

DIFFUSIONS FOR GLOBAL OPTIMIZATION*

STUART GEMAN† AND CHII-RUEY HWANG‡

Abstract. We seek a global minimum of $U: [0, 1]^n \rightarrow R$. The solution to $(d/dt)x_t = -\nabla U(x_t)$ will find local minima. The solution to $dx_t = -\nabla U(x_t) dt + \sqrt{2T} dw_t$, where w is standard (n -dimensional) Brownian motion and the boundaries are reflecting, will concentrate near the global minima of U , at least when "temperature" T is small: the equilibrium distribution for x_t is Gibbs with density $\pi_T(x) \propto \exp\{-U(x)/T\}$. This suggests setting $T = T(t) \downarrow 0$ to find the global minima of U . We give conditions on $U(x)$ and $T(t)$ such that the solution to $dx_t = -\nabla U(x_t) dt + \sqrt{2T} dw_t$ converges weakly to a distribution concentrated on the global minima of U .

Key words. global optimization, simulated annealing, diffusion, reflecting boundaries

AMS(MOS) subject classifications. 60J60, 60J70

1. Introduction. We can find a local minimum of a function U on R^n by starting at an arbitrary $x_0 \in R^n$ and solving the equation

$$\frac{dx_t}{dt} = -\nabla U(x_t).$$

A continuous path, x , seeking a *global* minimum will in general be forced to "climb hills" as well as follow down-hill gradients. One way of introducing hill-climbing, while preserving the tendency to descend along gradients, is to introduce random fluctuations into the path of x :

$$(1.1) \quad dx_t = -\nabla U(x_t) dt + \sqrt{2T} dw_t$$

where w is a standard Brownian motion and T , the "temperature," controls the magnitude of the random fluctuations. Under suitable conditions on U , x_t approaches (weakly) an equilibrium, which is a *Gibbs distribution* with density

$$\pi_T(x) = \frac{1}{Z_T} \exp\{-U(x)/T\} \quad \text{where } Z_T = \int_{R^n} \exp\{-U(x)/T\} dx.$$

As $T \rightarrow 0$, π_T concentrates on the global minima of U . Hence, in low temperature equilibrium we can expect to find x_t near a global minimum.

Unfortunately, the time required to approach equilibrium increases exponentially with $1/T$; solutions to (1.1) with small T will be very slow to find the important minima of U . This suggests that (1.1) be integrated with a gradually decreasing temperature, $T = T(t) \downarrow 0$. The hope is that the early and large random fluctuations will allow x_t to quickly escape from local minima, whereas the later (large t) behavior will be essentially a gradient descent into a prominent minimum of U .

The theorem presented here gives sufficient conditions on U and $T(t)$ for the weak convergence of x_t to a measure concentrating on the global minimum of U . We have simplified the mathematics by confining x to a rectangle in R^n (the diffusion is "reflected at the boundaries"). The rectangle is taken for convenience to be the unit

* Received by the editors March 4, 1985, and in revised form July 17, 1985.

† Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912. This work was supported in part by the National Science Foundation under grants DMS-8352087 and DMS-8306507, and the U.S. Army Research Office under grant DAAG29-83-K-0116.

‡ Institute of Mathematics, Academia Sinica, Taipei, Taiwan, Republic of China.

cube. For illustration, let us assume that $U: [0, 1]^n \rightarrow \mathcal{R}$ has a unique global minimum at $x = \xi$. If U is sufficiently smooth, and properly-behaved at the boundaries (see § 2), and if $T(t) = c/\log(2+t)$ for c sufficiently large, then the solution to (1.1), with $T = T(t)$, converges to ξ :

$$P(|x_t - \xi| < \varepsilon) \rightarrow 1$$

for all $\varepsilon > 0$ and all starting points.¹

Our work was inspired by the “simulated annealing” recently proposed by Černý [2] and Kirkpatrick et al. [10]. Given a function U of n binary variables x_1, \dots, x_n they propose to find global minima of U by running the “Metropolis algorithm” [13] while gradually lowering the temperature. The Metropolis algorithm produces a Markov process with state space $\{0, 1\}^n$. As in (1.1), there is a “temperature”, T , and at fixed T the Metropolis algorithm also has the Gibbs distribution as equilibrium. The same heuristics, then, motivate gradually lowering $T = T(t)$. This is called simulated annealing since it copies the physical procedure, called annealing, of melting and then slowly cooling a physical substance (such as a crystal) in search of a low energy configuration. The latter typically corresponds to a high degree of spatial regularity, useful for some applications. Černý and Kirkpatrick apply their simulated annealing to certain combinatorial optimization problems, often with striking success.

