TR-87-45

A Combined Multistart-Annealing Algorithm for Continuous Global Optimization

by

Mauro Piccioni

A Combined Multistart-Annealing Algorithm for Continuous Global Optimization

Mauro Piccioni

Systems Research Center
The University of Maryland
College Park, Maryland 20742

and

Dipartimento di Matematica
II Università di Roma "Tor Vergata"
00173 Roma, Italy

Abstract

The application of simulated annealing to the global optimization of a function on a compact subset of $\Re^d$ is discussed. For the Langevin algorithm the class of convergent schedules depends on some a prior knowledge about the form of the function. It is shown that this problem disappears by using the simplest annealing algorithm of jump type, which can also be improved by performing local searches between two consecutive jump times. The resulting algorithm is essentially a multistart technique controlled by an annealing schedule.

Key Words: Monte Carlo Methods, Simulated Annealing, global Optimization.

This research was supported in part by the National Science Foundation under grant number CDR-85-00108 to The University of Maryland Systems Research Center.

1 Introduction

The simulated annealing approach has been recently proposed to design stochastic algorithms for general combinatorial optimization problems, by Kirkpatrick et al. [1]. It has been successfully applied to some particular problems by Bonomi and Lutton [2], Cerny [3] and Kirkpatrick [4], and various convergence proofs have already been given by Gelfand and Mitter [5], Geman and Geman [6], Gidas [7], Hajek [8] , Mitra et al [9] and Tsitsiklis [10]. In this paper we are interested in exploiting the same ideas to propose a new stochastic algorithm for the most classic continuous optimization problem, namely finding a global minimum point of a function on a compact subset of $\Re^d$.

Let us start with a brief discussion of such approach for the discrete case. We will consider only continuous time algorithms because they allow a simpler analysis. Let K be a finite set and V any real function defined on it. For any $\tau > 0$, which we will call the temperature by analogy with statistical mechanics, let us define the Boltzmann distribution on K

$$\pi^\tau(x) \propto e^{-V(x)/\tau}, x \in K. \tag{1}$$

If it is desired to simulate a sample from (1) a direct evaluation of V on the whole K is of course impractical when $|K|$ is large. This difficulty motivated Metropolis et al. [11] to introduce the following algorithm for simulation.

Let $(X_t^\tau, t \geq 0)$ be a homogeneous irreducible Markov process on K with transition function P^τ and define the intensity matrix R^τ by

$$R^\tau(x,y) = \frac{d}{dt} P^\tau(t,x,\{y\})(0), x \neq y.$$

Then, if the following detailed balance condition w.r.t. π^τ holds

$$e^{-V(x)/\tau} R^\tau(x,y) = e^{-V(y)/\tau} R^\tau(y,x), x \neq y \tag{2}$$

it results that π^τ is the limiting distribution of (X_t^τ), as $t \to \infty$, for any initial distribution. The Metropolis' algorithm simulates a sample path of (X_t^τ) with R^τ obtained from any symmetric irreducible intensity matrix S by

$$R^\tau(x,y) = e^{-[V(y)-V(x)]^+/\tau} S(x,y), x \neq y. \tag{3}$$

This has the meaning of decreasing the chance of transitions which increase the value of V, as τ decreases; it is immediately seen that (2) is satisfied, so that it is expected that after a sufficiently long period of time, the distribution of (X_t^τ) will be close to (1).

It is easy to verify that the distribution (1) converges weakly, as $\tau \to 0$, to a uniform distribution over the set Λ of global minimum points of V. Therefore, for τ small, the problem of sampling from (1) becomes nearly equivalent to the problem of finding a global minimum point for V.

The idea of simulated annealing is to let the temperature τ decrease to zero as the simulation of the Metropolis' algorithm proceeds. Let T be any temperature schedule, i.e. a function mapping $[0, +\infty)$ into $(0, +\infty)$. It is expected that if $T(t)$ decreases slowly

enough to zero as $t \to \infty$, the distribution of the non-homogeneous Markov process (X_t^T) with intensity $R^{T(t)}$ at time $t \geq 0$ will have enough time to approach the Boltzmann distribution with temperature $T(t)$ for t large and will finally reach a limit concentrated on Λ, as $t \to \infty$.

