Simulated Annealing via Sobolev Inequalities

Richard Holley[1]★ and Daniel Stroock[2]★★

[1] Department of Mathematics, University of Colorado, Boulder, Colorado 80309–0426, USA
[2] Department of Mathematics, M.I.T., Cambridge, MA 02139, USA

Abstract. We use Sobolev inequalities to study the simulated annealing algorithm. This approach takes advantage of the local time reversibility of the process and yields the optimal "freezing schedule" as well as quantitative information about the rate at which the process is tending to its ground state.

0. Introduction

Simulated annealing is a Monte Carlo method for locating the minimum of a complicated function U on some space $\mathbf{E}$. Thinking of $U(x)$ as the "energy" of the "state" $x \in \mathbf{E}$, the method is to run a time-inhomogeneous Markov process on $\mathbf{E}$ in such a way that at any time $t \geq 0$ the corresponding instantaneous time-homogenous Markov process has as its equilibrium distribution the Gibbs measure with energy U at inverse temperature $\beta(t)$. By choosing $t \to \beta(t)$ to go to infinity as $t \to \infty$, the idea is to force the process toward the minima of U.

Inherent in the simulated annealing procedure is a competition between two goals. On the one hand, one wants to make the temperature tend to 0 as fast as possible, thereby concentrating the Gibbs measures as fast as possible near the minima of U. On the other hand, having the Gibbs measure concentrated near the minima of U will do one no good unless the process at large times is close to equilibrium. Since, as we will see below (cf. Theorem (2.1)), decreasing the temperature inhibits equilibration, one must be careful not to adopt too fast a "freezing schedule" (i.e. rate at which $\beta(t) \to \infty$).

The problem raised above has been studied previously by many authors ([1–7]), and much of what we prove below is qualitatively similar to known results. Nonetheless, we believe that our approach to this problem has several features to recommend it. In the first place, it is simple and essentially context independent. Secondly, it yields quantitative information about the rate at which the process is tending to the minima of U. In fact, the results in Sect. 2 indicate

that at least asymptotically, our estimates may be optimal. Because in practice it is the most realistic situation and since its description involves the fewest technicalities, we will describe everything for the case when the state space $\mathbf{E}$ is finite. However, as will be apparent to anyone familiar with Dirichlet forms, only the results in Sect. 2 really depend on $\mathbf{E}$'s being finite, and even those results have extensions to other contexts.

Let μ_0 be a probability measure on $\mathbf{E}$ and assume that μ_0 charges every point (e.g. normalized counting measure). Next let $q_0(x, y)$ be a probability transition function on $\mathbf{E}$ which, in practice, should be chosen so that $U(y) - U(x)$ can be easily evaluated whenever x and y are points for which $q_0(x, y) \neq 0$. We assume about $q_0(x, y)$ that it is *irreducible* on $\mathbf{E}$ (i.e. $\sum_{n=0}^{\infty} q_0^{(n)}(x, y) = \infty$ for all $x, y \in \mathbf{E}$, where $q_0^{(n+1)}(x, y) = \sum_{\xi \in \mathbf{E}} q_0(x, \xi) q_0^{(n)}(\xi, y))$ and that it is μ_0-*reversible* in the sense that

$$\alpha(x, y) \equiv \mu_0(x) q_0(x, y) = \alpha(y, x), \quad x, y \in \mathbf{E}. \tag{0.1}$$

(One may think of μ_0 as the Gibbs measure at infinite temperature and $q_0(x, y)$ as controlling the one-step transitions that the annealing process may make.)

Next, define

$$\mu_\beta = \frac{e^{-\beta U}}{Z_\beta} \mu_0, \quad \beta \geq 0,$$

where $Z_\beta \equiv \int e^{-\beta U} d\mu_0$. Then μ_β is the Gibbs measure at inverse temperature β. Corresponding to μ_β we define the transition probability function $q_\beta(x, y)$ by

$$q_\beta(x, y) = \begin{cases} \exp\left(-\beta(U(y) - U(x))^+\right) q_0(x, y) & \text{if } y \neq x \\ 1 - \sum_{\xi \neq x} q_\beta(x, \xi) & \text{if } y = x. \end{cases} \tag{0.2}$$

With the preceding conventions in mind, we can now define the transition function $P_{s,t}(x, y)$ for our simulated annealing process. Namely, define

$$\mathbf{L}_\beta \phi(x) = \sum_{y \in \mathbf{E}} (\phi(y) - \phi(x)) q_\beta(x, y), \quad x \in \mathbf{E}, \tag{0.3}$$

for $\phi: \mathbf{E} \to R_1$. Then the simulated annealing process corresponding to the freezing schedule $t \to \beta(t)$ is to have its transition probabilities determined by the forward (i.e. Fokker–Planck) equation

$$\frac{\partial}{\partial t} P_{s,t}(x, y) = [\mathbf{L}_{\beta(t)}^* P_{s,t}(x, \cdot)](y), \quad t \geq s, \tag{0.4}$$

$$P_{s,s}(x, y) = \delta_{x,y},$$

where the meaning of the first equation in (0.4) is that

$$\frac{\partial}{\partial t} [P_{s,t} \phi](x) = [P_{s,t} \mathbf{L}_{\beta(t)} \phi](x), \quad t \geq s, \tag{0.4'}$$

for $\phi: \mathbf{E} \to R_1$. (In (0.4') and elsewhere, $[P_{s,t} \phi](x) \equiv \sum_{y \in \mathbf{E}} \phi(y) P_{s,t}(x, y)$ is the operator determined by $P_{s,t}(x, y)$.)

As is well-known, $P_{s,t}(x,y)$ satisfies the Chapman–Kolmogorov equation

$$P_{r,t}(x,y) = \sum_{\xi \in E} P_{r,s}(x,\xi) P_{s,t}(\xi,y), \quad 0 \leq r < s < t;$$

which, in conjunction with (0.4'), yields the backward equation:

$$\frac{\partial}{\partial s}[P_{s,t}\phi](x) = -[\mathbf{L}_{\beta(s)}P_{s,t}\phi](x), \quad 0 \leq s \leq t. \tag{0.5}$$

Also, once the initial distribution v has been specified, the transition probabilities $P_{s,t}(x,y)$ completely determine our simulated annealing process $X(t)$ via the equation

$$\text{Prob}(X(t_0)\in\Gamma_0, \ldots, X(t_n)\in\Gamma_n) = \sum_{y_n\in\Gamma_n} \cdots \sum_{y_0\in\Gamma_0} v(y_0)P_{t_0t_1}(y_0,y_1)\cdots P_{t_{n-1,n}}(y_{n-1},y_n) \tag{0.6}$$

for $0 = t_0 < \cdots < t_n$ and $\Gamma_0, \ldots, \Gamma_n \subset \mathbf{E}$.

The intuitive description of the simulated annealing process is as follows: If the process is at site x, then it waits an exponential time with mean one and at the end of that time it chooses a site y with probability $q_0(x,y)$ and attempts to move there. If $U(y) \leq U(x)$ then the process moves to y, but if $U(y) > U(x)$ and if the attempted move occurs at time t, then, having chosen y with probability $q_0(x,y)$, the process actually moves to y with probability $e^{-\beta(t)(U(y)-U(x))}$, and with probability $1 - e^{-\beta(t)(U(y)-U(x))}$ it remains at x. This is just the continuous time version of the usual discrete time simulated annealing process (see [8]).

