Proceedings of 24th Conference
on Decision and Control
Ft. Lauderdale, FL • December 1985

GLOBAL OPTIMIZATION VIA THE LANGEVIN EQUATION

Basilis Gidas(*)

Division of Applied Mathematics
Brown University
Providence, RI 02912

ABSTRACT

We provide a simple proof of the convergence of the cooling algorithms, i.e., the annealing algorithm and the Langevin equation. The convergence is established for temperature schedules which are very near to optimal ones. Our methods are based on Differential Equations techniques.

I. Introduction

In contrast to the multitude of local optimization methods, there exist only few global optimization techniques. Broadly speaking, global optimization algorithms are divided into deterministic and stochastic algorithms [1,16].

The deterministic methods either they aim at finding the exact global minimum (or minima), or they use heuristic techniques which try to find the global minima by enumerating all local minima. In both cases, there are some drawbacks: the exact methods require a priori knowledge of certain properties of the cost function to be minimized (e.g. the Lipschitz constant), and their computational cost in realistic problems is formidable; the heuristic methods also make implicit assumptions about the cost function and they do not guarantee success.

The stochastic methods are in general more efficient. The main stochastic methods are variants of the so-called multistart technique: If the cost function $f(x)$ is defined in a region $\Omega\subset\mathbb{R}^d$ (usually taken to be convex and compact), then points are sampled from a uniform distribution over Ω, after which local minima are located by a local search procedure. A very interesting multistart stochastic method has been developed in [16]. Stochastic algorithms converge, in general, with probability one (or almost surely).

Recently, a new global optimization technique, the Annealing Algorithm (AA), has been introduced in combinatorial optimization problems [12], image processing problems [2], and machine learning [10]. The annealing algorithm is a variant (with a time-dependent temperature) of the Metropolis algorithm [14], and is appropriate for systems where the state space of the underlying random variable is discrete. Motivated by image processing problems with continuous grey-levels, U. Grenander [7] and S. Geman, proposed the use of the Langevin equation (with time-dependent temperature) as a global minimization algorithm. Independently, G. Parisi [15] proposed the Langevin equation as a tool in Lattice Guage theories. We will refer to this method as the Langevin Algorithm (LA).

We have performed computer experiments using the LA for minimizing certain test function [1], and a one-dimensional function with nearly 400 local minima and one global minimum. We believe that for low dimensional problems, a combination of the LA with the multistart techniques would improve the computational cost of both methods. However, for large dimensional problems (like gauge field theories), the LA appears to be a more efficient method.

The first rigorous result concerning the convergence of the AA was obtained in [2]. In [4], the annealing algorithm was treated as a special case of non-stationary Markov chains, and some optimal annealing schedules and an ergodic theorem were established. Optimal annealing schedules for the AA have recently been obtained by nice intuitive arguments in [8].

A convergence theorem for the LA in bounded domains of $\mathbb{R}^d$, $d\geq 1$, was first obtained in [3]. In [13] (also [11]), the method of Large Deviations was applied to obtain results in the entire of $\mathbb{R}^d$. In [6] (see also [5]), we use methods from Partial Differential Equations to analyse the Fokker-Planck equation associated with the LA. We have established [6] the convergence of the LA in the entire of $\mathbb{R}^d$, and have determined (in certain cases) the optimal temperature schedules.

Here we present an elementary version of the techniques of [6], and establish the convergence of the cooling algorithms (i.e., the AA and the LA) for temperature schedules which are not optimal, but very near to the optimal ones.

