ inneres Produkt auf dem Hilbertraum $\mathfrak{l}^2(\pi)$, wobei π ein Wahrscheinlichkeitsmaß auf dem Zustandsraum $\mathcal{S}$. Dann sei $\mathbf{1}^\perp$ das orthogonale Komplement der Konstanten Funktionen, das heißt

$$\mathbf{1}^\perp := \{f \in \mathfrak{l}^2(\pi) \mid (f,1)_\pi = 0\}$$

A und B seien selbstadjungierte Operatoren in $\mathfrak{l}^2(\pi)$, die folgenden Bedingungen genügen:

i) A und B haben 1 als größten Eigenwert,

ii) A ins invariant auf $\mathbf{1}^\perp$ (d.h. $Af \in \mathbf{1}^\perp$ falls $f \in \mathbf{1}^\perp$),

iii) $||A|| \leq 1$ und

iv) B ist positiv (d.h. $(Bf,f)_\pi \geq 0$ für alle $f \in \mathfrak{l}^2(\pi)$)

Lemma 6. *Unter den obigen Voraussetzungen gilt:*

$$Gap(ABA) \geq Gap(B)$$

Beweis. Sei $f \in \mathbf{1}^\perp$, dann gilt nach ii), dass auch $Af \in \mathbf{1}^\perp$ ist. Per Definition des Spectral Gap erhält man:

$$\begin{aligned}
Gap(ABA) &= \inf_{f\in\mathbf{1}^\perp} \tfrac{(f,(I-ABA)f)_\pi}{(f,f)_\pi} && (14)\\
&= \inf_{f\in\mathbf{1}^\perp} \tfrac{(f,f)_\pi-(f,ABAf)_\pi}{(f,f)_\pi} && (15)\\
&= \inf_{f\in\mathbf{1}^\perp} \left\{1 - \tfrac{(Af,BAf)_\pi}{(f,f)_\pi}\right\} && (16)\\
&\overset{\text{iv)}}{\geq} \inf_{f\in\mathbf{1}^\perp} \left\{1 - \tfrac{(Af,BAf)_\pi}{(Af,Af)_\pi}\right\} && (17)\\
&= \inf_{f\in\mathbf{1}^\perp, g=Af} \left\{1 - \tfrac{(g,Bg)_\pi}{(g,g)_\pi}\right\} && (18)\\
&\geq \inf_{g\in\mathbf{1}^\perp} \left\{1 - \tfrac{(g,Bg)_\pi}{(g,g)_\pi}\right\} && (19)\\
&= Gap(B) && (20)
\end{aligned}$$

□

4 Beweis der Hauptaussage

Der Beweis der Hauptaussage über die Konvergenz der Markov Kette basiert grundlegend auf der Zerlegung der einzelnen Markov Ketten und der Anwendung eines Theorems von Caracciolo, Pelisetto und Sokal. Nachdem man die Zustandsräume und die Markov Ketten zerlegt hat, kann man Aussagen über diese Teile treffen und weiter zeigen, dass entsprechende Grenzen auch für die Zusammengesetzten Ketten ihre Gültigkeit nicht verlieren. Der Beweis teilt sich somit in 4 Schritte nachdem das Theorem vorgestellt wurde. Der Beweis dazu ist nachzulesen in "Two remarks on simulated tempering" der genannten Autoren.
Nun folgt eine genauere Bechreibung der Zerlegung, die dann im nächsten Theorem ihre Anwenung findet. Sei θ eine Wahrscheinlichkeitsverteilung auf einem endlichen Zustandsraum $\mathcal{S}$ und P eine Übergangsmatrix, die reversibel ist zu θ. Der Zustandsraum besitze eine Partition aus m Teilmengen, sodass

$$\mathcal{S} = \bigcup_{i=1}^{m} \mathcal{S}_i \text{ wobei } \mathcal{S}_i \cap \mathcal{S}_j = \emptyset \text{ falls } i \neq j$$

Definition 2. *Die* **Einschränkung** *P_i von P auf $\mathcal{S}_i$ für jedes $i = 1, \ldots N$ ist*

$$P_i(x, B) = P(x, B) + 1_{\{x \in B\}} P(x, \mathcal{S} \backslash \mathcal{S}_i) \text{ für } x \in \mathcal{S}_i \text{ und } B \subset \mathcal{S}_i.$$

Dann ist auch P_i reversibel zu der Einschränkung $\theta_i(x) = \frac{\theta(x)}{\theta(\int_i)}$ von θ auf $\mathcal{S}_i$. Sei $b_i = \theta(\mathcal{S}_i) = \Sigma_{x \in \mathcal{S}_i} \theta(x)$. Dann ist $(b_1, \ldots, b_m)$ eine Wahrscheinlichkeitsverteilung auf $\{1, \ldots, m\}$.