Since what we want to test is whether the simulated annealing process is having time to equilibrate, the quantity of interest is the Radon–Nikodym derivative, f_t, of the distribution of $X(t)$ with respect to the Gibbs measure $\mu_{\beta(t)}$. Then the degree to which equilibrium has been achieved can be measured by keeping track of the size of f_t. For example, let $q\in(1,\infty]$ and $T > 0$ and suppose that $\|f_t\|_{L^q(\mu_{\beta(t)})} \leq C < \infty$ for all $t \geq T$. We would then have that

$$\text{Prob}(U(X(t)) \geq U_{\min} + \delta) \leq C(\mu_{\beta(t)}(\{x: U(x) \geq U_{\min} + \delta\}))^{1-1/q}, \quad t \geq T, \tag{0.7}$$

where $U_{\min} = \min_{\mathbf{E}} U$. Obviously, (0.7) gets better as q gets larger. However, one should expect that one will pay for large q with slower freezing schedules $\beta(t)$ and larger constants C.

Our analysis of $\|f_t\|_{L^q(\mu_{\beta(t)})}$ will be based entirely on consideration of the Dirichlet forms E_β associated with the operators $\mathbf{L}_\beta$. (It is this fact which makes most of analysis context independent.) Namely, define

$$E_\beta(\phi,\psi) = \frac{1}{2Z_\beta} \sum_{x,y\in\mathbf{E}} e^{-\beta(U(x)\vee U(y))}(\phi(x)-\phi(y))(\psi(x)-\psi(y))\alpha(x,y) \tag{0.8}$$

(cf. (0.1)) for functions ϕ and ψ on $\mathbf{E}$. It is then an easy matter to check that

$$-\int \phi\mathbf{L}_\beta\psi\,d\mu_\beta = E_\beta(\phi,\psi). \tag{0.9}$$

In particular,

$$\gamma(\beta) \equiv \inf\{E_\beta(\phi,\phi): \|\phi\|_{L^2(\mu_\beta)} = 1 \text{ and } \langle\phi\rangle_\beta \equiv \int \phi\,d\mu_\beta = 0\}$$

is the gap between 0 and the rest of the $L^2(\mu_\beta)$ spectrum of $-\mathbf{L}_\beta$; and, as such, $\gamma(\beta)$

controls the L^2 – rate at which e^{tL_β} approach equilibrium. Because $q_0(x,y)$ is irreducible, we know that $\gamma(0) > 0$. However, as β gets larger, $\gamma(\beta)$ will get smaller at an exponential rate (cf. Theorem (2.1)).

Assume that we know that

$$\gamma(\beta) \geq \gamma e^{-\beta m}, \quad \beta \geq 0, \tag{0.10}$$

for some $\gamma > 0$ and $m > 0$. (We can always take $\gamma = \gamma(0)$ and $m = M$, where

$$M \equiv \max_E U - \min_E U, \tag{0.11}$$

but sometimes (cf. Theorem (2.1)) one can show that $m < M$.) We will now give a simple example of our ideas as they apply to the study of $\| f_t \|_{L^2(\mu_t)}$. To this end, note that (cf. Lemma (1.6))

$$\frac{d}{dt} \| f_t - 1 \|^2_{L^2(\mu_{\beta(t)})} = -2E_{\beta(t)}(f_t, f_t) + \beta'(t) \int (U - \langle U \rangle_{\mu_{\beta(t)}}) f_t^2 d\mu_{\beta(t)}$$

$$= -2E_{\beta(t)}(f_t, f_t) + \beta'(t) \int (U - \langle U \rangle_{\beta(t)}) (f_t - 1)^2 d\mu_{\beta(t)}$$
$$+ 2\beta'(t) \int (U - \langle U \rangle_{\mu_{\beta(t)}}) (f_t - 1) d\mu_{\beta(t)} \leq [-2\gamma e^{-\beta(t)m}$$
$$+ M\beta'(t)] \| f_t - 1 \|^2_{L^2(\mu_{\beta(t)})} + 2\beta'(t) M \| f_t - 1 \|_{L^2(\mu_{\beta(t)})}.$$

Thus if we take $\beta(t) = 1/m \log(1 + m\gamma t/M)$, and we set $u(t) = \| f_t - 1 \|^2_{L^2(\mu_{\beta(t)})}$, then we see that

$$u'(t) \leq \frac{\gamma M}{M + m\gamma t}(-u + 2u^{1/2});$$

from which it is easy to estimate $\sup_{t \geq 0} \| f_t \|_{L^2(\mu_{\beta(t)})}$ in terms of $\| f_0 \|_{L^2(\mu_0)}$, γ, m, and M.

If one wants to estimate $\| f_t \|_{L^q(\mu_{\beta(t)})}$ for $q > 2$, then one must bring other considerations to bear. In Sect. 1 we give an abstract account of what can be said about $\| f_t \|_{L^\infty(\mu_{\beta(t)})}$ on the basis of the behavior of the Sobolev constant for E_β. Our conclusions are contained in Theorem (1.10). In Sect. 2 we see how (at least when E is finite) one must choose the freezing schedule $t \to \beta(t)$ in order for the results of Sect. 1 to be applicable. What we find (cf. Theorem (2.11) is that we can take $\beta(t) = \log(1 + t)/(m + \varepsilon)$ for any $\varepsilon > 0$, where m is the optimal choice in (0.10). In addition, we calculate m in terms of U and $q_0(x, y)$ (cf. Theorem (2.1)). Finally, in Sect. 3 we replace the Sobolev inequality by a logarithmic Sobolev inequality and see what can then be said. The result here, when applied to finite state spaces, shows that $\| f_t \|_{L^{q(t)}(\mu_{\beta(t)})}$ stays bounded when one chooses $\beta(t) = 1/m \log(1 + \gamma(t))$, where $\lim_{t \to \infty} t^\varepsilon/\gamma(t) = 0$ for every $\varepsilon < 1$ and $q(t)$ grows like $\beta(t)$ (cf. Theorems (3.9) and (3.21) and (3.23)). However, we expect that the considerations in Sect. 3 are only really interesting when one considers E's of arbitrarily large dimension.

1. General Results

The results in this section are based on some differential inequalities which are collected together in the first lemma.

Lemma (1.1). *Let a, b, c, and ε be non-negative numbers and assume that both a and*

ε *are strictly positive. Let* $u \in C^1((0, \infty); [0, \infty))$.
   i) *If*

$$u' \leq \frac{-a}{1+t} u^{1+\varepsilon} + \frac{bu}{1+t} + \frac{cu^{1/2}}{1+t}, \quad t > 0,$$

then

$$u(t) \leq \left[\frac{4\lambda}{a} \frac{1}{1-(1+t)^{-\varepsilon\lambda}} \right]^{1/\varepsilon}, \quad t > 0,$$

where $\lambda \equiv b\backslash(a + c)$.
   ii) *If* $T > s \geq 0$ *and* $u \in C^1([s, T); (0, \infty))$ *satisfies*

$$u' \geq \frac{a}{1+t} u^{1+\varepsilon} - \frac{b}{1+t} u, \quad t \in [s, T),$$

then

$$u(s) \leq \left[\frac{b}{a \left(1 - \left(\dfrac{1+s}{1+T} \right)^{\varepsilon b} \right)} \right]^{1/\varepsilon}.$$

Proof. We begin with i).

$$2(u^{1/2})' \leq \frac{-a}{1+t}(u^{1/2})^{1+2\varepsilon} + \frac{b}{1+t} u^{1/2} + \frac{c}{1+t}$$

$$= -\frac{a}{1+t}(1 + (u^{1/2})^{1+2\varepsilon}) + \frac{bu^{1/2} + a + c}{1+t}$$

$$\leq -\frac{a}{4^\varepsilon(1+t)}(1 + u^{1/2})^{1+2\varepsilon} + \frac{\lambda}{1+t}(1 + u^{1/2}).$$

