ON A RANDOMIZED VERSION OF PARALLEL LOCAL SEARCH

Matthias Löwe

Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
Germany

Abstract. We introduce a stochastic version of the parallel local search algorithm. Using techniques similar to those used to establish convergence of the simulated annealing algorithm we are able to give a cooling schedule for which the process converges and to analyze the speed of convergence by spectral gap methods.

Key words and phrases. Markov processes, Spectral gap, Optimization, Simulated Annealing.

1. Introduction

Let $G = (X, E)$ be a finite connected graph with vertex set X and edge set E. Let $U : X \to \mathbb{R}$ be a real valued function. A problem which often appears in applications is to determine the minimum value of U or, equivalently, the vertex where it is attained. This problem becomes extremely difficult when the vertex set X is very large as e. g. in the most typical examples, in the Traveling Salesman Problem, (where X consists of all closed tours through a given set of cities) or in the problem of restoring a disturbed image (where X is the set of all possible images, so $|X| \sim 10^{10^4}$ in this case). In these cases it is plain that an application of the so called global search algorithm, which calculates explicitly all values $U(x)$, is hopeless. On the other hand global search is one of the few deterministic strategies that definitely find the global minimum.

In contrast the versions of the local search algorithm often get trapped in a local minimum which usually is not the global one. To be more precise there are essentially two forms of local search, a sequential one and a second one which is basically parallel in concept. The first one, roughly speaking, works in the following way: Start in a point x_0; when having arrived at a vertex x_t at time t, choose a vertex $y \in \mathcal{N}(x_t)$ (where $\mathcal{N}(x)$ denotes the neighborhood of a vertex x in (X, E), and if $U(y) \leq U(x_t)$ set $x_{t+1} = y$ and stay in x_t otherwise. In contrast to that the second algorithm when having arrived at a vertex x_t at time t evaluates $U(y)$ for all $y \in \mathcal{N}(x_t)$ (which is the parallel component of the algorithm) and then steps to one of those vertices in $\{x_t\} \cup \mathcal{N}(x_t)$ that minimize $U(\cdot)$.

Now for the first type of local search (the sequential one) there has been introduced a stochastic version by Kirkpatrick et.al. [K83], to overcome the difficulty of convergence to suboptimal points. This stochastic version is the well known Simulated Annealing (SA) Algorithm. The idea behind SA is the following: modify the local search scheme described above in such a way that it induces an irreducible and reversible, time inhomogeneous random walk on X with an invariant measure that is Gibbsian with energy U and at some inverse temperature β_t. While running this random walk we slowly cool down the process (i.e. we increase $\beta_t \to \infty$) such that the invariant measure finally becomes concentrated on the set of minima of $U(\cdot)$.

More precisely SA modifies the original random walk created by the first version of local search in such a way that each transition probability from x to y is multiplied by a factor $\exp(-\beta_t(U(x) - U(y)^+)$, where $x^+ = \max(x, 0)$.

Basically there are two questions intrinsic in the description of this algorithm, namely how can the cooling schedule β_t be chosen to ensure convergence of the algorithm to the set of global minimum of $U(\cdot)$, and in case of convergence, how fast does the process become concentrated on the set of minima of U?

The first convergence results of SA have been established by Mitra et al. [M85] and Gidas [G85]. Later Hajak [Ha88] was able to solve the problem of an asymptotically optimal cooling schedule by proposing a functional dependence of β_t from t of the form $\beta_t = \frac{1}{m} log(A + B_t)$, where A and B are positive constants and m is approximately the largest hill to be climbed to reach the global minimum. The right exponential order of the speed of convergence has been investigated by Catoni [C90] by using Freidlin–Wentzell techniques [FW84].

A new light was shed on the mechanism behind the convergence of SA by the work of Holley and Stroock [HS88] who investigated the speed of convergence by establishing and solving a differential inequality for the Radon–Nikodym density of the distribution of the process at time t with respect to the Gibbs measure at time t with the help of Sobolev inequalities. A refined analysis of these differential inequalities by Götze [G92] led to applicable bounds for the entrance time into typical subsets of the state space X.