Simulated annealing has also played a role in overcoming some of the computational problems that arise in image processing (Geman and Geman [4], Grenander [8], Marroquin [12]). In these applications, the procedure is modified to accommodate arbitrary discrete variables x_1, \dots, x_n with finite state spaces (rather than binary), and Geman and Geman have established weak convergence to the global minima of U , provided again that the temperature is lowered sufficiently slowly. Unfortunately, the extension of the Metropolis algorithm to *continuous* variables, x_1, \dots, x_n , involves some awkward computational problems. Nevertheless, many of the variables that arise in image processing are most naturally modelled as continuous, such as pixel grey levels, line orientations, and the sizes and orientations of objects. This motivated both Grenander (in [8]) and us to look at a diffusion-process alternative. In future image processing experiments, we will be comparing the computational performance of the continuous-valued Metropolis scheme to the diffusion scheme presented here.

Some encouraging simulation results have been recently obtained by Aluffi-Pentini, Parisi, and Zirilli [1]. They study the performance of a modified version of (1.1), which includes repeated runs, and an interactive “annealing schedule” $T = T(t)$. The experiments involve 22 different test functions U . These are defined on \mathcal{R}^n , with n ranging from one to fourteen, and have multiple local minima. Properly tuned, the algorithm finds a global minimum for each test function.

2. Statement of result. Given a real-valued function U on the unit cube

$$U: [0, 1]^n \rightarrow \mathcal{R},$$

and an “annealing schedule” $T(t) \downarrow 0$, we define a diffusion x :

$$dx_t = -\nabla U(x_t) dt + \sqrt{2T(t)} dw_t$$

¹ B. Gidas [6] and H. Kushner [11] have recently improved on our result. Gidas gets a tight characterization of the minimum allowed c in the schedule $T(t) = c/\log(2+t)$, and removes the reflecting boundaries. Kushner generalizes to a richer class of diffusions, allowing state-dependent diffusion coefficients and a random drift. The latter makes the connection to “stochastic approximation” in which U , or its functionals, cannot be directly observed.

where $w_t \in R^n$ is standard Brownian motion. x is confined to $[0, 1]^n$ “by reflection,” which will be made precise shortly. The theorem gives conditions on $T(t)$ and U which insure the convergence of x_t to the set of global minima of U , in a suitable (weak) sense. Conditions on $T(t)$ will be given later. As for U , the conditions include:

- (A) There exists an extension of U to an open set $S \supseteq [0, 1]^n$, which is twice continuously differentiable, and whose gradient has zero normal component at all noncorner boundary elements of $[0, 1]^n$.²

There are many equivalent ways to make precise the notion of a reflected diffusion. We will proceed in a manner that best fits with the methods to be used later in the proof of the theorem. First, we extend U “periodically” to \hat{U} , defined on all of R^n . Let Z denote the integers, and for every $(i_1, \dots, i_n) \in Z^n$ define

$$S_{i_1, \dots, i_n} = \prod_{k=1}^n [i_k, i_k + 1]$$

and define $G_{i_1, \dots, i_n}: [0, 1]^n \rightarrow S_{i_1, \dots, i_n}$ by

$$(G_{i_1, \dots, i_n}(x))_k = \begin{cases} i_k + x, & i_k \text{ even,} \\ i_k + 1 - x, & i_k \text{ odd.} \end{cases}$$

Finally, define $\hat{U}: R^n \rightarrow R$ by

$$x \in S_{i_1, \dots, i_n} \Rightarrow \hat{U}(x) = U(G_{i_1, \dots, i_n}^{-1}(x)).$$

If x is “on a boundary” (i.e. $x_k = l$ some $l \in Z, 1 \leq k \leq n$), then x is an element of two or more cubes: for example $x \in S_{i_1, \dots, i_n}$ and $x \in S_{j_1, \dots, j_n}$ where $(i_1, \dots, i_n) \neq (j_1, \dots, j_n)$. But then

$$G_{i_1, \dots, i_n}^{-1}(x) = G_{j_1, \dots, j_n}^{-1}(x),$$

and hence \hat{U} is well-defined.