Set $\omega(t) = (1 + t)^{-\lambda/2}(1 + u^{1/2}(t))$. Then

$$\omega' = (1 + t)^{-\lambda/2}\left[(u^{1/2})' - \frac{\lambda}{2(1+t)}(1 + u^{1/2}) \right]$$

$$\leq -\frac{a}{8(1+t)^{1+\lambda/2}}(1 + u^{1/2})^{1+2\varepsilon}$$

$$= -\frac{a}{8}(1+t)^{\varepsilon\lambda-1}\omega^{1+2\varepsilon},$$

and so

$$(\omega^{-2\varepsilon})' = -2\varepsilon \frac{\omega'}{\omega^{1+2\varepsilon}} \geq \frac{a\varepsilon}{4}(1+t)^{\varepsilon\lambda-1}.$$

Hence

$$\omega(t)^{-2\varepsilon} \geq \frac{a}{4\lambda}[(1+t)^{\varepsilon\lambda} - 1],$$

$$(1+t)^{\varepsilon\lambda} u(t)^{-\varepsilon} \geq \frac{a}{4\lambda}[(1+t)^{\varepsilon\lambda} - 1],$$

$$u(t) \leq \left[\frac{4\lambda}{a} \frac{1}{1-(1+t)^{-\varepsilon\lambda}} \right]^{1/\varepsilon}.$$

To prove ii) we compute as follows:

$$((1+t)^b u)' = (1+t)^b\left[u' + \frac{b}{1+t}u\right] \geqq a(1+t)^{b-1}u^{1+\varepsilon} = a(1+t)^{-\varepsilon b-1}((1+t)^b u)^{1+\varepsilon}.$$

Therefore

$$\frac{d}{dt}[(1+t)^b u]^{-\varepsilon} = -\varepsilon\frac{((1+t)^b u)'}{((1+t)^b u)^{1+\varepsilon}} \leqq -a\varepsilon(1+t)^{-\varepsilon b-1}.$$

Thus

$$-(1+s)^{-b\varepsilon}u(s)^{-\varepsilon} \leqq \frac{a}{b}((1+T)^{-\varepsilon b} - (1+s)^{-\varepsilon b}),$$

i.e.

$$u(s)^{-\varepsilon} \geqq \frac{a}{b}\left(1 + \left(\frac{1+s}{1+T}\right)^{\varepsilon b}\right).$$

ii) follows easily from this.   Q.E.D.

Throughout the rest of this section we assume that the following hypotheses hold.

(1.2) Assumption. U is bounded above and below and $\sup_{E} U - \inf_{E} U \equiv M < \infty$.

(1.3) Assumption. There is a constant $B < \infty$, such that $d\beta(t)/dt \leqq B/(1+t)$.

(1.4) Assumption. For some finite A, some $2 < p < \infty$, and all $0 \leqq t$,

$$\|f - \langle f\rangle_{\beta(t)}\|^2_{L^p(\mu_{\beta(t)})} \leqq A(1+t)E_{\beta(t)}(f,f).$$

We also use the following notation: For $t \geqq 0$, $\|\cdot\|_{p,t}$ is the L^p norm on the space $L^p(\mu_{\beta(t)})$, $\langle\ \rangle_t$ is the integral of f with respect to $\mu_{\beta(t)}$, $E_t = E_{\beta(t)}$, $\mu_t = \mu_{\beta(t)}$, and $\mathbf{L}_t = \mathbf{L}_{\beta(t)}$.

If p is as in Assumption (1.4) then we let $\varepsilon = (p-2)/p$ and note that $p = 2/(1-\varepsilon)$.

(1.5) Lemma. *If (1.4) holds, then for all $f \in L^2(\mu_\beta(t))$,*

$$\|f - \langle f\rangle_t\|^{2+2\varepsilon}_{2,t} \leqq A(1+t)E_t(f,f)\|f - \langle f\rangle_t\|^{2\varepsilon}_{1,t}.$$

Proof. Set $\theta = 1/(1+\varepsilon)$, then $1/2 = \theta/p + (1-\theta)$, and so by Holder's inequality,

$$\|f - \langle f\rangle_t\|_{2,t} \leqq \|f - \langle f\rangle_t\|^\theta_{p,t}\|f - \langle f\rangle_t\|^{1-\theta}_{1,t} \leqq [A(1+t)E_t(f,f)]^{1/(2(1+\varepsilon))}$$
$$\cdot \|f - \langle f\rangle_t\|^{\varepsilon/(1+\varepsilon)}_{1,t}.$$ Q.E.D.

Let $f_0 \in L^1(\mu_0)^+$ with $\langle f_0\rangle_0 = 1$, and set $f_t = (d(m_0 P_{0,t}))/d\mu_t$, where $dm_0 = f_0 d\mu_0$.

(1.6) Lemma. *With f_t defined as above, $d/dt\, f_t = \mathbf{L}_t f_t + \beta_t'(U - \langle U\rangle_t)f_t$.*

Proof. Let $\phi: \mathbf{E} \to (-\infty, \infty)$. Then, denoting $d\mu_t/d\mu_0$ by v_t, we have

$$\frac{d}{dt}\int \phi f_t v_t d\mu_0 = \frac{d}{dt}\int P_{0,t}\phi\, dm_0 = \int P_{0,t}\mathbf{L}_t\phi\, dm_0 = \int(\mathbf{L}_t\phi)f_t d\mu_t$$
$$= \int \phi(\mathbf{L}_t f_t)d\mu_t = \int \phi(\mathbf{L}_t f_t)v_t d\mu_0.$$

Thus $d/dt \, (f_t v_t) = (\mathbf{L}_t f_t) v_t$, and so

$$\frac{d}{dt} f_t = \frac{d}{dt}\left(\frac{f_t v_t}{v_t}\right) = \frac{(\mathbf{L}_t f_t) v_t}{v_t} - \frac{f_t v_t}{v_t^2} v_t' = \mathbf{L}_t f - f_t \frac{v_t'}{v_t}.$$

Finally

$$\frac{v_t'}{v_t} = -\beta_t' U - \frac{Z_t'}{Z_t} = -\beta_t'(U - \langle U \rangle_t). \qquad\qquad \text{Q.E.D.}$$

(1.7) Lemma. *Set* $K_1 = 4(1 + 4\text{ABM})/(1 - e^{-(1/2A + 2BM)})$, *and let* f_t *be as above. Then*

$$\| f_t \|_{2,t} \leqq 1 + K_1^{1/(2\varepsilon)}, \quad t \geqq e^{1/\varepsilon} - 1.$$

Proof. Set $u(t) = \| f_t - 1 \|_{2,t}^2 = \| f_t \|_{2,t}^2 - 1$. Then

$$
\begin{aligned}
u'(t) &= -2E_t(f_t, f_t) + \beta'(t)\int (U - \langle U \rangle_t) f_t^2 d\mu_t \\
&= -2E_t(f_t, f_t) + \beta'(t)\int (U - \langle U \rangle_t)(f_t - 1)^2 d\mu_t + 2\beta' \int (U - \langle U \rangle_t)(f_t - 1) d\mu_t \\
&\leqq -\frac{2}{4^\varepsilon A(1 + t)} u^{1+\varepsilon} + \frac{BM}{1+t} u + \frac{2BM}{1+t} u(t)^{1/2}.
\end{aligned}
$$

Hence the estimate follows from i) in Lemma (1.1).   Q.E.D.

(1.8) Remark. Since $\mathbf{E}$ is a finite set, f_0 is necessarily in $L^2(\mu_0)$, and so we do not need to run the process for time $e^{1/\varepsilon} - 1$ before $f_t \in L^2(\mu_t)$. In fact $\| f_0 \|_{2,0} \leqq 1/\min_{x \in \mathbf{E}} \mu_0(x)^{1/2}$. However, the bound in Lemma (1.7) will become important as $\mathbf{E}$ gets large.