The techniques developed in [G92] were generalized by Deuschel and Mazza [DM92] to a class of Markov processes of the Freidlin–Wentzell type and have also been applied in [L94a] and [L94b] to analyze the convergence of SA with time-dependent energy function and a wide class of genetic algorithms, respectively.

The aim of this paper is to introduce a randomized version of the second type of local search (the parallel one) similar to SA, to give good cooling schedules for it and to investigate its convergence properties. More precisely the algorithm we have in mind works in the following way: Start in a vertex $x_0 \in X$. When having arrived in x_t at time t, calculate $U(y)$ for all $y \in \mathcal{N}(x_t)$. Then select x_{t+1} with the help of a Gibbs distribution on $\{x_t\} \cup \mathcal{N}(x_t)$ with energy function $U(\cdot)$ and at inverse temperature β_t. While continuing with this procedure, increase β_t according to a fixed cooling schedule. Note that similar algorithms have already been investigated in [L94b]. But there we had to suppose a symmetry condition, which usually is not fulfilled for the process described above.

The paper is organized as follows: In Section 2 we will give a more formal description of the algorithm. In Section 3 we show that if we choose the algorithm in such a way that it induces a random walk on X that is irreducible and reversible, then the invariant measure becomes concentrated on the set of global minima of $U(\cdot)$ as $\beta_t \to \infty$. This makes it possible to apply theorems for the convergence of a class of Markov processes of the Freidlin-Wentzell type by Deuschel and Mazza [DM92]. Together with the speed of convergence these theorems also yield good cooling schedules.

2. The Model

Recall that X was a finite set and $U : X \to \mathbb{R}$ was a real valued function. Without loss of generality we may assume that $\min_x U(x) = 0$ and that the edge set E is induced by a random walk $q(\cdot, \cdot)$ on X, i. e. the (oriented) edge $(x, y) \in E$ if and only if $q(x, y) > 0$. So, first of all E is a set of oriented edges.

Assume that the following two conditions a fulfilled:

(I) The random walk $(q(x, y)_{x,y \in X}$ is irreducible, i. e. for all x, y there is an integer N such that $q^{(N)}(x, y) > 0$.

(Here $q^{(N)}(x, y)$ is defined inductively by $q^{(1)}(x, y) = q(x, y)$ and the Chapman–Kolmogorov equations $q^{(N)}(x, y) = \sum_{z \in X} q^{(N-1)}(x, z)q(z, y)$)

(R) The random walk $(q(x, y))_{x,y \in X}$ is reversible with respect to a probability distribution π on X, i. e.

$$\alpha(x, y) := \pi(x)q(x, y) = \pi(y)q(y, x) = \alpha(y, x).$$

Conditions (I) and (R) imply that $(q(x,y))_{x,y\in X}$ possesses $\pi(\cdot)$ as its invariant measure. Moreover $\pi(x) > 0$ for all $x \in X$ and hence $q(x,y) > 0$ if and only if $q(y,x) > 0$. Therefore we may regard (X,E) as an ordinary graph (i. e. the edges are not oriented).

The similarity of our algorithm to SA is now introduced by the following time-dependent version of $(q(x,y))$. Let us choose a monotonely increasing function β_t, with $\beta_t \geq 0$ for all $t > 0$ and $\beta_t \underset{t\to\infty}{\rightarrow} \infty$, and define the random walk $(q_t(x,y))_{x,y\in X}$ via

$$q_t(x,y) = \frac{q(x,y)e^{-\beta_t(U(y)-U(\kappa_x))}}{Z_t^x},$$

for $y \in \{x\} \cup \mathcal{N}(x), y \neq x$, and

$$q_t(x,x) = 1 - \sum_{y\neq x} q_t(x,y).$$

Here

$$Z_t^x = \sum_{y\in\{x\}\cup\mathcal{N}(x)} e^{-\beta_t(U(y)-U(\kappa_x))}q(x,y)$$

and κ_x is given by the equation

$$U(\kappa_x) = \min_{z\in\{x\}\cup\mathcal{N}(x)} U(z).$$

¿From the point of view of functional analysis the random walks $(q(x,y))$ and $(q_t(x,y))$ induce self-adjoint contractions Q and Q_t on the space of all functions $\phi : X \to \mathbb{R}$ via