II. The Annealing Algorithm

We shall treat the continuous time case only. Similar arguments apply to the discrete time case. We begin by considering a non-stationary Markov chain with finite state space $\Omega = (s_1,\dots,s_n)$, and transition rate matrix $\mathcal{L}(t)$. Let $\pi(t) = (\pi_1,\dots,\pi_n)$ be a probability vector with $\pi_j(t) > 0$, $j = 1,\dots,n$, satisfying the "detailed balance" relation

$$\pi_i\mathcal{L}_{ij} = \pi_j\mathcal{L}_{ji}. \tag{2.1}$$

The matrix $\mathcal{L}$ satisfies

$$\sum_j\mathcal{L}_{ij}(t) = 0. \tag{2.2}$$

The transition matrix $P(t)$ defined by

$$P_{ij}(t) = P\{X(t) = s_j|X(0) = s_i\}$$

satisfies

$$\frac{dP}{dt} = -P(t)\mathcal{L}(t),\ P_{ij}(0) = \delta_{ij}. \tag{2.3}$$

We assume that the entries of $\mathcal{L}(t)$ are bounded, so that there exists a solution of (2.3) satisfying

$$\sum_j P_{ij}(t) = 1.$$

It is easily verified that the matrix $\mathcal{L}$ is a Hermitean matrix on $\ell^2(\Omega)$ with weight $\pi(t)$, i.e. for any two

(*) Partially supported by NSF Grant DMS 85-16230 and U.S. ARO DAAG-29-83-K-0116.

functions f and g on Ω, we have

$$\sum_i \pi_i f_i (\mathcal{L}g)_i = \sum_i \pi_i (\mathcal{L}f)_i g_i \tag{2.4}$$

where $f_i = f(s_i)$, $g_i = g(s_i)$, $i = 1,\ldots,n$. Therefore, $\mathcal{L}$ has real eigenvalues $\lambda_i(t)$, $i = 1,\ldots,n$, which we may arrange in increasing order $\lambda_1 < \lambda_2 \le \lambda_3 \le \cdots \le \lambda_n$. The first eigenvalue is zero, i.e. $\lambda_1(t) = 0$, and the corresponding normalized eigenvector is $\varphi_1 = (1,\ldots,1)$. The second eigenvalue is strictly positive, and plays a crucial role in our analysis. Here is our basic theorem.

Theorem A

Consider a non-stationary Markov chain as above. Suppose that

$$\left|\frac{d\pi_j(t)}{dt}\right| \le \gamma(t)\pi_j(t), \quad j = 1,\ldots,n \tag{2.5}$$

where $\gamma(t)$ is independent of j. Then, if

$$\int_0^{+\infty} \lambda_2(t)dt = +\infty \tag{2.6}$$

and

$$\lim_{t\to+\infty} \frac{\gamma(t)}{\lambda_2(t)} = 0 \tag{2.7}$$

then

$$\sum_j |P_{ij}(t) - \pi_j(t)| \to 0 \quad \text{as} \quad t \to +\infty. \tag{2.8}$$

Proof: Set

$$P_{ij}(t) = \pi_j(t)Q_{ij}(t)$$

then

$$\sum_j |P_{ij} - \pi_j| \le (\sum_j \pi_j (Q_{ij}-1)^2)^{1/2}. \tag{2.9}$$

We will show that the right-hand side of (2.9) goes to zero as $t \to \infty$. For a fixed i, we set $Q_{ij}(t) = q_j(t)$, and easily derive the equation

$$\frac{dq_j}{dt} = -(\mathcal{L}q)_j - \frac{d\pi_j}{dt}\pi_j^{-1}q_j. \tag{2.10}$$

From this we obtain

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\sum_{j=1}^n \pi_j(q_j-1)^2 &= -\sum_j \pi_j q_j(\mathcal{L}q)_j - \frac{1}{2}\sum_j \frac{d\pi_j}{dt}q_j^2 \\ &= -\sum_j \pi_j(q_j-1)(\mathcal{L}(q-\varphi_1))_j - \frac{1}{2}\sum_j \frac{d\pi_j}{dt}q_j^2.\end{aligned} \tag{2.11}$$

Note that $q-\varphi_1$ is orthogonal to the eigenspace spanned by $\varphi_1 = (1,\ldots,1)$. Since $\lambda_2(t)$ is the smallest eigenvalue of $\mathcal{L}$ in this orthogonal subspace, we have

$$\sum_j \pi_j(q_j-1)(\mathcal{L}(q-\varphi_1))_j \ge \lambda_2(t)\sum_j \pi_j(q_j-1)^2. \tag{2.12}$$