The definition of the reflected process, x , is in terms of a “free” (ordinary diffusion) process \hat{x} :

$$d\hat{x}_t = -\nabla \hat{U}(\hat{x}_t) dt + \sqrt{2T(t)} dw_t.$$

Fix $t \geq s \geq 0$ and $x \in [0, 1]^n$. The conditional distribution on x_t given $x_s = x$ is the same as if we had set $\hat{x}_s = x$ and then defined $x_t = G_{i_1, \dots, i_n}^{-1}(\hat{x}_t)$ whenever $\hat{x}_t \in S_{i_1, \dots, i_n}$. In other words, we reflect \hat{x}_t at the boundaries of the unit cube. More precisely, let $\hat{p}(s, x, t, y)$ be transition probability densities for the process \hat{x} (density on x_t evaluated at y , given that $x_s = x$). Then x is the Markov process with the following transition probability densities:

$$p(s, x, t, y) = \sum_{i_1, \dots, i_n} \hat{p}(s, x, t, G_{i_1, \dots, i_n}(y))$$

$\forall x, y \in [0, 1]^n, t > s \geq 0$. To check that these actually satisfy the Markov property, first observe that for any $(i_1, \dots, i_n), (j_1, \dots, j_n) \in Z^n, x, y \in [0, 1]^n$, and $t > s \geq 0$,

$$\hat{p}(s, G_{j_1, \dots, j_n}(x), t, G_{i_1, \dots, i_n}(y)) = \hat{p}(s, x, t, G_{k_1, \dots, k_n}(y))$$

² We believe, but are not certain, that the theorem still holds when the normal component of ∇U does not vanish on the boundaries of $[0, 1]^n$.

where

$$k_p = \begin{cases} i_p - j_p & \text{if } j_p \text{ is even,} \\ j_p - i_p & \text{if } j_p \text{ is odd,} \end{cases}$$

for each $1 \leq p \leq n$. Thus, for any $x, y \in [0, 1]^n$, $t > s \geq 0$, and $t_0 \in (s, t)$,

$$\begin{aligned} p(s, x, t, y) &= \sum_{i_1, \dots, i_n} \hat{p}(s, x, t, G_{i_1, \dots, i_n}(y)) \\ &= \sum_{i_1, \dots, i_n} \int_{\mathbb{R}^n} \hat{p}(s, x, t_0, z) \hat{p}(t_0, z, t, G_{i_1, \dots, i_n}(y)) dz \\ &= \sum_{i_1, \dots, i_n} \sum_{j_1, \dots, j_n} \int_{[0,1]^n} \hat{p}(s, x, t_0, G_{j_1, \dots, j_n}(z)) \hat{p}(t_0, G_{j_1, \dots, j_n}(z), t, G_{i_1, \dots, i_n}(y)) dz \\ &= \sum_{j_1, \dots, j_n} \int_{[0,1]^n} \hat{p}(s, x, t_0, G_{j_1, \dots, j_n}(z)) \sum_{i_1, \dots, i_n} \hat{p}(t_0, G_{j_1, \dots, j_n}(z), t, G_{i_1, \dots, i_n}(y)) dz \\ &= \sum_{j_1, \dots, j_n} \int_{[0,1]^n} \hat{p}(s, x, t_0, G_{j_1, \dots, j_n}(z)) \sum_{k_1, \dots, k_n} \hat{p}(t_0, z, t, G_{k_1, \dots, k_n}(y)) dz \\ &= \sum_{j_1, \dots, j_n} \int_{[0,1]^n} \hat{p}(s, x, t_0, G_{j_1, \dots, j_n}(z)) p(t_0, z, t, y) dz \\ &= \int_{[0,1]^n} p(s, x, t_0) p(t_0, z, t, y) dz. \end{aligned}$$

If the temperature were constant, then x would have a unique equilibrium distribution (as will be clear from the proof of the theorem):

$$(2.1) \quad \pi_T(B) = \int_{B \cap [0,1]^n} \frac{1}{Z_T} \exp \{-U(x)/T\} dx$$

where

$$Z_T = \int_{[0,1]^n} \exp \{-U(x)/T\} dx.$$

As $T \rightarrow 0$, π_T concentrates on the global minima of U . In fact, for well-behaved functions U , $\{\pi_T\}_{T>0}$ has a unique weak limit, call it π_0 , and this satisfies

$$\pi_0(\{x: U(x) = \inf_y U(y)\}) = 1$$

(see Hwang [9]). If, for example, the global minimum of U is attained at a finite number of points, then $\pi_T \xrightarrow{w} \pi_0$, where π_0 concentrates on the global minima and has a simple characterization in terms of the Hessian of U . On the other hand, if the global minima of U form a set of positive Lebesgue measure, then again $\pi_T \xrightarrow{w} \pi_0$, but the latter is *uniform* on the set of global minima. In any case, we will assume the existence of a unique weak limit π_0 :

(B) There exists π_0 such that $\pi_T \xrightarrow{w} \pi_0$ as $T \rightarrow 0$.