(1.9) Lemma. *Set* $K_2 = 2\text{ABM}e^{2BM}/(1 - e^{-BM})$. *Then for all* $T > e^{1/\varepsilon}$ *and* $\phi \in L_2(\mu_T)$:

$$\| P_{t,T}\phi \|_{2,t} \leqq (K_2^{1/\varepsilon} + e^{2BM/\varepsilon})^{1/2} \| \phi \|_{1,T}$$

if $t = (1 + T)e^{-1/\varepsilon} - 1$.

Proof. We may assume that $\phi \geqq 0$ and that $\langle \phi \rangle_T = 1$. Set $\phi_s = P_{s,T}\varphi$. Then

$$\frac{d}{ds}\langle \phi_s \rangle_s = -\int \mathbf{L}_s \phi d\mu_s - \beta'(s)\int (U - \langle U \rangle_s)\phi_s d\mu_s \geqq -\frac{BM}{1+s}\langle \phi_s \rangle_s,$$

and so

$$\langle \phi_t \rangle_t \leqq \left(\frac{1+T}{1+t}\right)^{BM} \leqq e^{BM/\varepsilon}, \quad (1+t) \geqq e^{-1/\varepsilon}(1+T).$$

Next, set $u(s) = \| \phi_s - \langle \phi_s \rangle_s \|_{2,s}^2 = \| \phi_s \|_{2,s}^2 - \langle \phi_s \rangle_s^2$. Then

$$u'(t) \geqq 2E_t(\phi_t, \phi_t) - \frac{BM}{1+t} u(t).$$

Applying Lemma (1.5) we bound this below by

$$\frac{2e^{-2BM}}{4^\varepsilon A(1+t)} u^{1+\varepsilon} - \frac{BM}{1+t} u(t), \quad (1+t) \geqq e^{-1/\varepsilon}(1+T).$$

Hence, by ii) in Lemma (1.1),

$$u(e^{-1/\varepsilon}(1+T)-1) \leqq \left[\frac{2ABMe^{2BM}}{1+e^{-BM}}\right]^{1/\varepsilon}.\qquad\text{Q.E.D.}$$

We are now ready to prove the main result of this section.

(1.10) Theorem. *Assume that (1.2), (1.3), and (1.4) hold. Set*

$$K_\varepsilon = (1 + K_1^{1/(2\varepsilon)})(K_2^{1/\varepsilon} + e^{2BM_1/\varepsilon})^{1/2}.$$

Then for all $T \geqq e^{2/\varepsilon}$,

$$\|f_T\|_{\infty,T} \leqq K_\varepsilon.$$

Proof. Note that if ϕ is a bounded function on $\mathbf{E}$ and $T \geqq e^{2/\varepsilon}$, then

$$|\int \phi f_T d\mu_T| = |\int (P_{t,T}\phi)f_t d\mu_t| \leqq \|P_{t,T}\phi\|_{2,t}\|f_t\|_{2,t},$$

where $t = e^{-1/\varepsilon}(1+T) - 1 \geqq e^{1/\varepsilon} - 1$.   Q.E.D.

2. The Finite Case

In this section we will check the hypotheses of Theorem (1.10) for the model of simulated annealing given in the introduction. Since the problem and the simulated annealing process remain unchanged if a constant is added to the energy, we assume throughout this section that the minimum of U is 0. This is just to simplify the resulting expressions, and the reader can easily supply the necessary changes if he is bothered by the addition of an unknown constant to the energy.

From Theorem (1.10) we see that we want to take the inverse temperature at time t to be $1/m \log(t+1)$ for some constant $1/m$, and $1/m$ must be chosen so that (1.4) holds. At the same time, we would like to choose $1/m$ as large as we can. From (0.8) we see that $\mathbf{L}_\beta$ is self-adjoint on $L^2(\mu_\beta)$. The constant m is closely related to the gap between 0 and the rest of the spectrum of $\mathbf{L}_\beta$. Thus we denote this distance by $\gamma(\beta)$. An easy argument based on (0.8) shows that $\gamma(\beta) \geqq e^{-\beta M}\gamma(0)$. Although this is qualitatively the sort of estimate which we seek, M is only an upper bound for m. We want to show that one can do better.

For $x, y \in \mathbf{E}$ we define a path from x to y to be any sequence of points $x = x_0, x_1, x_2, \ldots, x_n = y$, such that $q_0(x_{i-1}, x_i) > 0$ (and hence $q_\beta(x_{i-1}, x_i) > 0$) for $i = 1, \ldots, n$. Let $P_{x,y}$ denote the set of paths from x to y. Elements of $P_{x,y}$ will be denoted by $p = \{p_i\}_{i=0}^n$. If we think of the elevation at x as being given by the function U, then the highest elevation along the path $p \in P_{x,y}$ is given by $\text{Elev}(p) = \max\{U(p_i): p_i \in p\}$, and the lowest possible highest elevation along any path from x to y is given by $H(x,y) = \min\{\text{Elev}(p): p \in P_{x,y}\}$. Note that for any $x, y, z \in \mathbf{E}$, $H(x,y) = H(y,x)$ and $H(x,y) \leqq \max\{H(z,x), H(z,y)\}$. Finally let

$$m \equiv \max\{H(x,y) - U(x) - U(y): x, y \in \mathbf{E}\}.\qquad(2.1)$$

Note that if x_0 and y_0 are such that $H(x_0, y_0) - U(x_0) - U(y_0) = m$, then by the above observations either $U(x_0) = 0$ or $U(y_0) = 0$. The intuitive description of m is as follows. For each x_0 with $U(x_0) = 0$ and each $y \in \mathbf{E}$ consider the path from y to

x_0 which requires the least total elevation gain when beginning at y. Pick x_0 and y to make this least total elevation gain as large as possible. Then m is the elevation that must be gained along the route that takes the lowest passes between y and x_0.

Our immediate goal is to prove the following theorem.

Theorem (2.1). *There exist constants $0 < c \leq C < \infty$ such that for all $\beta \geq 0$,*

$$ce^{-\beta m} \leq \gamma(\beta) \leq Ce^{-\beta m}. \tag{2.2}$$

This theorem is an immediate consequence of the next two lemmas.

(2.3) Lemma. *There is a constant $C \leq \infty$ such that for all $\beta \geq 0$,*

$$\gamma(\beta) \leq Ce^{-\beta m}. \tag{2.4}$$

Proof. We first assume that $m > 0$.