Necessary and sufficient conditions have been given on the schedule T to determine a process (X_t^T) converging in probability to Λ. In particular local minimum points and their depths are defined in terms of the undirected graph induced by non-zero intensities of S. It turns out that T is convergent in the above sense if and only if $\int_0^{+\infty} e^{-d^*/T(t)}dt = +\infty$, where d^* is the maximum depth of a local minimum point which is not global [8]. Of course the set of convergent schedules enlarges as more non-zero intensities are allowed in S, but at the price of a growing complexity for the simulation of each jump. So in combinatorial optimization problems the one-jump transitions are usually constrained within an a priori specified graph G; e.g., when $K = \{0,1\}^D$ for some finite set D, a reasonable choice will be to define $(x,y) \in G$ if and only if there exists a unique $i \in D$ such that $x(i) \neq y(i)$. Once the graph induced by S is specified, the canonical choice is to take $S(x,y) = |G[x]|^{-1}$, where $G[x] = \{y : (x,y) \in G\}$. As a consequence, in order to prescribe convergent schedules for simulation, it is needed to analyze the specific problem under investigation to get an upper estimate for d^*.

2 Annealing Analysis of Pure Random Search

In this section the simulated annealing approach will be applied to the problem of minimizing a continuous function V on a compact subset K of $\Re^d$ with dense interior. No assumption is made on the structure of the set Λ of global minimum points. Without losing any generality it is assumed that the minimum value of V in K is zero, and the Lebesgue measure (denoted by m) of K is 1.

For each $\tau > 0$ the corresponding Boltzmann distribution on K is defined by the density

$$\frac{d\pi^\tau}{dm}(x) \propto e^{-V(x)/\tau} \tag{4}$$

It is possible to show that again π^τ tends to be concentrated on Λ, as $\tau \to 0$; more precisely $V(U^\tau)$ converges, in probability to zero, if U^τ is distributed according π^τ. Moreover, under some regularity assumptions there is a weak limit of π^τ having Λ as support. This makes reasonable to follow the same argument of the previous section to design stochastic algorithms for our global optimization problem. However, we will not rely on π^τ for our convergence proof because a more general approach is possible.

The basic step is again the choice of a family of Markov generators satisfying the general detailed balance condition (see Gardiner [12]) w.r.t. the distribution π^τ in (4). Of course we have the choice between a diffusion and a jump model. In the former case the canonical example is the Langevin algorithm, for which some convergence results have already been established by Chiang et al. [13], Geman and Hwang [14], Gidas [15] and Kushner [16]. In the next section we will show that, even if such a model is certainly interesting from a purely mathematical standpoint, it is not equally meaningful for application.

For this reason, we devote our attention to the simplest family of generators of jump Markov processes on K, defined for any bounded measurable function f on K

$$(A^\tau f)(x) = \int_K e^{-[V(y)-V(x)]^+/\tau}[f(y) - f(x)]dy \tag{5}$$

for $\tau > 0$, which is the most immediate extension of (3) in our case, with the canonical choice $S(x,y) \equiv 1$ corresponding to have the limiting jump process as $\tau \to \infty$ distributed according a Poisson measure on $[0, +\infty) \times K$. Of course the underlying assumption is that the complexity of sampling uniformly in K is negligible, which makes the problem entirely different from the combinatorial optimization case.