Let $F(t) = \sum_j \pi_j(q_j-1)^2$. Then using (2.12) and bounding the last term in (2.11) by (2.5), we obtain

$$\frac{1}{2}\frac{dF(t)}{dt} \le -\lambda_2(t)F(t) + \frac{1}{2}\gamma(t)F(t) + \frac{1}{2}\gamma(t). \tag{2.13}$$

If $\lambda_2(t)$ and $\gamma(t)$ satisfy (2.6) and (2.7), then (2.13) implies $F(t) \to 0$ as $t \to +\infty$. This completes the proof of the theorem.

Next, we apply Theorem A to the annealing algorithm. Suppose that a function f defined on Ω is to be minimized. Suppose $U_1,\ldots,U_m$ are the positive values of f on Ω, ordered so that $U_1 < U_2 < \ldots < U_m$. The non-diagonal elements $\mathcal{L}_{ij}$, $i \ne j$ of are given by

$$\mathcal{L}_{ij}(t) = \begin{cases} R_{ij} & \text{if } f_j \le f_i \\ R_{ij}e^{-\frac{f_j-f_i}{T(t)}} & \text{if } f_j > f_i \end{cases} \tag{2.14}$$

where R_{ij} is a transition matrix of an arbitrary symmetric $(R_{ij} = R_{ji})$ and irreducible Markov chain (this matrix is called the "proposal" matrix). The temperature $T(t)$ is assumed to go to zero monotonically as $t \to \infty$. The probability vector $\pi_j(t)$ is given by

$$\pi_j(t) = \frac{e^{-\frac{1}{T(t)}f_j}}{\sum_i e^{-\frac{1}{T(t)}f_i}}. \tag{2.15}$$

This yields

$$\frac{d\pi_j}{dt} = \frac{T'}{T^2}(f_j - \overline{f}(t))\pi_j(t) \tag{2.16}$$

where

$$\overline{f}(t) = \sum_j \pi_j(t)f_j.$$

Thus

$$\left|\frac{d\pi_j}{dt}\right| \le \gamma(t)\pi_j(t) \tag{2.17}$$

with

$$\gamma(t) = -\frac{T'}{T^2}(U_m - U_1). \tag{2.18}$$

Under the above conditions on the proposal matrix R_{ij} and π we can show that there exists a constant Δ such that

$$\lambda_2(T(t)) = \text{const } e^{-\frac{\Delta}{T(t)}} + O\left(e^{-\frac{2\Delta}{T}}\right) \tag{2.19}$$

as $t \to +\infty$. This is established by the methods of [18], and the constant Δ has an explicit representation in terms of graphs as in [18]. If

$$T(t) \ge \frac{c_0}{\log t}, \quad t \text{ large} \tag{2.20}$$

with $c_0 > \Delta$, then $\lambda_2(T(t))$ and $\gamma(t)$ (given by (2.18)) satisfy (2.6) and (2.7). Thus one obtains the convergence of the annealing algorithm for temperature schedules satisfying (2.20) with $c_0 > \Delta$.

Remarks 1) The above result can be generalized in two ways: first, it holds [6] for a more general class of proposal matrices R, and second it holds for other variants [4] of the annealing of algorithm.

2) The schedules (2.20) are not optimal (in fact the above proof fails even for $c_0 = \Delta$), but they are much better than the ones in [2]. In [6], we have shown that for the annealing algorithm (under appropriate

proposal matrices), condition (2.6) is necessary and sufficient for convergence. In [6], we again start with equation (2.10), but we perform a diagonalization of the matrix $\mathcal{L}(t)$.