Set $\mathrm{Var}_\beta(f) = \sum_{x \in E} f^2(x)\mu_\beta(x) - (\sum_{x \in E} f(x)\mu_\beta(x))^2$. Then since

$$\gamma(\beta) = \inf_{f \in L^2(\mu_\beta)} \frac{-\sum_{x \in E} f(x)\mathbf{L}_\beta f(x)\mu_\beta(x)}{\mathrm{Var}_\beta(f)}, \tag{2.5}$$

we will have proved (2.4) once we exhibit a function F and a constant $C < \infty$, such that for all $\beta \geq 0$,

$$\frac{-\sum_{x \in E} F(x)\mathbf{L}_\beta F(x)\mu_\beta(x)}{\mathrm{Var}_\beta(F)} \leq Ce^{-\beta m}.$$

Let x_0 and y_0 be two points for which $H(x_0, y_0) - U(x_0) - U(y_0) = m$ and set $\mathbf{A} = \{z \in \mathbf{E}: H(y_0, z) < H(y_0, x_0)\}$. Note that $x_0 \notin \mathbf{A}$, and $y_0 \in \mathbf{A}$ (otherwise m would be 0). Note also that if $x \in \mathbf{A}$, and $y \notin \mathbf{A}$, and $q_0(x, y) > 0$, then $U(y) \geq H(y_0, x_0)$ (consider the path $p \in P_{y_0, x}$ for which $H(y_0, x) = \mathrm{Elev}(p)$, and extend it one more step to y). Thus for all $x \in \mathbf{A}$, $y \notin \mathbf{A}$, and all $\beta \geq 0$, $q_\beta(x, y)\mu_\beta(x) = q_\beta(y, x)\mu_\beta(y) \leq \alpha(x, y)e^{-\beta H(y_0, x_0)}/Z(\beta)$. Let $F(x) = \mathbf{I}_\mathbf{A}(x)$, the indicator function of $\mathbf{A}$. Then

$$-\sum_{x \in E} F(x)\mathbf{L}_\beta F(x)\mu_\beta(x) = \tfrac{1}{2}\sum_{x \in E}\sum_{x \in E} q_\beta(x, y)(F(y) - F(x))^2\mu_\beta(x)$$

$$\leq \sum_{x \in A}\sum_{y \notin A} \alpha(x, y)e^{-\beta H(y_0, x_0)}/Z_\beta. \tag{2.6}$$

On the other hand

$$\mathrm{Var}_\beta(F) = \mu_\beta(\mathbf{A})\mu_\beta(\mathbf{A}^c) \geq \mu_0(x_0)\mu_0(y_0)e^{-\beta U(y_0)}e^{-\beta U(x_0)}/Z_\beta^2.$$

Thus there is a constant C which is independent of β such that

$$\gamma(\beta) \leq CZ_\beta e^{-\beta m}.$$

Since $Z_\beta \leq 1$ for all $\beta \geq 0$, (2.4) is proved in case $m > 0$.

If $m = 0$, let F be the indicator function of $\{x: U(x) = \max_{y \in E} U(y)\}$. Then just repeat the above argument with this F.   Q.E.D.

(2.7) Lemma. *There is a constant $c > 0$, such that for all $\beta \geqq 0$,*

$$ce^{-\beta m} \leqq \gamma(\beta). \tag{2.8}$$

Proof. From (2.5) we see that it will be enough to find a constant $c > 0$ such that for all $\beta \geqq 0$ and all $f \in L^2(\mu_\beta)$,

$$\frac{-\sum\limits_{x \in E} f(x) \mathbf{L}_\beta f(x) \mu_\beta(x)}{\mathrm{Var}_\beta(f)} \geqq ce^{-\beta m}. \tag{2.9}$$

We first bound $\mathrm{Var}_\beta(f)$ above as follows:

If $x, y \in \mathbf{E}$ let $p^{x,y} \in P_{x,y}$ be a path from x to y for which $\mathrm{Elev}(p^{x,y}) = H(x, y)$. Let $n(x, y)$ be the length of $p^{x,y}$, and define

$$N \equiv \max_{x,y \in \mathbf{E}} n(x, y).$$

If $z, w \in \mathbf{E}$ let

$$\chi_{z,w}(x, y) = \begin{cases} 1 & \text{if for some } 0 \leqq i < n(x, y) \ p_i^{x,y} = z \text{ and } p_{i+1}^{x,y} = w \\ 0 & \text{otherwise.} \end{cases}$$

Note that if $\alpha(z, w) = 0$, then $\chi_{z,w}(x, y) = 0$ for all $x, y \in \mathbf{E}$. In the following we interpret $\chi_{z,w}(x, y)/\alpha(z, w) = 0$ if $\chi_{z,w}(x, y) = 0$. Now

$$
\begin{aligned}
2\,\mathrm{Var}_\beta(f) &= \sum_{x \in \mathbf{E}} \sum_{y \in \mathbf{E}} (f(y) - f(x))^2 \mu_\beta(y) \mu_\beta(x) \\
&= \sum_{x \in \mathbf{E}} \sum_{y \in \mathbf{E}} \left(\sum_{i=1}^{n(x,y)} f(p_i^{x,y}) - f(p_{i-1}^{x,y}) \right)^2 \mu_\beta(y) \mu_\beta(x) \\
&\leqq \sum_{x \in \mathbf{E}} \sum_{y \in \mathbf{E}} n(x, y) \sum_{i=1}^{n(x,y)} (f(p_i^{x,y}) - f(p_{i-1}^{x,y}))^2 \mu_\beta(y) \mu_\beta(x) \\
&\leqq N \sum_{x \in \mathbf{E}} \sum_{y \in \mathbf{E}} \sum_{z \in \mathbf{E}} \sum_{w \in \mathbf{E}} \chi_{z,w}(x, y)(f(z) - f(w))^2 \alpha(z, w) e^{-\beta U(z) \backslash / U(w)} \frac{\mu_\beta(x) \mu_\beta(y)}{\alpha(z, w) e^{-\beta U(z) \backslash / U(w)}} \\
&\leqq N \left[\max_{z,w \in \mathbf{E}} \sum_{x \in \mathbf{E}} \sum_{y \in \mathbf{E}} \chi_{z,w}(x, y) \frac{\mu_\beta(x) \mu_\beta(y) Z(\beta)}{\alpha(z, w) e^{-\beta U(z) \backslash / U(w)}} \right] \\
&\quad \cdot \sum_{z \in \mathbf{E}} \sum_{w \in \mathbf{E}} \alpha(z, w)(f(z) - f(w))^2 e^{-\beta U(z) \backslash / U(w)}/Z_\beta. \tag{2.10}
\end{aligned}
$$

Also

$$
\begin{aligned}
\chi_{z,w}(x, y) \frac{\mu_\beta(x) \mu_\beta(y) Z_\beta}{\alpha(z, w) e^{-\beta U(z) \backslash / U(w)}} &= \frac{\chi_{z,w}(x, y)}{\alpha(z, w)} \frac{\mu_0(x) \mu_0(y)}{\mathbf{Z}_\beta} e^{\beta[U(z) \backslash / U(w) - U(x) - U(y)]} \\
&\leqq e^{\beta m} \frac{\chi_{z,w}(x, y)}{\alpha(z, w)} \frac{\mu_0(x) \mu_0(y)}{\sum\limits_{v: U(v) = 0} \mu_0(v)}.
\end{aligned}
$$

Thus (2.9) follows from (0.7) and (2.10).   Q.E.D.

(2.11) Theorem. *Let $\varepsilon > 0$, and set $\beta(t) = \log(1 + t)/(m + \varepsilon)$ and $p = 2M/(M - \varepsilon)$, then there is a constant $A < \infty$ such that (1.4) holds.*

Proof. We fix $t \geqq 0$ and set $\beta = \beta(t)$. Then

$$\int (f - \langle f \rangle_\beta)^p d\mu_\beta = \int (f - \langle f \rangle_\beta)^2 |f - \langle f \rangle_\beta|^{p-2} d\mu_\beta \leqq \int (f - \langle f \rangle_\beta)^2 d\mu_\beta \| f - \langle f \rangle_\beta \|_\infty^{p-2}$$

$$\leqq \frac{1}{c} e^{m\beta} E_\beta(f, f) \| f - \langle f \rangle_\beta \|_\infty^{p-2}.$$

Now

$$\| f - \langle f \rangle_\beta \|_\infty^p e^{-\beta M} \min_{x \in E} \mu_0(x) \leqq \| f - \langle f \rangle_\beta \|_\infty^p \min_{x \in E} \frac{e^{-\beta U(x)} \mu_0(x)}{Z_\beta} \leqq \| f - \langle f \rangle_\beta \|_{L^p(\mu_\beta)}^p.$$

Thus

$$\| f - \langle f \rangle_\beta \|_{L^p(\mu_\beta)}^p \leqq \frac{1}{c} e^{m\beta} E_\beta(f, f) e^{\beta(M/p)(p-2)} \frac{\| f - \langle f \rangle_\beta \|_{L^p(\mu_\beta)}^{p-2}}{(\min_{x \in E} \mu_0(x))^{(p-2)/p}}.$$

That is

$$\| f - \langle f \rangle_\beta \|_{L^p(\mu_\beta)}^2 \leqq A e^{\beta(m + (M(p-2)/p))} E_\beta(f, f)$$

for some finite constant A. Since $M(p-2)/p = \varepsilon$, and $\beta = \log(1+t)/(m+\varepsilon)$, we have the desired result. Q.E.D.