$$[Q\phi](x) = \sum_{y\in X} \phi(y)q(x,y) \quad \text{for all } x \in X$$

and

$$[Q_t\phi](x) = \sum_{y\in X} \phi(y)q_t(x,y) \quad \text{for all } x \in X,$$

respectively. Let us also introduce the generator $L_t := Q_t - I$ of the random walk $(q_t(x,y))$, which is

$$[L_t\phi](x) = \sum_{y\in X} (\phi(y) - \phi(x))q_t(x,y)$$

for all $x \in X$ and $\phi : X \to \mathbb{R}$. With these definitions we are able to give a formal description of the search process we have in mind. If we denote by $P_{s,t}(x,y)$ the transition probabilities to arrive in y at time t when starting in x at time s, $s \leq t$, the search process is determined by the following forward equation

$$\frac{\partial}{\partial t}[P_{s,t}\phi](x) = [P_{s,t}[L_t\phi]](x) \quad \text{for } s < t \tag{1}$$

and

$$[P_{s,s}\phi](x) = \phi(x)$$

for $\phi : X \to \mathbb{R}$ and $x \in X$. Here the operators $[P_{s,t}\cdot]$ associated with the transition probabilities $P_{s,t}(x,y)$ are defined by

$$[P_{s,t}\phi](x) = \sum_{y \in X} \phi(y) P_{s,t}(x,y) \quad \text{for all } x \in X$$

and for $\phi : X \to \mathbb{R}$.

In the following let us reserve the symbol $(Y_t)_{t \geq 0}$ for the process defined by (1). The intuitive description of Y_t is the following. When starting in a vertex x_s at time s, wait an exponential time with mean one, and after having waited, say at time t, choose a new vertex x_t with probability $q_t(x_s, x_t)$, i.e. with the appropriate Gibbs distribution with energy $U(\cdot)$ on $\{x_s\} \cup \mathcal{N}(x_s)$, etc..

3. On the Convergence of the Algorithm

In this section we are going to investigate the ergodic behavior of Y_t. To make Y_t a genuine search process we, of course, should show that Y_t converges to the set of minima of $U(\cdot)$. The only chance to obtain such a result is to show that Y_t follows an invariant measure $\pi_t(\cdot)$, if β_t is chosen appropriately (i. e. if the process cools down slowly enough) and that $\pi_t(\cdot)$ becomes concentrated on the set of minima of $U(\cdot)$. The latter is the content of the following lemma.

Lemma 3.1. *The random walk $(q_t(x,y))_{x,y \in X}$ possesses an invariant measure π_t. π_t is of the form*

$$\pi_t(x) = \exp(-\beta_t(V(x) - V_{\min}))\Gamma_x(\beta_t)$$

for each $x \in X$, some functions $\Gamma_x(\beta) : \mathbb{R}^+ \to \mathbb{R}^+$ bounded uniformly from above and away from 0, and a function $V : X \to \mathbb{R}$ defined by (2) in the proof. U and V are minimized by the same set of vertices. Especially for the set of minima of U

$$U_{\min} = \{x \in X : U(x) = 0\}$$

it holds

$$\pi_t(U_{\min}^c) \leq A \exp(-\beta_t x^*)$$

for a constant $A > 0$ and x^* is defined by

$$x^* = \min\{V(x) : x \notin U_{\min}\}.$$

The proof of Lemma 3.1 is essentially based on the classical Matrix Tree Theorem by Bott and Mayberry [BM54]. In order to formulate this theorem we have to introduce some notation:

Let Y be a finite set and $(R(x, y))$ be a random walk on Y. $(R(x, y))$ defines a set of directed edges $F \subset Y \times Y$ on Y by $(x, y) \in F$ if and only if $R(x, y) > 0$.

For $W \subset Y$ we call a graph (Y, F') a W-graph, if its vertex set is Y, its set of edges is $F' \subset F$, and it satisfies the following two conditions:

(1) Every vertex $x \in X \setminus W$ is the initial vertex of one (oriented) edge.
(2) The graph (Y, F') is cycle-free.