Most commonly, U will possess only one global minimum, in which case (B) is trivial. Of course, π_0 necessarily concentrates on the global minima.

The following theorem gives conditions for the weak convergence of x_t to π_0 .

THEOREM. Assume (A) and (B) and that $T(t) = c/\log(2+t)$. For all c sufficiently large,

$$P(x_t \in \cdot | x_0 = x) \xrightarrow{w} \pi_0(\cdot) \quad \forall x \in [0, 1]^n.$$

Remarks. (1) Actually, the result holds if $T(t) \geq c/\log(2+t)$, c sufficiently large, and if (1) $T(t) \downarrow 0$; (2) $T(t)$ is continuously differentiable; and (3)

$$\frac{dT(t)/dt}{T(t)^3} e^{2\Delta/T(t)} \rightarrow 0$$

where $\Delta = \sup_{x,y \in [0,1]^n} (U(x) - U(y))$. The proof is the same.

(2) Almost sure convergence to the set minimizing U is, in general, impossible, as can be demonstrated already with $n = 1$ and a very simple function U . The reason can be put loosely as follows. If $T(t) \downarrow 0$ sufficiently slowly to guarantee escape from local minima, then repeated escapes from global minima are also guaranteed (albeit with increasing rareness).

(3) For most problems, the constant c necessary to guarantee convergence to global minima will most likely be too large to be practical. But if the discrete case is any guide, then our image processing experiments suggest that significant improvement is obtained over greedy algorithms (such as zero-temperature gradient descent) with a constant far too small to invoke the theorem. (See Geman and Geman [4] for further discussion. Hajek (personal communication) and Gidas [5] have actually identified the needed constant for the discrete case.)

3. Proof of the theorem. For any $x \in [0, 1]^n$, $t > s \geq 0$, $f \in C[0, 1]^n$, and μ a probability measure on $[0, 1]^n$, we give the following definitions:

(i) $\pi^s = \pi_{T(s)}$, π_T as in (2.1).

Notice that π^s has a density for all $0 \leq s < \infty$. We will use the same symbol, π^s , to denote this density:

(ii) $\pi^s(x) = \frac{\exp\{-U(x)/T(s)\}}{\int_{[0,1]^n} \exp\{-U(x)/T(s)\} dx}$,

(iii) $\mu(f) = \int_{[0,1]^n} f(x) \mu(dx)$,

(iv) $p(s, x, t, f) = \int_{[0,1]^n} f(y) p(s, x, t, y) dy$,

(v) $p(s, \mu, t, f) = \int_{[0,1]^n} \int_{[0,1]^n} f(y) p(s, x, t, y) \mu(dx) dy$.

The proof of the theorem is based upon the following two lemmas.

LEMMA 1. $\forall f \in C[0, 1]^n$, $s \geq 0$,

$$\lim_{t \rightarrow \infty} \sup_{w \in [0,1]^n} |p(s, v, t, f) - p(s, w, t, f)| = 0.$$

LEMMA 2. $\forall f \in C[0, 1]^n$,

$$\lim_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |p(s, \pi^s, t, f) - \pi^t(f)| = 0.$$

Assuming the validity of these, we establish the theorem as follows: fixing $x \in [0, 1]^n$ and $f \in C[0, 1]^n$,

$$\begin{aligned} & \overline{\lim}_{t \rightarrow \infty} |p(0, x, t, f) - \pi_0(f)| \\ & \cong \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |p(s, p(0, x, s, \cdot), t, f) - p(s, \pi^s, t, f)| \\ & \quad + \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |p(s, \pi^s, t, f) - \pi^t(f)| + \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |\pi^t(f) - \pi_0(f)| \\ & \hspace{15em} \text{(by Lemma 2 and } \pi^t = \pi_{T(t)} \xrightarrow{w} \pi_0) \\ & = \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |p(s, p(0, x, s, \cdot), t, f) - p(s, \pi^s, t, f)| \\ & = \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \left| \int_w \int_z [p(0, x, s, z) - \pi^s(z)] p(s, z, t, w) f(w) dz dw \right| \\ & = \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \left| \int_z [p(0, x, s, z) - \pi^s(z)] p(s, z, t, f) dz \right| \\ & \cong \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \sup_{v, w} |p(s, v, t, f) - p(s, w, t, f)| = 0 \end{aligned}$$

by Lemma 1.

