Publ. I.R.M.A. Strasbourg, 1988, 361/S-19
Actes 19[e] Séminaire Lotharingien de Combinatoire, p. 23-29

KONVERGENZVERHALTEN GEWISSER MARKOV-PROZESSE

VON

ULRICH FAIGLE[†]

1. Simulated Annealing

In der diskreten Optimierung betrachtet man eine (endliche) Menge S, deren Elemente durch eine reellwertige Funktion $c: S \to \mathbb{R}$ gewichtet sind. Das Problem besteht darin, ein Element $s^* \in S$ mit der Eigenschaft

$$c(s^*) = \min\{c(s): s \in S\}$$

zu bestimmen.

Probabilistische Lösungsmethoden, die unter dem Begriff Simulated Annealing (= simulierte Temperierung) zusammengefasst werden, wollen die Minimierung von Energiezuständen in der Thermodynamik modellieren und basieren auf folgendem Iterationsprinzip.

Ausgehend vom Element $s \in S$ wird ein Element $t \in S$ mit einer Wahrscheinlichkeit a_{st} erzeugt. Der tatsächliche Übergang von s nach t erfolgt dann mit einer Wahrscheinlichkeit $p_{st}(T)$, wobei

$$p_{st}(T) = \begin{cases} 1 & \text{falls } c(t) \le c(s) \\ e^{-(c(t)-c(s))/T} & \text{falls } c(t) > c(s). \end{cases}$$

Dabei ist $T > 0$ ein Parameter ("Temperatur"), der im Verlauf des Algorithmus auf 0 abgesenkt wird.

[†] Fakultät für Angewandte Mathematik, Universität Twente, Enschede, Niederlande.

Die Idee dieser Methode geht auf Metropolis et al. [1953] zurück, die beobachtet hatten, dass bei symmetrischen Auswahlwahrscheinlichkeiten (d.h. $a_{st} = a_{ts}$) und festem T der Iterationsprozess im Limes ein Element $s \in S$ gemäss einer Boltzmann-Verteilung erzeugt ($c(s)$ ist in diesem Sinn die Energie von s). Die Wiederentdeckung und Anwendung auf komplexe diskrete Optimierungsprobleme erfolgte durch Kirkpatrick et al [1983] und Černý [1985].

Die bisherige praktische Erfahrung mit Simulated Annealing zeigt, dass der Erfolg sehr implementationsabhängig ist. Der theoretische Hintergrund für gute Implementierungen fehlt jedoch noch fast vollständig. In dieser Arbeit sollen kurz einige neue Resultate über die Konvergenz von Simulated-Annealing-Verfahren diskutiert werden. Dabei erweist es sich als bequem, Simulated Annealing im Rahmen von Markov-Prozessen zu interpretieren.

2. Markov-Prozesse

Es sei $A = (a_{st})$ eine stochastische Matrix, deren Zeilen und Spalten mit S indiziert sind. $t \in S$ ist ein *Nachbar* von $s \in S$, falls

$$a_{st} > 0.$$

Somit kann A der Nachbarschaftsgraph $G(A) = (S,E)$ zugeordnet werden mit

$$(s,t) \in E \iff a_{st} > 0.$$

Es soll hier grundsätzlich angenommen werden, dass G(A) stark zusammenhängend und folglich der durch A definierte (homogene) Markov-Prozess irreduzibel ist.

Unter Berücksichtigung der Akzeptanzwahrscheinlichkeiten $p_{st}(T)$ kann ein Iterationsschritt in einem Simulated-Annealing-Verfahren somit durch die stochastische Matrix $B(T) = \left(b_{st}(T)\right)$ beschrieben werden, wobei

$$b_{st}(T) = \begin{cases} p_{st}(T) \cdot a_{st} & \text{falls } s \neq t \\ 1 - \sum\limits_{r \neq s} b_{st}(T) & \text{falls } s = t. \end{cases}$$