III. The Langevin Equation

The Langevin equation (LE) in $\mathbb{R}^d$, $d \geq 1$ reads

$$dX(t) = \sqrt{2T}\, dw(t) - \nabla f(X(t))dt \tag{3.1}$$

where $w(t)$ is the standard Brownian motion in $\mathbb{R}^d$, and T is the temperature. We assume that the function is C^2, bounded below, and grows at infinity properly (see below). Equation (3.1) defines a d-dimensional diffusion process $X(t)$.

Equation (3.1) was proposed by Langevin in 1908 as a generalization of Einstein's theory of Brownian motion, and it describes the motion of a particle in a viscuous fluid (Brownian motion assumes zero viscosity). Langevin's equation was the first mathematical equation describing non-equilibrium thermodynamics. The first term, $\sqrt{2T}\, dw(t)$, corresponds to microscopic fluctuations caused by the Brownian force, while the second term, $-\nabla f dt$, called the "drag" force, is generated by the viscosity of the fluid. For T independent of t, the LE has two time sclaes: the short time scale where the Brownian fluctuations dominate, and the long time scale where the drag force dominates.

Onsager gave a new interpretation to the LE as an irreversible process by a simple correspondence of the two terms: The velocity of the particle is interpreted as the deviation of a thermodynamic quantity from its equilibrium, while the drag force is interpreted as the "drift" of the thermodynamic system towards its equilibrium. In Onsager's interpretation, the thermal fluctuations dominate in the short time, while the drift dominate in the long time scale.

When T depends on t, and $T(t) \to 0$ (at an appropriate rate) as $t \to +\infty$, the LE may be interpreted as the decay of "metastable" states onto the ground state(s) of the thermodynamic system. In this case, the LE has many time scales, corresponding to the lengths of time the non-stationary diffusion process takes to get out of local minima and eventually fall into a global minimum. Our mathematical analysis isolates all these time scales - the worst of which determines the best annealing schedule.

Let

$$p(t,x) = P\{X(t) = x \mid X(t_0) = x_0\} \tag{3.2}$$

be the transition function corresponding to (3.1) with $T = T(t)$. It satisfies the "forward" (or Fokker-Planck) equation

$$\frac{\partial p}{\partial t} = T(t)\Delta p + \nabla(\nabla f p), \quad p(t,x)\big|_{t \downarrow t_0} = \delta(x-x_0). \tag{3.3}$$

Let

$$\pi_T(x) = \frac{e^{-\frac{1}{T}f(x)}}{\int e^{-\frac{1}{T}f(x)}dx}$$

and $\pi(t,x) = \pi_{T(t)}(x)$. We transform (3.3) by introducing

$$p(t,x) = \pi(t,x)q(t,x) = (\pi(t,x))^{1/2}r(t,x).$$

A straightforward computation yields

$$\frac{\partial q}{\partial t} = T\Delta q - \nabla f\cdot\nabla q - \frac{T'}{T^2}(f(x)-\bar{f}(t))q \tag{3.4$_a$}$$

$$= T\pi^{-1}\nabla(\pi\nabla q) - \frac{T'}{T^2}(f(x)-\bar{f}(t))q \tag{3.4$_b$}$$

and

$$\frac{\partial r}{\partial t} = T\Delta r - \frac{T}{(2T)^2}\{|\nabla f|^2 - 2T\Delta f\}r - \frac{1}{2}\frac{T'}{T^2}(f(x)-\bar{f}(t))r, \tag{3.5}$$

where $T' = \frac{dT}{dt}$, and $\bar{f}(t) = \int \pi(t,x)f(x)dx$. The spectrum and the eigenfunctions of the operator

$$\mathcal{L}(t) = -T(t)\Delta + \nabla f\cdot\nabla \tag{3.6}$$

appearing in (3.4), play a crucial role in our analysis. This operator is self-adjoint on $L_2(\mathbb{R}^d, \pi dx)$, and it is unitary equivalent to the Schrödinger-type operator

$$H = -T\Delta + \frac{T}{(2T)^2}\{|\nabla f|^2 - 2T\Delta f\} \tag{3.7}$$

appearing in (3.5). We have $\mathcal{L} = \pi^{-\frac{1}{2}} H \pi^{\frac{1}{2}}$. The operator H is self-adjoint on $L_2(\mathbb{R}^d)$. Under the assumptions of Theorem B below, H (and hence $\mathcal{L}$) has a discrete spectrum $\lambda_1(T) < \lambda_2(T) \leq \lambda_3(T) \leq \ldots$. We note that $\lambda_1(T) = 0$, and the corresponding normalized eigenvector of $\mathcal{L}$ is 1 (and of H, $\pi^{1/2}$).