Proof of Lemma 1. $\forall t \geq 0$ let

$$\delta_t = \inf_{x, y} p(t, x, t+1, y).$$

Then

$$\begin{aligned} & \overline{\lim}_{t \rightarrow \infty} \sup_{v, w} |p(s, v, t, f) - p(s, w, t, f)| \\ & = \overline{\lim}_{t \rightarrow \infty} \sup_{v, w} \left| \int p(s, v, s+1, z) p(s+1, z, t, f) dz \right. \\ & \quad \left. - \int p(s, w, s+1, z) p(s+1, z, t, f) dz \right| \\ & = \overline{\lim}_{t \rightarrow \infty} \sup_{v, w} \left| \int (p(s, v, s+1, z) - \delta_s) p(s+1, z, t, f) dz \right. \\ & \quad \left. - \int (p(s, w, s+1, z) - \delta_s) p(s+1, z, t, f) dz \right| \\ & \cong \overline{\lim}_{t \rightarrow \infty} \sup_{v, w} |(1 - \delta_s) \sup_z p(s+1, z, t, f) - (1 - \delta_s) \inf_z p(s+1, z, t, f)| \\ & = \overline{\lim}_{t \rightarrow \infty} (1 - \delta_s) \sup_{v, w} |p(s+1, v, t, f) - p(s+1, w, t, f)| \\ & \dots \\ & \cong \overline{\lim}_{t \rightarrow \infty} \left\{ \prod_{k=0}^{[t-s]-1} (1 - \delta_{s+k}) \right\} \\ & \quad \cdot \sup_{v, w} |p(s + [t-s], v, t, f) - p(s + [t-s], w, t, f)| \\ & \cong 2 \|f\|_\infty \overline{\lim}_{t \rightarrow \infty} \prod_{k=0}^{[t-s]-1} (1 - \delta_{s+k}) = 2 \|f\|_\infty \prod_{k=0}^\infty (1 - \delta_{s+k}) \end{aligned}$$

(where $[x]$ is the greatest integer not exceeding x). Hence, for the proof of Lemma 1 it is sufficient to show that

$$\sum_{k=0}^{\infty} \delta_{s+k} = \infty \quad \forall s \geq 0.$$

Let $\hat{\delta}_t = \inf_{x,y \in [0,1]^n} \hat{p}(t, x, t+1, y)$. Notice that $\hat{\delta}_t \leq \delta_t$ for all $t \geq 0$. We will show that

$$\sum_{k=0}^{\infty} \hat{\delta}_{s+k} = \infty \quad \forall s \geq 0.$$

Define $\mathcal{H} = \{f: [t, t+1] \rightarrow R^n, f \text{ continuous}\}$, and let P_x and Q_x be the probability measures on \mathcal{H} induced by

$$dZ_u = -\nabla \hat{U}(Z_u) du + \sqrt{2T(u)} dw_u, \quad Z_t = x, \quad u \in [t, t+1],$$

and

$$dZ_u = \sqrt{2T(u)} dw_u, \quad Z_t = x, \quad u \in [t, t+1],$$

respectively. Then $P_x \ll Q_x$ and

$$(3.1) \quad \frac{dP_x}{dQ_x}(Z(\circ)) = \exp \left\{ \int_t^{t+1} \frac{1}{2T(u)} \langle -\nabla \hat{U}(Z(u)), dZ(u) \rangle - \frac{1}{2} \int_t^{t+1} \frac{1}{2T(u)} |\nabla \hat{U}(Z(u))|^2 du \right\}$$