The rest of the section is devoted to prove that, when the temperature τ in (5) is decreased to zero during the simulation, in any case the function V evaluated on the sample path of a Markov process with such a time-varying intensity will converge in probability to zero. Let us begin with the construction of a Markov process with intensity $A^{T(t)}$ at time $t \geq 0$, on the probability space $(\Omega, \Im, P)$, a product of mutually independent random samples $T_i, Z_i, i = 1, 2, \cdots$, respectively negative exponential with mean 1 and uniform in $[0, 1]$, and $Y_i, i = 0, 1, 2, \cdots$, uniform in K. Let us define $X_0^T = Y_0,\ S_0 = 0,\ S_i = \Sigma_{j=1}^i T_j, i = 1, 2, \cdots$, and consider the process $(X_t^T, t \geq 0)$ defined by

$$X_t^T = X_{S_i}^T,\ S_i \leq t < S_{i+1},\ i = 0, 1, \cdots \tag{6}$$

$$X^T_{S_{i+1}} = X^T_{S_i} + (Y_{i+1} - X^T_{S_i}) 1_{\{z: z \le e^{-[V(Y_{i+1}) - V(X^T_{S_i})]^+/T(S_{i+1})}\}}(Z_{i+1}) \tag{7}$$

It is easily verified that (X^T_t) is a non-homogeneous Markov process with intensity $A^{T(t)}$ at time $t \ge 0$, that is for each bounded measurable function $f : K \to R$ the process

$$M^f_t = f(X^T_t) - \int_0^t (A^{T(\varsigma)} f)(X^T_\varsigma) d\varsigma \tag{8}$$

is a martingale. From (8) the main result follows.

THEOREM 1. For any schedule T such that $T(t) \to 0$ as $t \to \infty$, the process $V(X^T_t), (X^T_t)$ being defined in (6)-(7), converges to zero in probability.
Proof. We are going to show that for any $\delta > 0$ the function

$$\Gamma^\delta(t) = P(V(X^T_t) > \delta)$$

converges to zero as $t \to \infty$. In fact, let us substitute $f = 1_{V_\delta}$ in (8), where $V_\delta = \{y : V(y) > \delta\}$. Since for any probabilty measure μ on K

$$\langle A^\tau 1_{V_\delta}, \mu \rangle = \int_{[V_\delta \times (K \setminus V_\delta)] \cup [(K \setminus V_\delta) \times V_\delta]} (-1)^{1_{V_\delta}}(x) e^{-[V(y) - V(x)]^+/\tau} (\mu \times m)(dx, dy) =$$

$$= \int_{V_\delta} e^{-V(y)/\tau} dy \int_{K \setminus V_\delta} e^{+V(x)/\tau} \mu(dx) - \mu(V_\delta)(1 - m(V_\delta)) <$$

$$\le \int_{V_\delta} e^{-[V(y) - \delta]/\tau} dy (1 - \mu(V_\delta)) - \mu(V_\delta)(1 - m(V_\delta))$$

By taking expectations in (8) it is obtained that for $t \ge 0$

$$\Gamma^\delta(t) \le \Gamma^\delta(0) + \int_0^t g(T(\varsigma)) d\varsigma - \int_0^t [g(T(\varsigma)) + (1 - m(V_\delta))] \Gamma^\delta(\varsigma) d\varsigma, \tag{9}$$

where $g(\tau) = \int_{V_\delta} e^{-[V(y) - \delta]/\tau} dy$, which goes to zero as $\tau \to 0$ by bounded convergence theorem. By continuity of V it is $m(V_\delta) < 1$ for any $\delta > 0$, so that, by applying Gronwall's inequality, the proof is finished. □

The conclusion of the above theorem is not surprising if it is considered that the algorithm (6)-(7) has a limiting version when $T(t) \equiv 0$, which is obtained by replacing (7) by

$$X^0_{S_{i+1}} = X^0_{S_i} + (Y_{S_{i+1}} - X^0_{S_i}) 1_{\{y: V(y) \le V(X^0_{S_i})\}}(Y_{i+1}) \tag{10}$$