Definition 3. *Gegeben sei eine weitere Übergangsmatrix L, die reversibel zu θ sei. Dann ist die* **kumulierte Übergangsmatrix** *$\overline{L}$ definieren als*

$$\overline{L}(i, j) := \frac{1}{b_i} \Sigma_{y \in \mathcal{S}_j} \Sigma_{x \in \mathcal{S}_i} \theta(x) L(x, y) \ (i, j = 1, \ldots, m)$$

Diese Markov Kette ist dann reversibel zu $b(= (b_1, \ldots, m))$.

Theorem 2 (Caracciolo-Pelissetto-Sokal)**.** *Sei $\mathcal{L}$ ein Operator mit positivem Spektrum und $\mathcal{L}^{1/2}$ so gegeben, dass $\mathcal{L}^{1/2}\mathcal{L}^{1/2} = \mathcal{L}$ ist. Dann gilt*

$$Gap(\mathcal{L}^{1/2} \mathcal{P} \mathcal{L}^{1/2}) \geq Gap(\overline{\mathcal{L}}) \min_{1 \leq i \leq m} Gap(\mathcal{P}_i) \tag{21}$$

Im Beweis des Haupttheorems geht es nun darum $\mathcal{L}$ und $\mathcal{P}$ passend zu wählen, um eine Anwendung des zweiten Theorems zu ermöglichen. Dazu werden zunächst die Zustandsräume symmetrisch in je eine "positive" und "negative" Hälfte zerlegt.
In Beispiel 1 gilt dann $\mathcal{A} = \mathcal{A}^{Int}$ und weiter $A_+ = \mathcal{A}_+^{Int} = \{x \in \mathcal{A}^{Int} \mid x > 0\}$.
Für das zweite Beispiel folgt entsprechend $\mathcal{A} = \mathcal{A}^{Ising}$ und $\mathcal{A}_+ = \mathcal{A}_+^{Ising} = \{x \in \mathcal{A}^{Ising} \mid \Sigma_{i=1}^{M} x_i > 0\}$.

Für beide Beispiel ist dann $\mathcal{A}_- = \mathcal{A} \backslash \mathcal{A}_+$ und wegen der Symmetrie gilt $\pi(\mathcal{A}_- = \frac{1}{2} = \mathcal{A}_+$.

Im **ersten Schritt** geht es darum $Gap(Q\tilde{P}Q)$ abzuschätzen um technisch die Anwendung von Theorem 2 vorzubereiten. $Q\tilde{P}Q$ ist eine irreduzible, aperiodische und reversible endliche Markov Kette. Daher gilt

$$\begin{aligned} Gap(Q\tilde{P}Q) \geq & \ \tfrac{1}{3} Gap(Q\tilde{P}QQ\tilde{P}QQ\tilde{P}Q) \text{ nach Lemma 5} && (22)\\ = & \ \tfrac{1}{3} Gap(Q\tilde{P}Q^{\frac{1}{2}}(Q^{\frac{1}{2}}Q\tilde{P}QQ^{\frac{1}{2}})Q^{\frac{1}{2}}\tilde{P}Q) && (23)\\ \geq & \ \tfrac{1}{3} Gap(Q^{\frac{1}{2}}(Q\tilde{P}Q)Q^{\frac{1}{2}}) \text{ nach Lemma 6} && (24) \end{aligned}$$

Später wird klar, dass man die zusätzlichen $Q^{\frac{1}{2}}$ noch benötigt und sie deshalb nicht auch gestrichen wurden.