(see Stroock and Varadhan [14]). We will bound the exponent on the right-hand side. Apply Ito's formula, for the zero drift equations (i.e. under Q_x):

$$\begin{aligned} \frac{1}{2T(u)} \langle -\nabla \hat{U}(Z(u)), dZ(u) \rangle &= \frac{1}{2} \sum_{i=1}^n \hat{U}_{x_i x_i}(Z(u)) du - \frac{1}{2T(u)} d\hat{U}(Z(u)) \\ &\Rightarrow \int_t^{t+1} \frac{1}{2T(u)} \langle -\nabla \hat{U}(Z(u)), dZ(u) \rangle \\ &= \frac{1}{2} \sum_{i=1}^n \int_t^{t+1} \hat{U}_{x_i x_i}(Z(u)) du - \int_t^{t+1} \frac{1}{2T(u)} d\hat{U}(Z(u)). \end{aligned}$$

Under the assumptions on U , $\sup_{z \in R^n} |\hat{U}_{x_i x_i}(z)| \leq C_1 < \infty$ for some C_1 , and consequently

$$\left| \frac{1}{2} \sum_{i=1}^n \int_t^{t+1} \hat{U}_{x_i x_i}(Z(u)) du \right| \leq \frac{nC_1}{2}.$$

Using again the assumptions on U , together with the monotonicity and smoothness of $T(t)$,

$$\begin{aligned} \left| \int_t^{t+1} \frac{1}{2T(u)} d\hat{U}(Z(u)) \right| &= \left| \frac{\hat{U}(Z(t+1))}{2T(t+1)} - \frac{\hat{U}(Z(t))}{2T(t)} \right. \\ &\quad \left. - \int_t^{t+1} \hat{U}(Z(u)) d\left(\frac{1}{2T(u)}\right) \right| \leq \frac{C_2}{T(t+1)}. \end{aligned}$$

Hence

$$\left| \int_t^{t+1} \frac{1}{2T(u)} \langle -\nabla \hat{U}(Z(u)), dZ(u) \rangle \right| \leq \frac{nC_1}{2} + \frac{C_2}{T(t+1)}.$$

As for the other term in the exponent of (3.1), we easily get a bound $C_3/T(t+1)$. Therefore, for some constant C_4 ,

$$\frac{dP_x}{dQ_x}(Z(\cdot)) \geq \exp \{-C_4/T(t+1)\}.$$

Consequently, for any $\varepsilon > 0$, $x, y \in R^n$,

$$P_x(|Z(t+1) - y| < \varepsilon) \geq e^{-C_4/T(t+1)} Q_x(|Z(t+1) - y| < \varepsilon).$$

Under Q_x , $\{Z_i(t+1)\}_{i=1}^n$ are independent normal with

$$Z_i(t+1) \sim N\left(x_i, \int_t^{t+1} 2T(u) du\right).$$

Taking $x, y \in [0, 1]^n$,

$$\begin{aligned} P_x(|Z(t+1) - y| < \varepsilon) &\geq e^{-C_4/T(t+1)} \int_{|z-y|<\varepsilon} \frac{1}{(2\pi \int_t^{t+1} 2T(u) du)^{n/2}} \\ &\cdot \exp\left(-|z-x|^2 / \left\{4 \int_t^{t+1} T(u) du\right\}\right) dz \\ &\geq C_5 e^{-C_4/T(t+1)} \int_{|z-y|<\varepsilon} \exp\left(-(\sqrt{n} + \varepsilon)^2 / \left\{4 \int_t^{t+1} T(u) du\right\}\right) dz. \end{aligned}$$

Finally, then,

$$\begin{aligned} \hat{\delta}_t &= \inf_{x,y \in [0,1]^n} \hat{p}(t, x, t+1, y) \\ &= C_6 \inf_{x,y \in [0,1]^n} \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^n} P_x(|z(t+1) - y| < \varepsilon) \\ &\geq e^{-C_7/T(t+1)} \end{aligned}$$

for a sufficiently large constant C_7 . It now follows that the condition

$$\sum_{k=0}^{\infty} \hat{\delta}_{s+k} = \infty \quad \forall s \geq 0$$

is satisfied for $T(t) \geq c/\log(2+t)$, provided c is sufficiently large.