It is almost clear that $V(X^0_t)$ converges a.s. to zero as $t \to \infty$, because for $\delta > 0$

$$P(V(X^0_t) > \delta) = m(V_\delta) e^{-(1 - m(V_\delta))t} \tag{11}$$

As a consequence, for the algorithm (6)-(7) there is no problem connected with the effect of decreasing τ too quickly, as actually stated in Theorem 1. Moreover for any schedule T, for any $t \ge 0$

$$\inf_{0 \le s \le t} V(X^T_s) = V(X^0_t) \tag{12}$$

which means that there is no need to choose positive temperatures once it is possible, as it is reasonable, to keep track of the minimum value of V along the observed sample path. The algorithm (6)-(10) is well known in the literature as the pure random search (see Dixon and Szegö [17]). We leave to the last section the discussion of how to exploit also positive temperatures by modifying (6)-(7) to get, under additional smoothness, a more efficient algorithm.

3 Spectral Analysis and the Langevin Algorithm

The particularly simple structure of the generator (5) has allowed us not to use explicitly the detailed balance condition during the proof of Theorem 1. However, the fact that such a condition plays a fundamental role for the analysis of the Langevin algorithm, requires a brief digression.

It is immediately verified that, since its intensity satisfies (2) the Markov generator A^τ has a continuous self-adjoint extension on $L^2(\pi^\tau)$ (which will be called again A^τ), the Hilbert space of measurable functions on K such that $\int_K f^2(x)e^{-V(x)/\tau}dx < +\infty$, with the scalar product

$$(f,g)_\tau = \alpha(\tau)^{-1}\int_K f(x)g(x)e^{-V(x)/\tau}dx,$$

where $\alpha(\tau) = \int_K e^{-V(x)/\tau}dx$. By consequence, for fixed $\tau > 0$, if X_0^τ has an absolutely continuous distribution w.r.t. π^τ with density in $L^2(\pi^\tau)$, the distribution of X_t^τ will have the same property with density $e^{A^\tau t}p_0$. Since A^τ is a contraction and π^τ is an invariant measure, 0 is the first eigenvalue, with constant eigenvector. The rest of the spectrum is of course contained in the negative real axis, and is bounded from above by

$$\lambda(A^\tau) = \inf\{\mu \le 0 : (A^\tau f, f)_\tau \le \mu \parallel f \parallel_\tau^2, f \in L^2(\pi^\tau), (f,1)_\tau = 0\}, \tag{13}$$

so that $\parallel e^{A^\tau t} - 1 \parallel_\tau = 0(e^{-\lambda(A^\tau)t})$. It seems therefore useful to compute $\lambda(A^\tau)$, which is done in the next.

THEOREM 2. For any $\tau > 0$ it is

$$-\lambda(A^\tau) \ge \alpha(\tau); \tag{14}$$

moreover if there exists at least one global interior minimum $\bar{x}$ around which V is C^2, then

$$\alpha(\tau) \ge 0(\tau^{d/2}). \tag{15}$$

Proof. For (14) let us compute

$$\begin{aligned}
-(A^\tau f, f)_\tau &= -\alpha(\tau)^{-1}\int_K\int_K e^{-\max(V(x),V(y))/\tau}(f(y)-f(x))f(x)dydx = \\
&= [2\alpha(\tau)]^{-1}\int_K\int_K (f(y)-f(x))^2 e^{-\max(V(x),V(y))/\tau}dydx \ge \\
&\ge [2\alpha(\tau)]^{-1}\int_K\int_K (f(y)-f(x))^2 e^{-V(x)/\tau}e^{-V(y)/\tau}dydx = \\
&= \alpha(\tau)(\parallel f \parallel_\tau^2 - (f,1)_\tau^2).
\end{aligned}$$