We denote the set of all W-graphs by $G(W)$ and its elements by g. With these notations the Matrix Tree Theorem reads as follows.

Theorem 3.2. If the random walk $(R(x, y))$ on Y is irreducible, i. e. if (Y, F) is connected, then $(R(x, y))$ has an invariant measure ζ and ζ is of the form

$$\zeta(x) = \frac{\bar{\zeta}(x)}{\sum_{y \in Y} \bar{\zeta}(y)},$$

with

$$\bar{\zeta}(x) = \sum_{g \in G(\{x\})} \Pi_{(y,z) \in g} R(y, z).$$

For a proof of Theorem 3.2 see e. g. [FW84], pp. 177-178. Let us now see how to apply Theorem 3.2 to prove Lemma 3.1.

Proof of Lemma 3.1. Since $(q_t(x, y))$ is irreducible it has an invariant distribution π_t and

$$\pi_t(x) = \frac{\sum_{g \in G(\{x\})} \Pi_{(y,z) \in g} q_t(y, z)}{\sum_{y \in X} \sum_{g \in G(\{y\})} \Pi_{(z,w) \in g} q_t(z, w)}.$$

The numerator of this fraction can be computed as

$$\begin{aligned} &\sum_{g \in G(\{x\})} \Pi_{(y,z) \in g} q_t(y, z) \\ = &\sum_{g \in G(\{x\})} \exp\left(-\beta_t \sum_{(y,z) \in g} (U(z) - U(\kappa_y))\right) \Pi_{(y,z) \in g} \frac{q(y, z)}{Z_t^y}. \end{aligned}$$

Observe that for each y and t we have the relation

$$c \leq Z_t^y \leq 1$$

where we have put $c := \min_{y \in X} \sum_{\substack{z \in \mathcal{N}(y): \\ U(z)=U(\kappa_y)}} q(y,z)$. Hence for each $x \in X$ we can find a function $\bar{\Gamma}_x(\beta) : \mathbb{R}^+ \to \mathbb{R}^+$ bounded from above and away from 0, such that for

$$V(x) = \min_{g \in G(\{x\})} \sum_{(y,z) \in g} U(z) - U(\kappa_y) \tag{2}$$

the numerator of the above fraction can be written in the form

$$\sum_{g \in G(\{x\})} \Pi_{(y,z) \in g} q_t(y,z) = \exp(-\beta_t V(x)) \bar{\Gamma}_x(\beta).$$

By the same arguments the denominator can be written as

$$\sum_{x \in X} \sum_{g \in G(\{x\})} \Pi_{(y,z) \in g} q_t(y,z) = \exp(-\beta_t V_{\min}) \Gamma(\beta),$$

for a function $\Gamma(\cdot)$ bounded from above and away from 0 and with $V_{\min} = \min_{x \in X} V(x)$. Thus $\pi_t(x)$ takes the form

$$\pi_t(x) = \exp(-\beta_t (V(x) - V_{\min})) \Gamma_x(\beta)$$

where $\Gamma_x(\beta) = \frac{\bar{\Gamma}_x(\beta)}{\Gamma(\beta)}$ is bounded from above and away from 0.

It remains to show that U and V are minimized by the same set of vertices. So take a vertex x with $U(x) = 0$ and a vertex y with $V(y) = V_{\min}$ and assume that $x \neq y$.

Take a path $g \in G(\{y\})$ with

$$\sum_{(w,z) \in g} U(z) - U(\kappa_w) = V(y).$$

Observe that g contains a unique trajectory t_{xy} leading from x to y. Reversing t_{xy} to t_{yx}, i.e. reversing the direction of every edge makes t_{yx} a trajectory from y to x and thus a graph $g \in G(\{y\})$ is mapped to a graph $\bar{g} \in G(\{y\})$ by building $g = g \setminus t_{xy} \cup t_{yx}$. (Also note that this mapping is one to one). Now

$$\left(\sum_{(w,z) \in g} U(z) - U(\kappa_w) \right) - \left(\sum_{(w,z) \in \bar{g}} U(z) - U(\kappa_w) \right) = U(y) - U(\kappa_x) - U(x) + U(\kappa_y) \geq 0$$

(since $U(\kappa_x) = U(x)$) with equality if and only if $U(y) = 0$. This shows that V can take its minimum only on vertices x with $U(x) = 0$ and for every such x there exists a path in $G(\{x\})$ minimizing V. This proves the lemma.