Offenbar gilt $G[B(T)] = G(A)$ für alle $T > 0$. Falls T genügend klein ist, besitzt B(T) positive Diagonalelemente. Somit ist Konvergenz

$$\lim_{k \to \infty} b_{st}^{(k)}(T) = \pi_t(T)$$

für alle $t \in S$ (unabhängig von s!) gegen die Grenzverteilung $\pi(T) = (\ldots, \pi_t(T), \ldots)$ mit

$$\sum_{t \in S} \pi_t(T) = 1$$

gewährleistet (vgl. z.B. Seneta [1981]). Hierbei bezeichnen die $b_{st}^{(k)}(T)$ die Elemente in der k-ten Matrixpotenz $B^k(T)$, also die Wahrscheinlichkeit, in genau k Iterationen bei fester Temperatur T von s nach t zu gelangen.

3. Inhomogene Markov-Prozesse

Sei T_i die Temperaturstufe des i-ten Iterationsschritts im betrachteten Simulated-Annealing-Algorithmus. Das Verhalten des Verfahrens ist dann durch

$$\lim_{k \to \infty} \prod_{i=1}^{k} B(T_i)$$

gegeben. Da man im Verlauf des Algorithmus das bislang beste Element leicht fortschreiben kann, stellt sich die Frage: Ist es möglich, dass kein Iterationsschritt ein minimales Element von S erzeugt?

Um die Ergebnisse von Anily und Federgruen [1987] zu beschreiben, bezeichnen wir mit S_* die Menge aller lokalen Optima von S und definieren

$$\underline{e}(T) = \min\{e^{-(c(t)-c(s))/T} : a_{st} > 0\}$$

$$\overline{e}(T) = \max\{e^{-(c(t)-c(s))/T} : s \in S_*,\ t \notin S_*,\ a_{st} > 0\}.$$

Ist nun $|S| = n$, so gilt

Theorem 1 (Anily und Federgruen [1987]:

(a) Falls $\sum_{i=1}^{\infty} \underline{e}^i(T_{in}) = \infty$, dann erzeugt der Algorithmus ein minimales Element. Ausserdem existiert $\lim_{i\to\infty} \pi_s(T_i)$ für alle $s \in S$.

(b) Falls $\sum_{i=1}^{\infty} \overline{e}(T_i) < \infty$ und suboptimale lokale Optima existieren, dann erzeugt der Algorithmus mit strikt positiver Wahrscheinlichkeit kein minimales Element.

Beweisskizze: Sei $s^* \in S$ optimal und $s \in S$ das Startelement des Algorithmus. Mit p_s^k bezeichnen wir die Wahrscheinlichkeit dafür, dass in den Iterationsschritten n, 2n, ... kn ein element $t \neq s^*$ erzeugt wird.

Wegen $T_1 \geq T_2 \geq \ldots \geq T_i \geq \ldots$ ergibt sich dann die Abschätzung

$$p_s^k \leq \left(1 - d^n \underline{e}^n(T_{kn})\right) p_s^{k-1},$$

wobei $d = \min\{a_{st} : a_{st} > 0\}$.
Folglich erhält man im Fall (a) die Aussage

$$p_s^k \leq \prod_{i=1}^{k} \left(1 - d^n \underline{e}^n(T_{in})\right) \to 0.$$

Die zweite Aussage resultiert aus der gleichen Abschätzung unter Verwendung des von Dobrushin eingeführten Ergodizitätskoeffizienten.

Ähnlich beweist man Teil (b) des Theorems. □

4. Homogene Markov-Prozesse

Anily und Federgruens allgemeines Konvergenztheorem macht im Teil (a) keine Aussage über die numerische Grösse von $\lim_{T\to 0} \pi_s(T)$. Für gute Konvergenz erwartet man intuitiv

$$\lim_{T\to 0} \pi_s(T) = 0 \quad \text{falls } s \text{ nicht optimal.}$$

Die Richtigkeit einer solchen Aussage ist im allgemeinen jedoch nicht gegeben.