Proof of Lemma 2. For $t > s \geq 0$ define

$$N(s, t) = \int \pi^t(x) \left(\frac{p(s, \pi^2, t, x)}{\pi^t(x)} - 1 \right)^2 dx.$$

We will show that

$$(3.2) \quad \lim_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} N(s, t) = 0.$$

From this, Lemma 2 is obtained as follows: For any $f \in [0, 1]^n$

$$\begin{aligned} & \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} |p(s, \pi^s, t, f) - \pi^t(f)| \\ &= \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \left| \int (p(s, \pi^s, t, x) - \pi^t(x))f(x) dx \right| \\ &\leq \|f\|_\infty \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \int |p(s, \pi^s, t, x) - \pi^t(x)| dx \\ &= \|f\|_\infty \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \int \pi^t(x) \left| \frac{p(s, \pi^s, t, x)}{\pi^t(x)} - 1 \right| dx \\ &\leq \|f\|_\infty \overline{\lim}_{s \rightarrow \infty} \overline{\lim}_{t \rightarrow \infty} \sqrt{\int \pi^t(x) \left(\frac{p(s, \pi^s, t, x)}{\pi^t(x)} - 1 \right)^2 dx} \\ &= 0. \end{aligned}$$

The proof of (3.2) rests upon the following lemma.

LEMMA 3. Let $\Delta = \sup_{x, y \in [0, 1]^n} (U(x) - U(y))$. For all $t > s \geq 0$

$$\frac{\partial}{\partial t} N(s, t) \leq \Delta \left(\frac{d}{dt} \left(\frac{1}{T(t)} \right) \right) (1 + N(s, t)) - 2T(t) e^{-2\Delta/T(t)} N(s, t).$$

Accept, for now, Lemma 3. We have with $T(t) = c/\log(2+t)$,

$$(3.3) \quad \frac{\partial}{\partial t} N(s, t) \leq \frac{\Delta}{c} \left(\frac{1}{2+t} \right) - \left\{ \frac{2c}{\log(2+t)} \left(\frac{1}{2+t} \right)^{2\Delta/c} - \frac{\Delta}{c} \left(\frac{1}{2+t} \right) \right\} N(s, t).$$

From this, and the observation that $\lim_{t \downarrow s} N(s, t) = 0$, (3.2) is easily established, provided that c is sufficiently large. We will forgo these details; they only involve integrating (3.3), with the inequality replaced by equality.

All that remains is the proof of Lemma 3.

Proof of Lemma 3. First, observe that

$$N(s, t) = \int \frac{p(s, \pi^s, t, x)^2}{\pi^t(x)} dx - 1.$$

Hence

$$\begin{aligned} N_t(s, t) &= \int \left(\frac{d}{dt} \frac{1}{\pi^t(x)} \right) p(s, \pi^s, t, x)^2 dx \\ &\quad + 2 \int p_t(s, \pi^s, t, x) p(s, \pi^s, t, x) \left(\frac{1}{\pi^t(x)} \right) dx \\ &= A(s, t) + B(s, t) \end{aligned}$$

where

$$A(s, t) = \int \left(\frac{d}{dt} \frac{1}{\pi^t(x)} \right) p(s, \pi^s, t, x)^2 dx$$

and

$$B(s, t) = 2 \int p_t(s, \pi^s, t, x) p(s, \pi^s, t, x) \left(\frac{1}{\pi^t(x)} \right) dx.$$

Let $g(t) = 1/T(t)$. Then

$$\begin{aligned}
 A(s, t) &= - \int \left(\frac{1}{\pi'(x)} \right)^2 \left(\frac{d}{dt} \pi'(x) \right) p(s, \pi^s, t, x)^2 dx \\
 &= - \int \left(\frac{1}{\pi'(x)} \right)^2 (-g_t(t) U(x) \pi'(x) + g_t(t) \pi'(x) \pi'(U)) p(s, \pi^s, t, x)^2 dx \\
 (3.4) \quad &= g_t(t) \int \frac{1}{\pi'(x)} (U(x) - \pi'(U)) p(s, \pi^s, t, x)^2 dx \\
 &\leq \Delta g_t(t) \int \frac{p(s, \pi^s, t, x)^2}{\pi'(x)} dx \\
 &= \Delta g_t(t) (1 + N(s, t)).
 \end{aligned}$$

The treatment of $B(s, t)$ is more involved. The first step will be to show that

$$B(s, t) = -2T(t) \int |\nabla [p(s, \pi^s, t, x) / \pi'(x)]|^2 \pi'(x) dx.$$

From this, we will then derive the bound

$$B(s, t) \leq -2T(t) e^{-2\Delta/T(t)} N(s, t),$$

which, together with (3.4), completes the proof.