Inhalt des **zweiten Schritts** wird sein den Zustandsraum Ω nach Anzahl der positiven Elemente unter den $x_1, \ldots, x_N$ zu zerlegen und dann Theorem 2 anzuwenden. Sei $\tilde{\Omega} = \{-,+\}^N$ und $\vec{x} = (x_0, \ldots, x_N) \in \Omega = \mathcal{A}^{N+1}$. Dann definiere $sgn(\vec{x})$ als den vektor $(v_1, v_N) \in \tilde{\Omega}$, sodass

$$v_i = \begin{cases} +, & \text{falls } x_i \in \mathcal{A}_+ \\ -, & \text{falls } x_i \in \mathcal{A}_- \end{cases} \quad (1 \leq i \leq N) \tag{25}$$

Anschaulich ist $sgn(\vec{x})$ der Vektor, der die Vorzeichen der Elemente 1 bis N beinhaltet.
Weiter sei für $k \in \{0, \ldots, N\}$ $\tilde{\Omega}_k = \{v \in \Omega \mid v \text{ hat genau } k \ +' s\}$
Dann ist $\{\tilde{\Omega}_k, 0 \leq k \leq N\}$ eine Partiton von $\tilde{\Omega}$, denn es gilt

$$\tilde{\Omega} = \bigcup_{k=0}^{N} \tilde{\Omega}_k \text{ und } \tilde{\Omega}_k \cap \tilde{\Omega}_j = \emptyset \text{ falls } j \neq k.$$

Analog lässt sich dann eine Zerlegung von Ω in Ω_k's definieren,

$$\Omega_k = \{\vec{x} \in \Omega \mid sgn(\vec{x}) \in \tilde{\Omega}_k\},$$

und man erhält eine Zerlegung von Ω ($\{\Omega_k, 0 \leq k \leq N\}$). Nun definiere wie vorher Beschrieben die kumulierte Übergangsmatrix $\overline{Q}$. Weil Q ein positives Spektrum hat, kann hier Theorem 2 angewendet werden:

$$Gap(Q^{\frac{1}{2}}(Q\tilde{P}Q)Q^{\frac{1}{2}}) \geq Gap(\overline{Q}) \cdot \min_{0 \leq k \leq N} Gap((Q\tilde{P}Q)|_{\Omega_k)} \tag{26}$$

Der **dritte Schritt** soll nun eine untere Schranke für $Gap((Q\tilde{P}Q)|_{\Omega_k})$ liefern. Diese Kette ist von recht ungewöhnlicher Gestalt und es ist nicht einfach diese abzuschätzen. Um nicht ein Produkt von 3 Ketten, welches dann auf Ω_k beschränkt

wird, betrachten zu müssen, werden die 3 Ketten jeweils auf Ω_k beschränkt und dann betrachtet.
Ziel ist es dann also ein untere Schranke für das Produkt der 3 beschränkten Ketten zu bestimmen. Hierbei sei $\psi(\vec{x}) = \Pi_{i=0}^N h_i(x_i)$ wie vorher definiert und π_k dann die normalisierte Einschränkung auf Ω_k

$$\psi_k(A) = \tfrac{\psi(A\cap\Omega_k)}{b_k}, \quad A \subset \Omega,$$

wobei $b_k = \psi(\Omega_k)$ die Normalisierungskonstante ist. Weiter sei $\tilde{P}_k$ die Einschränkung der Update-Kette $\tilde{P}$ auf Ω_k:

$$\tilde{P}_k(\vec{x}, A) = \tilde{P}(\vec{x}, A \cap \Omega_k) + 1_{\{\vec{x}\in A\}}\tilde{P}(\vec{x}, \Omega\backslash\Omega_k), \text{ für } \vec{x} \in \Omega_k,\ A \subset \Omega.$$

Nun fehlt nur noch die Einschränkung von Q auf Ω_k:

$$Q_k(\vec{x}, A) = Q(\vec{x}, A \cap \Omega_k) + 1_{\{\vec{x}\in A\}}Q(\vec{x}, \Omega\backslash\Omega_k), \text{ für } \vec{x} \in \Omega_k,\ A \subset \Omega.$$