Now assume that B_ρ is a ball of radius ρ around $\bar{x}$ on which V is C^2, so that $H(x) \le \omega I$ on B_ρ, too; then by a Taylor series expansion there exists a map ς of B_ρ into itself such that

$$\alpha(\tau) \ge \int_{B_\rho} e^{-V(x)/\tau}dx = \int_{B_\rho} e^{-(x-\bar{x})^T H(\varsigma(x))(x-\bar{x})/2\tau}dx \ge$$

$$\geq \int_{B_\rho} e^{-\omega\|x-\bar{x}\|^2/2\tau} dx = (2\pi)^{d/2}(\tau/\omega)^{d/2}\chi_d(\frac{\omega}{\tau}\rho^2),$$

where χ_d is the distribution function of a chi-square random variable with d degrees of freedom. Since $\chi_d(\mu)$ goes to 1 as $\mu \to \infty$, (15) is established. $\square$

It is very intuitive that the asymptotic behavior of $\lambda(A^{T(t)})$, as $t \to \infty$, will play a major role also for the asymptotics of the simulated annealing jump algorithm (6)-(7). This could be made rigorous by following the same argument developed in [15] for both the discrete case and the Langevin algorithm. In fact such analysis depends only on the fact that the detailed balance condition essentially implies the generation of a self-adjoint semigroup on $L^2(\pi^\tau)$ (see Graham [18]). However, in our case, this leads to results weaker than Theorem 1.

The Langevin algorithm has been considered up to now as the only stochastic algorithm for global optimization on $\Re^d$ based on the simulated annealing approach [5] [16], [15],; it requires V to be C^2 and choses the following diffusion for simulation

$$d\tilde{X}_t^T = -(\nabla V)(\tilde{X}_t^T)dt + \sqrt{2T(t)}\ dW_t \tag{16}$$

to which a normal reflection on the boundary must be added for the optimization on K [14]. The analogy with (6)-(7) is that for a fixed temperature $T(t) \equiv \tau > 0$ the Markov generator $\tilde{A}^\tau$ of (16) satisfies the detailed balance condition w.r.t. π^τ (of course, this has a different form for diffusions, see [12]). By consequence the analysis at the beginning of the section is true also for $\tilde{A}^\tau$ and so the computation of $\lambda(\tilde{A}^\tau)$ is of major interest.

However, it is not easy to obtain even an upper bound; the asymptotic behavior, as $\tau \to 0$, have been computed only in few cases, in particular when $K = \Re^1$ and there is a unique strong global minimum point by Matkowsky and Schuss [19]. It turns out that $\lambda(\tilde{A}^\tau) = 0(e^{-d^*/\tau})$, where d^* is defined as in the discrete case; now it is in general impossible to have a tight estimate for it. This result has allowed to prove that, at least for the above case, schedules of the type $T(t) = c/\log t$ for large t , with $c \geq d^*$, guarantee that $V(\tilde{X}_t^T)$ converges to zero in probability [15], which is the best convergence result for the Langevin algorithm up to now. In these cases it is generally believed that these are actually the convergent schedules more quickly converging to zero. Under weaker conditions the asymptotic behavior of $\lambda(A^\tau)$ as given by Theorem 2 shows a better rate of convergence for τ small, and by the results of [15], also for varying temperatures. So, even neglecting the difficulty of simulating a sample path of a reflected diffusion, the jump algorithm (6)-(7) seems to be more appropriate than (16) as an algorithm for global minimization. Moreover, even when $K = \Re^d$ it is easily checked that, by replacing the Lebesgue measure by a probability measure equivalent to it, Theorem 1 and 2 remain true.

4 Improvement of the Pure Random Search

The pure random search was proposed for continuous function not necessarily differentiable. If V is C^2 then for any schedule T the algorithm (6)-(7) (or (6-10)) can be improved in the following way. Let $\phi(t;x)$ be the flow of the vector field $-\nabla V$, stopped at the exit time from K, if any. It is clear that

$$V(\phi(t;x)) \leq V(x), t \geq 0. \tag{17}$$

Now the annealing algorithm (6)-(7) is modified by allowing local searches along the above flow. More generally any vector field $-H\nabla V$, where H is positive-semidefinite valued will ensure (17), which is the unique property we are going to use. The resulting algorithm, once a schedule T is fixed, produces the following random process on the space $(\Omega, \Im, P)$: $\bar{X}_0^T = Y_0$ and