□

Remark 3.3. Note $\Gamma_x(\beta)$ is differentiable with respect to β and that

$$\sup_\beta |\frac{d}{d\beta}\Gamma_x(\beta)| < \infty.$$

To be able to formulate our central convergence result we also need a result on the spectral gap of the operator L_t.
More precisely define the π_t-dual of $L_t = (l_t(x,y))_{x,y}$ by $L_t^* := (l_t^*(x,y))_{x,y}$ and

$$l_t^*(x,y) = \frac{l_t(x,y)\pi_t(y)}{\pi_t(x)}.$$

The symmetric operator $\bar{L}_t$ is then defined by $\bar{L}_t = \frac{1}{2}(L_t + L_t^*)$ and its associated Dirichlet form $\mathcal{E}_t$ is given by

$$\mathcal{E}_t(\phi,\phi) := - < [\bar{L}_t\phi], \phi >_t = \frac{1}{2}\sum_{x,y\in X}(\phi(x)-\phi(y))^2 R_t(x,y),$$

for $\phi : X \to \mathbb{R}$. Here $< \cdot, \cdot >_t$ denotes the scalar product in $L^2(\pi_t)$ and

$$\begin{aligned} R_t(x,y) = R_t(y,x) &= \frac{1}{2}(\pi_t(x)l_t(x,y) + \pi_t(y)l_t(y,x)) \\ &= \frac{1}{2}(\pi_t(x)q_t(x,y) + \pi_t(y)q_t(y,x)) \quad \text{if} \quad x \neq y. \end{aligned}$$

Then the spectral gap λ_t of the operator L_t is defined as

$$\lambda_t := \inf\left\{\frac{\mathcal{E}_t(\phi,\phi)}{Var_t(\phi)} : \phi \quad \text{non constant}\right\}. \tag{3}$$

Here $Var_t(\phi)$ denotes the variance of ϕ with respect to π_t. As a matter of fact it can be shown by elementary linear algebra (see e. g. [HJ85]) that λ_t is a genuine spectral gap, i. e. the distance between the largest and second largest eigenvalue of $\bar{L}_t$.

To estimate the spectral gap λ_t we introduce the following quantities. For any two vertices $x, y \in X$ let us define a path leading from x to y to be any sequence of vertices

$x = x_1, \ldots, x_k = y$, $k \in \mathbb{N}$ with $q(x_{i-1}, x_i) > 0, i = 2, \ldots k$. By Λ_{xy} we denote the set of all paths leading from x to y. For $\gamma \in \Lambda_{xy}$ let

$$e(\gamma) = \max\{U(x_{i+1}) - U(\kappa_{x_i}) + V(x_i) - V_{\min}, U(x_i) - U(\kappa_{x_{i+1}}) + V(x_{i+1}) - V_{\min} : x_i, x_{i+1} \in \gamma\}$$

where the enumeration of the x_i follows the natural way of the path from x to y. By $h(x, y)$ we denote the minimum $e(\cdot)$ we have to face to get from x to y, i. e.

$$h(x, y) = \min_{\gamma \in \Gamma_{xy}} e(\gamma) \quad x \neq y$$

and $h(x, x) = 0$. Finally put

$$m = \max_{x,y \in X} (h(x, y) - V(x) - V(y) + 2V_{\min}).$$

It is quickly checked that the max is attained only if $V(y) = V_{\min}$ or $V(x) = V_{\min}$ and thus $m \geq 0$. Then the order of the spectral gap λ_t is given by the following lemma.