Weil $\tilde{P}$ und Q jeweils reversibel zu ψ sind, sind es auch ihre Einschränkungen $\tilde{P}_k$ und Q_k.
Nun kann man beginnen $Gap((Q\tilde{P}Q)|_{\Omega_k})$ Stück für Stück abzuschätzen:
Zunächst gilt:

$$Q(\vec{x}, \vec{y}) \geq \tfrac{1}{2}Q_k(\vec{x}, \vec{y}) \text{ und } \tilde{P}(\vec{x}, \vec{y}) \geq \tfrac{1}{2}\tilde{P}_k(\vec{x}, \vec{y}) \text{ für alle } \vec{x}, \vec{y} \in \Omega_k$$

Diese Ungleicheiten sind klar, falls $\vec{x} \neq \vec{y}$ ist, denn dann gilt $Q(\vec{x}, \vec{y}) = Q_k(\vec{x}, \vec{y})$. Falls $\vec{x} = \vec{y}$ ist, denn dann ist $Q(\vec{x}, \vec{y}) \geq \frac{1}{2}$. Das gleiche Argument gilt ebenfalls für $\tilde{P}_k$ und $\tilde{P}$. Weiter ist dann für $\vec{x}, \vec{y} \in \Omega_k$

$$\begin{aligned}
(Q\tilde{P}Q)|_{\Omega_k})(\vec{x}, \vec{y}) &= (Q\tilde{P}Q)(\vec{x}, \vec{y}) && (27)\\
&= \Sigma_{\vec{z},\vec{w}\in\Omega}Q(\vec{x}, \vec{z})\tilde{P}(\vec{z}, \vec{w})Q(\vec{w}, \vec{y}) && (28)\\
&\geq \Sigma_{\vec{z},\vec{w}\in\Omega_k}Q(\vec{x}, \vec{z})\tilde{P}(\vec{z}, \vec{w})Q(\vec{w}, \vec{y}) && (29)\\
&\geq \tfrac{1}{8}\Sigma_{\vec{z},\vec{w}\in\Omega_k}Q_k(\vec{x}, \vec{z})\tilde{P}_k(\vec{z}, \vec{w})Q_k(\vec{w}, \vec{y}) && (30)\\
&= \tfrac{1}{8}(Q_k\tilde{P}_kQ_k)(\vec{x}, \vec{y}). && (31)
\end{aligned}$$

Da sowohl $(Q\tilde{P}Q)|_{\Omega_k}$ als auch $Q_kP_kQ_k$ reversibel zu π_k sind, gilt die gleiche Abschätzung auch für die Dirchiletformen. Dann folgt nach Lemma 4, dass die Ungleichung auch für die Spectal Gaps gilt, also folgt

$$Gap((Q\tilde{P}Q)|_{\Omega_k}) \geq \frac{1}{8}Gap(Q_k\tilde{P}_kQ_k) \qquad (32)$$

Im **vierten Schritt** wird noch eine weitere Zerlegung durchgeführt. Jedes Ω_k wird noch einmal zerlegt. Per Definition hat jedes Element in Ω_k genau k positive Komponenten. Nun geht es darum, welche der Komponenten positiv sind. Definiere also für $\sigma \in \tilde{\Omega}_k$

$$\Omega_\sigma = \{\vec{x} \in \Omega \mid sgn(\vec{x}) = \sigma\}.$$

Dann sei $\tilde{P}_\sigma$ die Einschränkung von $\tilde{P}_k$ auf Ω_σ, so dass

$$\tilde{P}_\sigma(\vec{x}, A) = \tilde{P}_k(\vec{x}, A \cap \Omega_\sigma) + 1_{\{\vec{x} \in A\}} \tilde{P}_k(\vec{x}, \Omega \backslash \Omega_\sigma)$$

ist für $\vec{x} \in \Omega_\sigma$ und $A \subset \Omega_k$. Dann gilt natürlich $\tilde{P}_\sigma(\vec{x}, \vec{y}) = \tilde{P}_k(\vec{x}, \vec{y})$ solange $\vec{x}$ und $\vec{y}$ in Ω_σ liegen und nicht gleich sind. Für jedes $k \in \{0, \ldots, N\}$ erhält man dann eine Partition