$$\bar{X}_t^T = \phi(\bar{X}_{S_i}^T; t - S_i), \quad S_i \leq t < S_{i+1} \quad i = 0, 1, \cdots \tag{18}$$

$$\bar{X}_{S_{i+1}}^T = \bar{X}_{S_{i+1}-}^T + (Y_{i+1} - \bar{X}_{S_{i+1}}^T) 1_{\{z: z \leq e^{-[V(Y_{i+1}) - V(\bar{X}_{S_{i+1}-}^T)]^+/T(S_{i+1})}\}}(Z_{i+1}) \tag{19}$$

Of course also zero temperatures can be equally well used, by defining for $T(S_{i+1}) = 0$

$$\bar{X}_{S_i+1}^T = \bar{X}_{S_{i+1}-}^T + (Y_{i+1} - \bar{X}_{S_{i+1}-}^T) 1_{\{y: V(y) \leq V(\bar{X}_{S_{i+1}-}^T)\}}(Y_{i+1}) \tag{20}$$

A complete convergence analysis of this algorithm is not the object of the present paper. Any improvement on the rate (11) will of course depend on the form of V, and the following can be regarded as a worst-case analysis. It results that the algorithm (18)-(19)-(20) cannot perform worse than (6)-(7)-(10) as it is shown by the following.

THEOREM 3. For any sample in $(\Omega, \Im, P)$, for any T mapping $[0, +\infty)$ into itself, it is

$$V(\bar{X}_t^T) \leq V(X_t^T). \tag{21}$$

Proof. Of course it suffices to prove such a statement only for $t = S_i, i = 0, 1, \cdots$ For $i = 0$ this is guaranteed by $X_0^T = \bar{X}_0^T = Y_0$. Now suppose that (21) is true at time $t = S_n$, and show that it remains true for $t = S_{n+1}$. By assumption $V(\bar{X}_{S_{n+1}-}^T) \leq V(X_{S_{n+1}-}^T)$, because of (17). Now let $T(S_{n+1}) = \bar{\tau}$ and observe that either $\bar{X}_{S_{n+1}}^T = \bar{X}_{S_{n+1}-}^T$, in which case $V(\bar{X}_{S_{n+1}}^T) \leq V(Y_{n+1})$ and therefore $V(\bar{X}_{S_{n+1}}^T) \leq V(X_{S_{n+1}}^T)$, or $\bar{X}_{S_{n+1}}^T = Y_{n+1}$, in which case:

a) either $\bar{\tau} = 0$, so that

$$V(Y_{n+1}) = V(\bar{X}_{S_{n+1}}^T) = V(X_{S_{n+1}}^T);$$

b) or it must be

$$Z_{i+1} \leq e^{-[V(Y_{n+1}) - V(\bar{X}_{S_{n+1}-}^T)]^+/\bar{\tau}} \leq e^{-[V(Y_{n+1}) - V(X_{S_{n+1}-}^T)]^+/\bar{\tau}},$$

from which $X_{S_{n+1}}^T = Y_{n+1}$, too, and again $V(X_{S_{n+1}}^T) = V(\bar{X}_{S_{n+1}}^T)$.□

As a consequence of the previous theorem and Theorem 1, the convergence in probability of $V(X_t^T)$ to zero is obtained for any schedule T such that $T(t) \to 0$ as $t \to \infty$. The choice of the schedule reflects the relative importance of local searches w.r.t. the proposals for jump. So the fact that it has to approach zero as the simulation proceeds is very intuitive. The optimal choice of T will rely on a more sharp convergence analysis, perhaps resulting in a Bayesian modelization of the a priori knowledge available on V, as suggested by Betrò [20].

The proposed algorithm resembles the behavior of one of the most used stochastic algorithm for global optimization, called multistart. Its simplest version is in fact a series of local searches started from points uniformly sampled in K, each one independently of the others [17]. In our algorithms these searches are stopped according to an annealing schedule, giving rise to a seemingly promising combination of both approaches. Of course, these conclusions must be validated by numerical results, which are currently under progress. These results will be compared with those obtained by Aluffi-Pentini et al. [21] by running the Langevin algorithm more than once, each time with a different constant temperature.