Remark. The reasoning used to derive (2.2) may be applicable to a wide range of situations. For example, let S^1 denote the circle of unit length, μ the uniform measure on S^1, and E the standard Dirichlet form

$$E(f, f) = \int_{S^1} |df|^2 d\mu.$$

Given a continuous function $U : S^1 \to [0, \infty)$ whose minimum value is 0, define μ_β accordingly. If x and y are distinct points on S^1, let $A_+(x, y)$ and $A_-(x, y)$ denote the arcs running, respectively, clockwise and counter-clockwise from x to y; and define $H_\pm(x, y)$ to be the maximum of the values that U takes on $A_\pm(x, y)$. Set $H(x, y) = \min\{H_+(x, y), H_-(x, y)\}$ and take $m = \sup\{H(x, y) - U(x) - U(y) : x, y \in S^1\}$. Finally, let $\Phi(\cdot, x, y)$ be the unit speed geodesic running clockwise (counter-clockwise) from x to y if $H_+(x, y) \leqq H_-(x, y)$ (if $H_+(x, y) > H_-(x, y)$). It is then clear that, for each $t \in [1/2, 1]$ and x ($t \in [0, 1/2]$ and y), the map $y \to \Phi(t, x, y)$ ($x \to \Phi(t, x, y)$) is smooth μ-almost everywhere and has Jacobian not less than $1/2$. Hence

$$\text{Var}_\beta(f) = \frac{1}{2} \int_{S^1} \int_{S^1} \left(\int_0^1 |df|^2(\Phi(t, x, y)) \Phi'(t, x, y) dt \right) \mu_\beta(dx) \mu_\beta(dy)$$

$$\leqq \exp(\beta m) \int_0^1 (\int \int |df|^2(\Phi(t, x, y)) \exp(-\beta U(\Phi(t, x, y)) \mu_\beta(dx) \mu_\beta(dy)) dt$$

$$\leqq 2 \exp(\beta m) E(f, f).$$

The problem in more general situations is, of course, that it is not so clear how one should choose the paths $\Phi(\cdot, x, y)$.

3. Logarithmic Sobolev Inequality

We continue to assume that $U_{\min} = 0$.

In this section we will first examine what can be said about simulated annealing

on the basis of a logarithmic Sobolev inequality of the form

$$\int \phi^2 \log\left(\left(\frac{\phi}{\|\phi\|_{2,\beta}}\right)^2\right) d\mu_\beta \leq \alpha(\beta) E_\beta(\phi, \phi). \tag{3.1}$$

We will then discuss what one can expect $\alpha(\beta)$ to look like in various contexts.

Let $\beta: [0, \infty) \to [0, \infty)$ and $q: [0, \infty) \to [2, \infty)$ be smooth non-decreasing functions satisfying $\beta(0) = 0$ and $q(0) = 2$. Using Lemma (1.6) above and following L. Gross [6] (cf. Theorem (9.10) in [9]) we see that

$$\frac{d}{dt}\|f_t\|_{q(t),t} \leq \frac{\|f_t\|_{q(t),t}^{1-q(t)}}{q(t)^2}\left[q'(t)\int f_t^{q(t)} \log\left(\frac{f_t}{\|f_t\|_{q(t),t}}\right)^{q(t)} d\mu_t - 2q(t) E_t(f_t^{q(t)/2}, f_t^{q(t)/2})\right]$$

$$+ \beta'(t)\left(1 - \frac{1}{q(t)}\right)\|f_t\|_{q(t),t}\int\left(\frac{f_t}{\|f_t\|_{q(t),t}}\right)^{q(t)}(U - \langle U \rangle_t) d\mu_t. \tag{3.2}$$

In order to derive a useful differential inequality from (3.2), we need the following lemma.

(3.3) Lemma. *If μ is a probability measure and $\theta \in L^1(\mu)^+$ satisfies $\int \theta d\mu = 1$, then for every $\psi \in L^\infty(\mu)$ satisfying $\int \psi d\mu = 0$, one has*

$$|\int \psi \theta d\mu| \leq 8^{1/2} \|\psi\|_{L^\infty(\mu)} (\int \theta \log \theta d\mu)^{1/2}.$$

Proof. By a theorem of Sanov (see Lemma (3.38) in [9]), for each $\theta \geq 0$ such that $\int \theta(x)\mu(dx) = 1$, we have

$$\int \theta(x) \log(\theta(x)) \mu(dx) = \sup_\phi\{\int \phi(x)\theta(x)\mu(dx) - \log(\int e^{\phi(x)}\mu(dx))\},$$

where the supremum is taken over all bounded measurable functions ϕ. Letting ϕ be of the form $\phi(x) = a\psi(x)$, we see that

$$\int \theta(x) \log(\theta(x)) \mu(dx) \geq \sup_a\{\int a\psi(x)\theta(x)\mu(dx) - \log(\int e^{a\psi(x)}\mu(dx))\}.$$

Note that $\int \psi(x)\mu(dx) = 0$ implies that $\log(\int e^{a\psi(x)}\mu(dx)) \geq 0$ for all a. Thus if in addition $\int \theta(x)\psi(x)\mu(dx) \geq \varepsilon \geq 0$, then we have

$$\int \theta(x) \log(\theta(x)) \mu(dx) \geq \sup_{a \geq 0}\{a\varepsilon - \log(\int e^{a\psi(x)}\mu(dx))\}.$$

Let $\varepsilon = \int \psi(x)\theta(x)\mu(dx)$. By replacing ψ by $-\psi$ if necessary, we may assume that $\varepsilon \geq 0$. We denote $\log(\int e^{a\psi(x)}\mu(dx))$ by $F(a)$.

Define $K(\varepsilon) = \sup_{a \geq 0}\{a\varepsilon - F(a)\}$. Since $F(0) = 0$ and $F'(0) = 0$ and $F(a) \geq 0$ for all a we have $K(0) = 0$ and $K(\varepsilon) > 0$ for all $\varepsilon > 0$. Note that if $G(x) \geq F(x)$ for all $x \geq 0$, then

$$K(\varepsilon) = \sup_{a \geq 0}(\varepsilon a - F(a)) \geq \sup_{a \geq 0}(\varepsilon a - G(a)).$$

Since $F(0) = F'(0) = 0$ and $F''(a) \leq 4\|\psi\|_\infty^2$ for all a, we have $F(a) \leq 2a^2\|\psi\|_\infty^2$ for all a. Thus $K(\varepsilon) \geq \varepsilon^2/(8\|\psi\|_\infty^2)$ for all $\varepsilon > 0$. From the first part of the proof we have $\int \theta(x) \log \theta(x)\mu(dx) \geq K(\varepsilon) \geq \varepsilon^2/(8\|\psi\|_\infty^2)$. (Q.E.D.)