<u>Theorem B</u>

Let $f(x)$, $x \in \mathbb{R}^d$, $d \geq 1$, be a C^2 function. Assume that $f(x)$ and $|\nabla f|$ go to infinity as $|x| \to +\infty$, and that $f(x)$ and $-\frac{\Delta f}{1+|\nabla f|^2}$ are bounded below. Furthermore, assume that the critical set of $f(x)$ consists of isolated non-degenerate points. Then, if $T(t)$ goes to zero monotonically as $t \to +\infty$, and

$$\int_{t_0}^{+\infty} \lambda_2(T(t))dt = +\infty \tag{3.8}$$

$$-\frac{T'}{T}\frac{1}{\lambda_2(T(t))} \to 0 \quad \text{as} \quad t \to +\infty \tag{3.9}$$

then in the weak sense

$$|p(t,\cdot)| - \pi(t,\cdot)| \to 0 \quad \text{as} \quad t \to +\infty. \tag{3.10}$$

<u>Remarks</u> 1) Note that condition (3.9) contains $-\frac{T'}{T}$, while condition (2.2) for the annealing algorithm contains $-\frac{T'}{T^2}$.

2) Following the same strategy as here but with more complicated technical estimates, we have shown [6] that condition (3.8) is necessary and sufficient for (3.10).

3) In [6], we have shown that $\lambda_2(T)$ converges to zero exponentially as $T \downarrow 0$, and have obtained the asymptotic behavior of $\lambda_2(T)$ as $T \downarrow 0$. These type of problems are standard in the "classical" limit $h \downarrow 0$ (Planck's constant) of Schrödinger type operators [9,17]. Under the assumptions of Theorem B, the asymptotic behavior of $\lambda_2(T)$ is as follows: Let $a_1, \ldots, a_N$ be the minima of $f(x)$, and for simplicity we assume that a_N is the only global minimum. For each a_i, $i = 1, \ldots, N-1$, we consider curves $\gamma(s)$, $s \in [0,1]$ such that $\gamma(0) = a_i$, $\gamma(1) = a_N$, and define

$$\Delta_i = \min_{\gamma} \max_{s\in[0,1]} f(\gamma(s))-f(a_i) \quad (3.11_a)$$

$$\Delta = \max_{i=1,\dots,N-1} \Delta_i \quad . \quad (3.11_b)$$

In one dimension ($d = 1$), or in $d \geq 2$ but with $N = 2$, we have

$$\lambda_2(T) = \text{const } e^{-\frac{\Delta}{T}} + O\, e^{-\frac{2\Delta}{T}} \quad . \quad (3.12)$$

In these cases, conditions (3.8) and (3.9) hold for

$$T(t) \geq \frac{\Delta}{\log t}\ , \text{ large } t\ . \quad (3.13)$$

The constant Δ is the optimal constant for schedules of the forms $T(t) = \frac{c}{\log t}$, t large. Note that for the annealing algorithm we established convergence only for $c > \Delta$. The stronger result here is obtained by an appropriate use of the maximum principle for parabolic equations.