Lemma 3.4. *There exists a constant $c > 0$ such that for all $t \geq 0$*

$$\lambda_t \geq c e^{-\beta_t m}.$$

Proof. The proof essentially follows the ideas of the proof of Lemma 2.7 in [HS88]. By the definition (3) of the spectral gap it suffices to prove the existence of a positive constant c, such that for all $t \geq 0$ and all $\phi : X \to \mathbb{R}$

$$\frac{\mathcal{E}_t(\phi, \phi)}{Var_t(\phi)} = \frac{-\sum_{x \in X} \phi(x)[L_t\phi](x)\pi_t(x)}{Var_t(\phi)} \geq c e^{-\beta_t m}.$$

This is done by bounding the variance as follows:
For any two $x, y \in X$ take a path $\gamma_{x,y} \in \Lambda_{xy}$ leading from x to y with vertices $\gamma_{x,y}(i), i = 1, \ldots, n(x, y)$ such that $e(\gamma_{x,y}) = h(x, y)$. Here $n(x, y)$ denotes the length of $\gamma_{x,y}$ and we put

$$N := \max_{x,y \in X} n(x, y).$$

Let us also introduce the indicator

$$\chi_{z,w}(x, y) = \begin{cases} 1 & \text{if } \exists\ 1 \leq i \leq n(x, y) - 1 : \gamma_{x,y}(i) = z \text{ and } \gamma_{x,y}(i+1) = w \\ 0 & \text{otherwise.} \end{cases}$$

Then for any $\phi : X \to \mathbb{R}$

$$
\begin{aligned}
2Var_t(\phi) &= \sum_{x,y\in X} (\phi(x)-\phi(y))^2 \pi_t(x)\pi_t(y) \\
&= \sum_{x,y\in X} \left(\sum_{i=2}^{n(x,y)} \phi(\gamma_{x,y}(i)) - \phi(\gamma_{x,y}(i-1)) \right)^2 \pi_t(x)\pi_t(y) \\
&\leq \sum_{x,y\in X} n(x,y) \sum_{i=2}^{n(x,y)} (\phi(\gamma_{x,y}(i)) - \phi(\gamma_{x,y}(i-1)))^2 \pi_t(x)\pi_t(y)
\end{aligned}
$$

by Jensen's inequality. Therefore

$$
\begin{aligned}
2Var_t(\phi) \leq N & \sum_{x,y,z,w\in X} \chi_{z,w}(x,y)(\phi(z)-\phi(w))^2 \frac{1}{2}(\pi_t(z)q_t(z,w)+\pi_t(w)q_t(w,z))\times \\
& \frac{\pi_t(x)\pi_t(y)}{\frac{1}{2}(\pi_t(w)q_t(w,z)+\pi_t(z)q_t(z,w))} \\
\leq N & \left[\max_{z,w\in X} \sum_{x,y\in X} \chi_{z,w}(x,y) \frac{2\pi_t(x)\pi_t(y)}{\pi_t(w)q_t(w,z)+\pi_t(z)q_t(z,w)} \right] \times \\
& \left[\sum_{z,w\in X} (\phi(z)-\phi(w))^2 R_t(z,w) \right].
\end{aligned}
$$

Since by definition the second term in brackets is twice the Dirichlet form $\mathcal{E}_t(\phi,\phi)$ it suffices to bound the first term. This is done by

$$
\begin{aligned}
& \chi_{z,w}(x,y) \frac{\pi_t(x)\pi_t(y)}{\frac{1}{2}(\pi_t(w)q_t(w,z)+\pi_t(z)q_t(z,w))} \\
& \leq \chi_{z,w}(x,y) e^{-\beta_t(V(x)-V(y)-2V_{\min})} \Gamma_x(\beta)\Gamma_y(\beta) \\
& \times \frac{1}{\min(\pi_t(w)q_t(w,z),\pi_t(z)q_t(z,w))} \\
& \leq \frac{\chi_{z,w}(x,y)}{\min(q(w,z),q(z,w))} \frac{1}{\min(\Gamma_z(\beta),\Gamma_w(\beta))} \\
& \times e^{\beta_t(\max(U(z)-U(\kappa_w)+V(w)-V_{\min}, U(w)-U(\kappa_z)+V(z)-V_{\min})} \\
& \times e^{\beta_t(2V_{\min}-V(x)-V(y))} \Gamma_x(\beta)\Gamma_y(\beta) \\
& \leq \frac{\lambda_{z,w}(x,y)}{\min(q(w,z),q(z,w))} \frac{\Gamma_x(\beta)\Gamma_y(\beta)}{\min(\Gamma_t(\beta),\Gamma_w(\beta))} e^{\beta_t m}.
\end{aligned}
$$

Here we put $\frac{\chi_{t,w}(x,y)}{\min(q(z,w),q(w,t))} = 0$, whenever $\chi_{z,w}(x,y) = 0$.
Since the $\Gamma_x(\beta)$ are bounded from above and away from 0, this proves the assertion of the lemma.