Using Lemma (3.3) with $\mu = \mu_t$, $\theta = (f_t/\|f_t\|_{q(t),t})^{q(t)}$, and $\psi = U - \langle U \rangle_t$, we see

from (3.1) and (3.2) that for any choice of $\rho:[0, \infty) \to (0, \infty)$,

$$\frac{d}{dt} \| f_t \|_{q(t),t} \leqq \frac{\| f_t \|_{q(t),t}^{1-q(t)}}{q(t)^2} \left[(q'(t) + \rho(\beta(t))\beta'(t)q(t)^2)\alpha(\beta(t)) - 2q(t) \right] \qquad (3.4)$$

$$\cdot E_t(f_t^{q(t)/2}, f_t^{q(t)/2}) + \frac{2M^2 \beta'(t)}{\rho(\beta(t))} \| f_t \|_{q(t),t}.$$

Hence, by taking $t \to q(t)$ so that

$$q'(t) + \rho(\beta(t))\beta'(t)q(t)^2 - \frac{2}{\alpha(\beta(t))} q(t) = 0, \qquad (3.5)$$

we get

$$\sup_{t \geqq 0} \| f_t \|_{q(t),t} \leqq \| f_0 \|_{2,0} \exp\left(2M^2 \int_0^\infty \frac{1}{\rho(\beta)} d\beta \right). \qquad (3.6)$$

In particular, we want to take ρ so that $\int_0^\infty (1/\rho(\beta))d\beta < \infty$.

We still need to choose $t \to \beta(t)$ in such a way that $\beta(t)$ and $q(t)$ tend to ∞ as $t \to \infty$. To this end, let $\sigma:[0, \infty) \to (0, \infty)$ be a smooth non-increasing function and determine $t \to \beta(t)$ by

$$\beta'(t) = \frac{\sigma}{\rho\alpha}(\beta(t)) \quad \text{and} \quad \beta(0) = 0.$$

Then

$$\beta(t) = B^{-1}(t), \quad B(\beta) \equiv \int_0^\beta \frac{\rho\alpha}{\sigma}(\xi)d\xi \qquad (3.7)$$

and

$$\frac{1}{q(t)} = \frac{1}{Q(\beta(t))}, \qquad (3.8)$$

$$\frac{1}{Q(\beta)} \equiv e^{-2\Gamma(\beta)} \left[\frac{1}{2} + \int_0^{\Gamma(\beta)} \sigma(\Gamma^{-1}(\gamma))e^{2\gamma}d\gamma \right], \quad \Gamma(\beta) \equiv \int_0^\beta \frac{\rho}{\sigma}(\gamma)d\gamma.$$

We have now proved the following theorem.

(3.9) Theorem. *Assume that (3.1) holds for some non-decreasing smooth $\alpha:[0, \infty) \to (0, \infty)$. Choose smooth functions $\rho:[0, \infty) \to (0, \infty)$ so that ρ is non-decreasing, σ is non-increasing, $R \equiv \int_0^\infty (1/\rho(\beta))d\beta < \infty$ and $\sigma(\beta) \to 0$ as $\beta \to \infty$. Define $t \to \beta(t)$ and $t \to q(t)$ as in (3.7) and (3.8). Then*

$$\| f_t \|_{q(t),t} \leqq \| f_0 \|_{2,0} \exp(2M^2 R), \quad t \geqq 0, \qquad (3.10)$$

where $M \equiv \max_E U - \min_E U$. In particular, if $\alpha(\beta) \leqq A(1 + \beta)^k e^{\beta m}$ for some $A < \infty$, $k \in Z^+$, and $m \in [0, \infty)$ and one takes $\rho(\beta) = (1 + \beta)^2$ and $\sigma(\beta) = (1 + \beta)^{-1}$, respectively,

then $q(t) \geq (1 + \beta(t))/4$ *and*

$$\frac{1}{m}\log\left(1 + \frac{mt}{A}\right) > \beta(t) \geq \begin{cases} \left(1 + \frac{k+4}{2A}t\right)^{1/(k+4)} - 1 & \text{if } t < \frac{A}{m} \\[2em] \frac{1}{m}\log\left[1 + \dfrac{mt}{A\left(\dfrac{1}{m}\log\left(1 + \dfrac{mt}{A}\right)\right)^{k+3}}\right] & \text{if } t \geq \frac{A}{m}. \end{cases}$$

The crucial assumption in the above theorem is the bound assumed for $\alpha(\beta)$. When the state space $\mathbf{E}$ is compact, it is reasonable to assume that

$$\int \phi^2 \log\left(\frac{\phi}{\|\phi\|_{2,0}}\right)^2 d\mu_0 \leq AE_0(\phi, \phi) \tag{3.11}$$

for some $A < \infty$. Indeed, as we will see below, (3.11) is implied by the ordinary Sobolev inequality

$$\|\phi - \langle\phi\rangle_0\|_{p,0}^2 \leq CE_0(\phi, \phi) \tag{3.12}$$

for some $p\in(2, \infty)$ and $C < \infty$; and (3.12) holds in a wide variety of finite dimensional compact situations. (For example, if $\mathbf{E}$ is a compact connected Riemanian manifold of dimension N and if E_0 is the Dirichlet form associated with the Laplacian on $\mathbf{E}$, then (3.12) will hold with $1/p = 1/2 - 1/N$.) On the other hand, there are interesting infinite dimensional examples for which (3.11) holds but (3.12) fails for every $p\in(2, \infty)$. Indeed, (3.11) for μ_0 and E_0 on $\mathbf{E}$ implies the same estimate (with A unchanged) for any situation which results by taking arbitrary tensor products of this one with itself.

The following lemma provides a crude means of passing from (3.11) to an estimate (3.1).

(3.13) Lemma. *If* (3.11) *holds, then* (3.1) *holds with* $\alpha(\beta) \leq Ae^{\beta M}$, *where* $M = \sup_{\mathbf{E}} U - \inf_{\mathbf{E}} U$.

Proof. It is clear from (0.8) that

$$E_0(\phi, \phi) \leq Z_\beta e^{\beta \sup_{\mathbf{E}} U} E_\beta(\phi, \phi).$$

Thus we need only show that

$$\int \phi^2 \log\left(\left(\frac{\phi}{\|\phi\|_{2,\beta}}\right)^2\right) d\mu_\beta \leq \frac{1}{Z_\beta} e^{-\beta \inf_{\mathbf{E}} U} \int \phi^2 \log\left(\left(\frac{\phi}{\|\phi\|_{2,0}}\right)^2\right) d\mu_0. \tag{3.14}$$

But for any probability measure m on $\mathbf{E}$ and any $f: \mathbf{E}\to(-\infty, \infty)$,

$$\int f^2(\xi)\log\left(\frac{f^2(\xi)}{\|f\|_{L^2(m)}^2}\right)m(d\xi) = \inf_{x>0}\int (f^2(\xi)\log(f^2(\xi)) - f^2(\xi)\log(x) - f^2(\xi) + x)m(d\xi),$$

and for each $x > 0$ the integrand on the right side of the above equation is non-negative. Also the infimum on the right side is achieved when $x = \int f^2(\xi)m(d\xi)$. Hence one easily checks that (3.14) holds.   Q.E.D

When (3.11) is all that one knows, then Lemma (3.13) seems to be optimal. However, as we will now show, one can do much better if one knows that (3.12) holds for some $p \in (2, \infty)$ and in addition, that there is a spectral gap estimate of the form:

$$\| \phi - \langle \phi \rangle_\beta \|_{2,\beta}^2 \leqq B e^{\beta m} E_\beta(\phi, \phi), \quad \beta \geqq 0. \tag{3.15}$$