Beispiel: $S = \{1,2,3\}$ mit $c(1) < c(2) < c(3)$.
Falls $a_{12} = a_{23} = a_{31} = 1$, dann ist

$$\lim_{T\to 0} \pi_1(T) \neq 1!$$

□

Definition: Das Simulated-Annealing-Verfahren heisst *stark konvergent*, falls für jedes nicht-optimale $s \in S$ gilt

$$\lim_{T\to 0} \pi_s(T) = 0.$$

□

Bei gewissen Symmetrieeigenschaften der Nachbarschaftsstruktur $G(A)$, die dem Algorithmus zugrunde liegt, kann nun starke Konvergenz garantiert werden.

Theorem 2 (Faigle und Schrader [1988]:

Ist $G(A)$ symmetrisch (d.h. $a_{st} > 0 \Leftrightarrow a_{ts} > 0$), dann ist das Simulated-Annealing-Verfahren stark konvergent. □

Der Beweis von Theorem 2 wird im wesentlichen kombinatorisch geführt. Grundlegend ist folgendes

<u>Lemma</u>: Es gibt eine Bijektion $W \longleftrightarrow W'$ auf der Menge der gerichteten Wege in G(A) so, dass

(i) Führt W von s nach t, dann führt W' von t nach s.

(ii) $|W| = |W'|$.

(iii) Es gibt eine Konstante a, die nur von A abhängt so, dass für alle Wege W

$$A(W) \leq a \cdot A(W'),$$

wobei A(W) das Produkt der Wahrscheinlichkeiten entlang W ist.

□

Da die Konstante a unabhängig von den betrachteten Weglängen ist, folgt für die Matrizenprodukte $B^k(T)$:

$$b_{st}^{(k)}(T) \leq a \cdot b_{ts}^{(k)}(T) \cdot e^{-(c(t)-c(s))/T}$$

wennimmer $c(t) > c(s)$ und T genügend klein ist. Also gilt bei $c(t) > c(s)$

$$\pi_t(T) = \lim_{k\to\infty} b_{st}^{(k)}(T) \leq a \cdot e^{-(c(t)-c(s))/T} \to 0.$$

5. Bemerkungen

Das Konvergenzresultat von Anily und Federgruen [1987] ist unabhängig von der Struktur von G(A) und macht eine Aussage über die Absenkgeschwindigkeit der Temperatur bei Akzeptanzwahrscheinlichkeiten $p_{st}(T)$. Eine genauere Analyse zeigt, dass das Argument dabei die $p_{st}(T)$ nicht unbedingt in exponentieller Form wie hier diskutiert voraussetzt. Entsprechende Verallgemeinerungen werden in Anily und Federgruen [1987] angegeben.

Die starke Konvergenz hängt entscheidend von der Struktur von G(A) ab. Eine Verallgemeinerung der hier behandelten Symmetrie von G(A) ist in

Hajek [1985] unter dem Begriff "weak reversibility" vorgeschlagen worden. Dabei wird nur verlangt, dass es zu jedem Weg W von s nach t, der das Gewicht c nirgendwo überschreitet, einen ebensolchen Weg W' von t nach s zurück gibt.

Die Analyse dieses Modells erscheint jedoch ziemlich involviert. Ein einfacher kombinatorische Beweis, der in diesem Modell starke Konvergenz nachweist, wäre sicherlich interessant.

6. Literatur

S. Anily und A. Federgruen [1987]: Simulated annealing methods with general acceptance probabilities.
J. Appl. Prob. 24 (1987), 657-667.

V. Černý [1985]: Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm.
J. Optimization Theory and Appl. 45 (1985), 41-51.

U. Faigle und R. Schrader [1988]: On the convergence of stationary distributions in simulated annealing algorithms.
Information Processing Letters 27 (1988), 189-194.

B. Hajek [1985]: Cooling schedules for optimal annealing.
Preprint, Coordinated Science Lab., Univ. of Illinois, 1985.

S. Kirkpatrik, S.D. Gelatt jr. und M.P. Vecchi [1983]: Optimization by simulated annealing.
Science 220 (1983), 671-680.

N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller und E. Teller [1953]: Equation of state calculation by fast computing machines.
J. Chem. Phys. 21 (1953), 1087-1092.

E. Senata [1981]: Non-negative Matrices and Markov Chains.
Springer, New York/Heidelberg/Berlin, 2nd ed., 1981.