□

Remark 3.5. It should be mentioned that similar bounds on the spectral gap for a class of operators associated with Markov processes of the Freidlin–Wentzell form are obtained in [DM92], Theorem 4.2.4. There the authors use the Poincaré method as in [DS91] or [G92].

Now we are ready to use a result by Deuschel and Mazza [DM92], to establish the convergence of our process.

Theorem 3.6. Consider the Markov process Y_t given by the forward equation (1) and choose a freezing schedule β_t such that

$$\beta_t = \frac{1}{m}\log\left(1+\frac{2mc}{3M}\right)$$

where m and c are chosen as in Lemma 3.4 and M is a constant such that

$$|\frac{d}{d\beta_t}\log\pi_t(x)| \leq M \qquad \forall x \in X.$$

Then for the Radon–Nikodym density f_t of the distribution of Y_t with respect to π_t it holds

$$||f_t - 1||_{2,t} \leq 1 + \exp(-M\beta_t)||f_0 - 1||_{2,t}.$$

Here $||\cdot||_{2,t}$ denotes the norm in $L^2(\pi_t)$.

Moreover if $\bar{X}$ is any subset of $X, \bar{X} \neq \emptyset, X$, then for any starting point $x \in X$ of Y_t

$$P^x(Y_t \notin \bar{X}) \leq (\pi_t(\bar{X}))^{\frac{1}{2}}\left(1 + e^{-M\beta_t} + (\frac{1}{\pi(x)} - 1)^{\frac{1}{2}}\right).$$

For a proof we just remark that we have already shown that the conditions of Lemma 2.1.6 and Corollary 2.1.12 of [DM92] are fulfilled.

References

[C90] O. Catoni, *Asymptotics of Simulated Annealing Algorithms*, Thèse de Doctorat, Université Paris-Sud, Orsey (1990).

[DM92] J.D. Deuschel, C. Mazza, *L^2 Convergence of Time Non-Homogeneous Markov Processes I:Spectral Estimates*, Preprint (1992).

[DS91] P. Diaconis, D. Stroock, *Geometric Bounds for Eigenvalues of Markov Chains*, Ann. Appl. Prob. **1** (1991), 36–61.

[FW84] M.I. Freidlin, A.D. Vencel, *Random Perturbations of Dynamical Systems,*, Grundlehren d. math. Wissenschaften **260** (1984), Springer.

[GG84] S. Geman, D. Geman, *Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images*, IEEE Trans. on Pattern Anal. and Machine Intelligence **PAMI-6** (1984), 721–741.

[Gi85] B. Gidas, *Nonstationary Chains and the Convergence of the Annealing Algorithm*, J. Stat. Phys. **39** (1985), 73–131.

[G 92] F. Götze, *Rate of Convergence for Simulated Annealing Processes*, Preprint 92-008 (1992).

[Ha88] B. Hajek, *Cooling Schedules for Optimal Annealing*, Math. of Operations Research **13** (1988), 311–329.

[HJ85] R. Horn, C. Johnson, *Matrix Analysis*, Camb. Univ. Press (1985).

[K83] S. Kirkpatrick, C. Gelatt, M.P. Vecchi, *Optimization by Simulated Annealing*, Science **220** (1983), 621–680.

[L94a] M. Löwe, *Simulated Annealing with Time Dependent Energy Function via Sobolev Inequalities*, Preprint 94-082, SFB 343 (Bielefeld) (1994).

[L94b] M. Löwe, *Stochastic Search on Graphs and Genetic Algorithms*, Preprint, Bielefeld (1994).

[M85] D. Mitra, F. Romeo, A. Sangiovanni-Vincentelli, Convergence and Finite Time Behaviour of Simulated Annealing, Proc. 24th Conf. Dec. a. Control (1985), 761–767.