(Since (3.12), itself, implies (3.15) with $B = C$ and $m = M$, we will always assume that the m in (3.15) is in $[0, M]$.) To see what can be said on the basis of (3.12) and (3.15), first note that (by Jensen's inequality)

$$\int \left(\frac{\psi}{\| \psi \|_2} \right)^2 \log \left(\frac{\psi}{\| \psi \|_2} \right)^2 d\mu = \frac{2}{p-2} \int \left(\frac{\psi}{\| \psi \|_2} \right)^2 \log \left(\frac{\psi}{\| \psi \|_2} \right)^{p-2} d\mu$$

$$\leqq \frac{2}{p-2} \log \left(\frac{\| \psi \|_p}{\| \psi \|_2} \right)^p = \frac{p}{p-2} \log \left(\frac{\| \psi \|_p}{\| \psi \|_2} \right)^2.$$

Hence, since $\log x \leqq \delta x + \log 1/\delta$, $x > 0$, for every $\delta > 0$,

$$\int \psi^2 \log \left(\frac{\psi}{\| \psi \|_2} \right)^2 d\mu \leqq \frac{p}{p-2} [\delta \| \psi \|_p^2 + (\log 1/\delta) \| \psi \|_2^2].$$

At the same time, it is easy to check that (3.12) implies

$$\| \phi - \langle \phi \rangle_\beta \|_{p,\beta}^2 \leqq 4 C e^{\beta M} E_\beta(\phi, \phi), \quad \beta \geqq 0.$$

Combining these, taking $\delta = e^{-\beta M}$, and using (3.15), we conclude that

$$\int (\phi - \langle \phi \rangle_\beta)^2 \log \left(\frac{\phi - \langle \phi \rangle_\beta}{\| \phi - \langle \phi \rangle_\beta \|_{2,\beta}} \right)^2 d\mu_\beta \leqq D(1 + \beta) e^{\beta m} E_\beta(\phi, \phi), \quad \beta \geqq 0, \tag{3.16}$$

where $D < \infty$ depends only on B, C, p, and M.

There are various ways to pass from (3.16) to the estimate we want. Perhaps the most efficient route is provided by the following lemma due to Jean-Dominique Deuchel (private communication).

(3.17) Lemma. *For any probability measure μ and any $\phi \in L^2(\mu)$,*

$$\int \phi^2 \log \left(\frac{\phi}{\| \phi \|_{2,\mu}} \right)^2 d\mu \leqq \int (\phi - \langle \phi \rangle_\mu)^2 \log \left(\frac{\phi - \langle \phi \rangle_\mu}{\| \phi - \langle \phi \rangle_\mu \|_{2,\mu}} \right)^2 d\mu + 2 \| \phi - \langle \phi \rangle_\mu \|_{2,\mu}^2.$$

Proof. Let $\phi \in L^2(\mu)$ with $\langle \phi \rangle_\mu = 0$ and $\| \phi \|_{2,\mu} = 1$. What we have to show is that if $\int \phi^2 \log \phi^2 < \infty$, then

$$F(\alpha, \phi, \mu) \equiv \int (\phi + \alpha)^2 \log \left(\frac{(\phi + \alpha)^2}{1 + \alpha^2} \right) d\mu - \int \phi^2 \log \phi^2 d\mu \leqq 2, \tag{3.18}$$

for every $\alpha \neq 0$. By an easy approximation argument, one sees that it suffices to check (3.18) for the class $\Phi(\mu)$ of ϕ's such that $\langle \phi \rangle_\mu = 0$, $\| \phi \|_{2,\mu} = 1$, and $\varepsilon \leqq |\phi| \leqq 1/\varepsilon$ for some $0 < \varepsilon < 1$.

First note that if $\phi \in \Phi(\mu)$ then $1/\phi \in \Phi(\phi^2 \mu)$ and

$$F(\alpha, \phi, \mu) = \alpha^2 F \left(\frac{1}{\alpha}, \frac{1}{\phi}, \phi^2 \mu \right) + 2\alpha \int \phi \log \phi^2 d\mu. \tag{3.19}$$

[1], ι_{2m-1}, and $Q_{I_l}(\iota_{2m-1}) = Q_1 Q_1 \ldots Q_1(\iota_{2m-1})$, where I_l has length l.
Second, check that for any probability measure v and $\psi \in \Phi(v)$:

$$\frac{d}{d\beta} F(\beta, \psi, v) = 2 \int (\psi + \beta) \log \frac{(\psi + \beta)^2}{1 + \beta^2} dv,$$

$$\frac{d^2}{d\beta^2} F(\beta, \phi, v) = 2 \int \log \frac{(\psi + \beta)^2}{1 + \beta^2} dv + \frac{4}{1 + \beta^2} \leq 4,$$

and so:

$$F(\beta, \psi, v) \leq 2\beta \int \psi \log \psi^2 dv + 2\beta^2. \tag{3.20}$$

Taking $v = \phi^2 \mu$, $\psi = 1/\phi$ and $\beta = 1/\alpha$ in (3.20) and using (3.19), we get (3.18).

<div align="right">Q.E.D.</div>

Summarizing, we have now proved the following.

(3.21) Theorem. *Assume that (3.12) and (3.15) hold for some B, $C \in (0, \infty)$, $p \in (2, \infty)$, and $0 \leq m \leq M \equiv \sup_E U - \inf_E U$. Then there is an $A = A(p, B, C, M) < \infty$, such that*

$$\int \phi^2 \log \left(\frac{\phi}{\|\phi\|_{2,\beta}} \right)^2 d\mu_\beta \leq A(1 + \beta) e^{\beta m} E_\beta(\phi, \phi), \quad \beta \geq 0. \tag{3.22}$$

<div align="right">Q.E.D.</div>

We combine Theorem (3.9), Lemma (3.13), Theorem (3.21) and (0.7) for our final result.

(3.23) Theorem. *If (3.11) holds but (3.12) does not hold, let $B(\beta) = A \int_0^\beta (1 + \xi)^3 e^{\xi M} d\xi$.*

If (3.12) and (3.15) hold, let $B(\beta) = A \int_0^\beta (1 + \xi)^4 e^{\xi m}$, where A is as in (3.22). Let $\beta(t) = B^{-1}(t)$ be the cooling schedule for the simulted annealing process. Then if f_0 is the density of the initial distribution with respect to μ_0,

$$\text{Prob}(U(X(t)) \geq U_{\min} + \delta) \leq \|f_0\|_{2,0} e^{2M^2} (\mu_{\beta(t)}(\{x : U(x) \geq U_{\min} + \delta\}))^{\frac{\beta(t)-3}{\beta(t)+1}}.$$

References

1. Chiang, T.-S., Chow, Y.: On eigenvalues and annealing rates, (to appear in Math. Oper. Res.): On the convergence rate of a special class of annealing processes. (to appear in Soochow J. Math.): On the convergence rate of annealing process. Preprint
2. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
3. Geman, S., Hwang C.-R.: Diffusions for global optimization: Preprint (1984)
4. Gidas, B.: Non-stationary Markov chains and convergence of the annealing algorithm. J. Stat. Phys. **39**, 73–131 (1985)
5. Goldstein, L.: Mean square rates of convergence in the continuous time simulated annealing algorithm on R^d. Preprint (1986)
6. Gross, L.: Logarithmic Sobolev inequalities, Am. J. Math. **97**, 1061–1083 (1976)

7. Hajek, B.: Cooling schedules for optimal annealing. Preprint submitted to Math. Oper. Res. (1985)
8. Kirkpatrick, S., Gelett, C. D., Vecchi, M. P.: Optimization by simulated annealing. Science **220**, 621–680 (1983)
9. Stroock, D.: An introduction to the theory of large deviations. Berlin, Heidelberg, New York: Springer 1984