Proof of Theorem A Let $\eta(x)$ be a smooth function of compact support. Then

$$\left|\int p(t,x)\eta(x)dx - \int \pi(t,x)\eta(x)dx\right| \leq \left(\int \pi\eta^2\right)^{\frac{1}{2}}\left(\int \pi(q-1)^2\right)^{\frac{1}{2}}.$$

We will show that

$$\int \pi(t,x)(q(t,x)-1)^2 dx \to 0 \quad \text{as} \quad t \to +\infty. \quad (3.14)$$

From (3.4) we obtain

$$\frac{1}{2}\frac{d}{dt}\int \pi(q-1)^2 dx = -T\int q\nabla(\pi\nabla q)dx - \frac{1}{2}\frac{T'}{T^2}\int \pi(f(x)-\overline{f}(t))q^2 dx. \quad (3.15)$$

We will derive below the following a priori estimate: There exists a $\theta(t)$ tending to zero as $t \to +\infty$ such that

$$q(t,x) \leq c\ e^{\theta(t)f(x)} \quad (3.16)$$

for some constant c. This estimate together with the fact that $|\nabla q| \in L_2(\mathbb{R}^d, \pi(t,x)dx)$ (which is easily obtained from (3.4)), allow us to integrate by parts and obtain

$$-\int q\nabla(\pi\nabla q)dx = \int \pi q^2 dx = \int \pi|\nabla(q-1)|^2 dx. \quad (3.17)$$

The growth assumptions of $f(x)$, allow us to apply a standard Poincaré inequality for weighted spaces, and obtain

$$T\int \pi|\nabla(q-1)|^2 dx \geq \lambda_2(T)\int \pi(q-1)^2 dx \quad (3.18)$$

noting that the average $\int \pi q dx$ of q is equal to 1. Now, we use estimate (3.16) to bound the last term in (3.15) as follows

$$-\frac{T'}{T^2}\int \pi(f(x)-\overline{f}(t))q^2 dx \leq -c^2\frac{T'}{T^2}\int \pi(f-\overline{f}(t))e^{2\theta(t)f(x)}dx \quad (3.19_a)$$

$$\leq -\tilde{c}\ \frac{T'}{T}\ . \quad (3.19_b)$$

In the last step above we have used the fact (easily established) that the integral on the right-hand side of (3.19_a) goes to zero like T. Inserting (3.18) and (3.19_b) into (3.15), and setting $F(t) = \int \pi(q-1)^2 dx$, we obtain

$$\frac{1}{2}\frac{dF}{dt} \leq -\lambda_2(T(t))F(t) - \frac{1}{2}\tilde{c}\ \frac{T'}{T}. \quad (3.20)$$

This together with (3.8) and (3.9) yield that $F(t) \to 0$ as $t \to +\infty$, which in turn proves (3.14), and hence (3.10). It remains to prove (3.16). We set

$$q(t,x) = \varphi(t,x)e^{\theta(t)f(x)}. \quad (3.21)$$

Then

$$L\varphi \equiv \frac{\partial\varphi}{\partial t} - T(t)\Delta\varphi + (1-2\theta T)\nabla f\cdot\nabla\varphi + c(t,x)\varphi = 0 \quad (3.22_a)$$

$$c(t,x) = -[\theta|\nabla f|^2 + \theta'(t)f(x) + \frac{T'}{T^2}(f-\overline{f}(t)) - \theta^2 T|\nabla f|^2 - \theta T\Delta f] \quad (3.22_b)$$