Acknowledgements

The Author is grateful to Prof. S.K. Mitter for having addressed him to the problem and to Prof. J.S. Baras for having invited him to the Systems Research Center, University of Maryland at College Park, where the present work has been performed.

References

[1] Kirkpatrick, S., Gelatt, Jr., C. D. & Vecchi, M. P., "Optimization by simulated annealing," *Science 220*, (1983), pp. 621-680.

[2] Bonomi, E. & Lutton, J.-L., "The N-city travelling salesman problem: statistical mechanics and the Metropolis algorithm," *SIAM Review 26*, (1984), pp. 551-568.

[3] Cerny, V., "Thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm," *J. Opt. Theor. Appl. 45*, (1985), pp. 41-51.

[4] Kirkpatrick, S., "Optimization by simulated annealing," *J. Stat. Phys. 34*, (1984), pp. 975-986.

[5] Gelfand, S. B. & Mitter, S. K., "Analysis of simulated annealing for optimization," *Proc. 24th CDC*, Ft. Lauderdale (1985), IEEE, pp. 779-786.

[6] Geman, S. & Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Rec. Mach. Intell. 6*, (1984), pp. 721-741.

[7] Gidas, B., "Non stationary Markov chains and the convergence of the annealing algorithm," *J. Stat. Phys. 39*, (1985), pp. 73-131.

[8] Hajek, B., "Cooling schedules for optimal annealing," to appear on Math. Oper. Res.

[9] Mitra, D., Romeo, F. & Sangiovanni-Vincentelli, A., "Convergence and finite-time behavior of simulated annealing," *Proc. 24th CDC*, Ft. Lauderdale (1985), IEEE, pp. 761-767.

[10] Tsitsiklis, J., "Markov chains with rare transitions and simulated annealing," Cambridge, 1985, preprint, Laboratory for Information and Decision Systems, MIT.

[11] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E., "Equations of state calculations by fast computing machines," *J. Chem. Phys. 21*, (1953), pp. 1087-1091 .

[12] Gardiner, C. W., *Handbook of Stochastic Methods*, Berlin (1983).

[13] Chiang, T. S., Hwang, C. R. & Sheu, S. J., "Diffusion for global optimization in $\Re^n$," preprint.

[14] Geman, S. & Hwang, C. R., "Diffusions for global optimization," to appear on SIAM J. Contr. Opt.

[15] Gidas, B., "Global optimization via the Langevin equation," *Proc. 24th CDC*, Ft. Lauderdale (1985), IEEE, pp. 774-778.

[16] Kushner, H. J., "Asymptotic behaviour for stochastic approximations and diffusion with slowly decreasing noise: global minimization via Monte Carlo," *LCDS Rep. 85-7*, Div. App. Math., Brown University, (1985).

[17] Dixon, L. C. W. & Szegö, G. P., "The global optimization problem: an introduction," in *Towards Global Optimization 2, pp. 1-15*, North-Holland, Amsterdam, 1978.

[18] Graham, R., "Solution of the Fokker-Planck equations with and without manifest detailed balance," *Z. Phys. B 40*, (1980), pp. 1499-1505.

[19] Matkowski, B. J. & Schuss, Z., "Eigenvalues of the Fokker-Planck operator and the approach to equilibrium for diffusions in potential fields," *SIAM J. Appl. Math. 40*, (1981), pp. 242-254.

[20] Betrò, B., "Bayesian testing of nonparametric hypotheses and its application to global optimization," *J. Opt. Theor. Appl. 42*, (1984), pp. 31-50.

[21] Aluffi-Pentini, F., Parisi, V. & Zirilli, F., "Global optimization and stochastic differential equations," *J. Opt. Theor. Appl. 47*, (1985), pp. 1-16.