$$\Omega_k = \bigcup_{\sigma \in \tilde{\Omega}_k} \Omega_\sigma.$$

Wie im zweiten Schritt wird hier auch wieder eine kumulierte Übergangsmatrix $\overline{Q_k}$ eingeführt. Die zu $\overline{Q_k}$ gehörigen Markov Kette simuliert die Bewegung von positiven Koordinaten, die mit negativen vertauscht werden. Man hat also k positive Koordinaten, die alle jeweils eine Irrfahrt auf $\{1, \ldots, N\}$ absolvieren.
Dann lässt sich aufgrund der Positivität des Spektrums von Q_k wieder Lemma 6 anwenden:

$$Gap(Q_k \tilde{P}_k Q_k) \geq Gap((Q_k)^{\frac{1}{2}} \tilde{P}_k (Q_k)^{\frac{1}{2}}) \tag{33}$$

Darauf lässt sich nun wieder Theorem 2 anwenden:

$$Gap((Q_k)^{\frac{1}{2}} \tilde{P}_k (Q_k)^{\frac{1}{2}}) \geq Gap(\overline{Q_k}) \cdot \min_{\sigma \in \Omega_k} Gap(\tilde{P}_\sigma) \tag{34}$$

für jedes $k \in \{0, \ldots, N\}$. Abschließend lassen sich alle Abschätzungen aus den vier Schritten zusammenfassen und man erhält:

$$\begin{aligned}
& Gap(Q\tilde{P}Q) && (35)\\
\geq\ & \tfrac{1}{3} Gap(Q^{\frac{1}{2}}(Q\tilde{P}Q)Q^{\frac{1}{2}}) \text{ (nach Schritt 1)} && (36)\\
\geq\ & \tfrac{1}{3} Gap(\overline{Q}) \cdot \min\nolimits_{0 \leq k \leq N} Gap((Q\tilde{P}Q)|_{\Omega_k}) \text{ (nach Schritt 2)} && (37)\\
\geq\ & \tfrac{1}{24} Gap(\overline{Q}) \cdot \min\nolimits_{0 \leq k \leq N} Gap(Q_k \tilde{P}_k Q_k) \text{ (nach Schritt 3)} && (38)\\
\geq\ & \tfrac{1}{24} Gap(\overline{Q}) \cdot \min\nolimits_{0 \leq k \leq N} Gap(\overline{Q_k}) \cdot \min\nolimits_{\sigma \in \Omega_k, 0 \leq k \leq N} Gap(\tilde{P}_\sigma) \text{ (nach Schritt 4)} && (39)
\end{aligned}$$

Nun fehlen lediglich noch Abschätzungen bzw. untere Schranken für die einzelnen Teile der letzten Ungleichung. Im einzelnen wird in Proposition 10 aus Kapitel 5 von "On the Swapping Algorithm"

$$Gap(\overline{Q}) \geq \frac{1}{4N^2} \Theta^{M/N}, \tag{40}$$

in Proposition 11 aus Kapitel 6

$$Gap(\overline{Q_k}) \geq \frac{4}{N^3} \Theta^{M/N}, \tag{41}$$

und in Proposition 12 aus Kapitel 7

$$Gap(Q\tilde{P}Q) \geq \begin{cases} \frac{1}{16M^2(N+1)}, & \text{in Beispiel 1} \\ \frac{e^{-4\beta^*}}{32M^3(N+1)}, & \text{in Beispiel 2} \end{cases} \tag{42}$$

gezeigt. Wenn man diese unteren Schranken in die nach Schritt 4 gezeigte Ungleichung einsetzt, ergibt sich der Beweis von Theorem 1, der Hauptaussage. Die Beweise der oben genannten Propositionen sind sehr lang und teilweise sehr technisch. Daher wird im Rahmen dieser Seminararbeit auch auch im Vortrag nicht näher auf sie eingegangen.

5 Quellen

- Neal Madras, Zhongrong Zheng, "On the Swapping Algorithm", Department of Mathematics and Statistics, York University, 2002, S. 66-97
- S. Caracciolo, A. Pelissetto, und A. D. Sokal, "Two remarks on simulated tempering", unpublished manuscript, 1992
- P. Diaconis und L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains", Ann Appl Probab 6, 1996, S.695-750