Next, we choose $\theta(t)$ so that $\theta(t) \to 0$ monotonically as $t \to +\infty$, and $c(t,x) \geq 0$. If this is achieved, then the Maximum Principle for parabolic equations applies and we conclude that $\varphi(t,x) \leq c$ for a sufficiently large constant c. This gives (3.16). First we note that since $\lambda_2(T)$ goes to zero exponentially with T, condition (3.9) implies that $\frac{T'}{T^2}$ goes to zero as $t \to +\infty$. For large enough $|x|$, we can choose $\theta(t) = -c_1\frac{T'}{T^2}$, with c_1 such that the term $\theta|\nabla f|^2$ dominates the term $\frac{T'}{T^2}(f(x)-\overline{f}(t))$ (the remaining terms are of lower order). In fact this choice of $\theta(t)$ fails only in a small neighborhood of the critical points of $f(x)$ (which are assumed to be isolated and non-degenerate). If x_0 is a critical point, then the neighborhood where we modify our choice of $\theta(t)$ is given by $|x-x_0| \leq c\sqrt{T}$ because of the non-degeneracy of the point x_0. In this neighborhood, we choose $\theta(t)$ so that $-\theta T\Delta f$ dominates the term $\frac{T'}{T^2}(f(x)-\overline{f}(t))$. If the point is local maximum, we choose $\theta = -c_2\frac{T'}{T^3}$ (note again that $\frac{T'}{T^3} \to 0$ as $t \to +\infty$), with some appropriate constant $c_2 > 0$. If the point is a local minimum, we choose $\theta_2 = c_2\frac{T'}{T^3} \leq 0$. Note that in the neighborhood $|x-b| \leq c\sqrt{T}$ of a global minimum b, $|f(x)-\overline{f}(t)| \leq \tilde{c}\ T$, and we may take $\theta = -c\frac{T'}{T^2}$. Thus, we have chosen a (not necessarily continuous) $\theta(t)$ such that $c(t,x) \geq 0$, which completes the proof of the theorem.

REFERENCES

1. Dixon, L.C.W. and G.P. Szegö (eds.): Towards Global Optimization 2, North-Holland (1978).
2. Geman, S. and D. Geman: "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", IEEE Transactions, PAMI 6 (1984), 721-741.
3. Geman, S. and C.R. Hwang: "Diffusions for Global Optimization", to appear SIAM Jour. on Control and Optimization (1985).
4. Gidas, B.: "Non-Stationary Markov Chains and Convergence of the Annealing Algorithm", J. Stat. Phys. 39 (1985), 73-131.

5. Gidas, B.: "The Langevin Equation as a Global Minimization Algorithm", to appear in Disordered Systems and Biological Organization, Springer-Verlag (1985), eds. E. Bienenstock, et. al.
6. Gidas, B.: "Global Minimization via the Langevin Equation", in preparation.
7. Grenander, U.: Tutorial in Pattern Theory, Brown University (1983).
8. Hajek, B.: "Cooling Schedules for Optimal Annealing", to appear in Mathematics of Oper. Research.
9. Helffer, B. and I. Sjöstrand: "Multiple Wells in the Semi-classical Unit I", Comm. Part. Diff. Equ. 9 (1984), 337-408.
10. Hinton, G., T. Sejnowski and D. Ackley: "Boltzmann Machine: Constraint Satisfaction Networks that Learn", preprint 1984.
11. Chiang, T.S., C.R. Hwang and S.J. Sheu: "Diffusion for Global Optimization in $\mathbb{R}^n$", preprint (received August 1985).
12. Kirkpatrick, S., C.D. Gebatt, and M. Vecchi: "Optimization by Simulated Annealing", Science 220, 13 May (1983), 621-680.
13. Kushner, H.: "Asymptotic Behavior for Stochastic Approximations and Diffusion with slowly decreasing noise effects: Global Minimization via Monte Carlo", preprint (1985).
14. Metropolis, W., et. al.: "Equations of State Calculations by Fast Computing Machines", J. Chem. Phys. 21 (1953), 1087-1091.
15. Parisi, G.: "Prolegomena to any Further Computer Evaluation of the QCD Mass Spectrum", in Progress in Gauge Field Theory, Cargese (1983).
16. Rinnooy Kan, A.H.G., C.G.E. Boender and G.T. Timmer: "A Stochastic Approach to Global Optimization", preprint (1984).
17. Simon, B.: "Semiclassical Analysis of Low Lying Eigenvalues I. Non-degenerate Minima: Asymptotic Expansions", Ann. Inst. Henri Poincaré 38 (1983), 295-307.
18. Ventcel, A.D.: "On the Asymptotics of Eigenvalues of Matrices with Elements of Order $\exp\{-v_{ij}/2\varepsilon^2\}$", Dokl. Akad. Nauk SSR 202 (1972), 65-68.