We rewrite $B(s, t)$ with the help of the forward equation for the original (\wedge) process: for $t > s \geq 0$,

$$\hat{p}_t(s, y, t, x) = \sum_{k=1}^n \{ T(t) \hat{p}_{x_k x_k}(s, y, t, x) + \hat{U}_{x_k}(x) \hat{p}_{x_k}(s, y, t, x) + \hat{U}_{x_k x_k}(x) \hat{p}(s, y, t, x) \}.$$

Integration over y , with respect to π^s , gives

$$\begin{aligned}
 (3.5) \quad \hat{p}_t(s, \pi^s, t, x) &= \sum_{k=1}^n \{ T(t) \hat{p}_{x_k x_k}(s, \pi^s, t, x) \\
 &\quad + \hat{U}_{x_k}(x) \hat{p}_{x_k}(s, \pi^s, t, x) + \hat{U}_{x_k x_k}(x) \hat{p}(s, \pi^s, t, x) \}.
 \end{aligned}$$

We wish to convert (3.5) into a similar equation for p . This conversion is based upon the following identities, which are justified by the assumed smoothness of U , and the resulting smoothness of \hat{p} (see, for example, [3]). For each integer i define

$$P(i) = \begin{cases} 1 & \text{if } i \text{ is even,} \\ -1 & \text{if } i \text{ is odd.} \end{cases}$$

Recalling that

$$x, y \in [0, 1]^n \Rightarrow p(s, y, t, x) = \sum_{i_1, \dots, i_n} \hat{p}(s, y, t, G_{i_1, \dots, i_n}(x)),$$

we have, for each $1 \leq k \leq n$ and each $x, y \in [0, 1]^n$:

$$\begin{aligned}
 \hat{U}_{x_k}(G_{i_1, \dots, i_n}(x)) &= P(i_k) U_{x_k}(x), \\
 \hat{U}_{x_k x_k}(G_{i_1, \dots, i_n}(x)) &= U_{x_k x_k}(x), \\
 p(s, \pi^s, t, x) &= \sum_{i_1, \dots, i_n} \hat{p}(s, \pi^s, t, G_{i_1, \dots, i_n}(x)),
 \end{aligned}$$

$$\begin{aligned}
 p_t(s, \pi^s, t, x) &= \sum_{i_1, \dots, i_n} \hat{p}_t(s, \pi^s, t, G_{i_1, \dots, i_n}(x)), \\
 p_{x_k}(s, \pi^s, t, x) &= \sum_{i_1, \dots, i_n} p(i_k) \hat{p}_{x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x)), \\
 p_{x_k x_k}(s, \pi^s, t, x) &= \sum_{i_1, \dots, i_n} \hat{p}_{x_k x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x)).
 \end{aligned}$$

Now take $x, y \in [0, 1]^n$ and replace x by $G_{i_1, \dots, i_n}(s)$ in (3.5):

$$\begin{aligned}
 \hat{p}_t(s, \pi^s, t, G_{i_1, \dots, i_n}(x)) &= \sum_{k=1}^n \{T(t) \hat{p}_{x_k x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x)) \\
 &\quad + U_{x_k}(x) P(i_k) \hat{p}_{x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x)) \\
 &\quad + U_{x_k x_k}(x) \hat{p}(s, \pi^s, t, G_{i_1, \dots, i_n}(x))\}.
 \end{aligned}$$

Summation over i_1, \dots, i_n yields:

$$\begin{aligned}
 p_t(s, \pi^s, t, x) &= \sum_{k=1}^n \{T(t) p_{x_k x_k}(s, \pi^s, t, x) + U_{x_k}(x) p_{x_k}(s, \pi^s, t, x) + U_{x_k x_k}(x) p(s, \pi^s, t, x)\} \\
 (3.6) \quad &= T(t) \sum_{k=1}^n \frac{\partial}{\partial x_k} \left\{ \pi^t(x) \frac{\partial}{\partial x_k} [p(s, \pi^s, t, x) / \pi^t(x)] \right\}.
 \end{aligned}$$

The associated boundary conditions are

$$p_{x_k}(s, \pi^s, t, x) = 0 \quad \text{whenever } x_k = 0 \text{ or } 1.$$

To see how these arise, take, for example, $x_k = 0$: letting $x = (x_1, x_2, \dots, x_{k-1}, 0, x_{k+1}, \dots, x_n)$, and letting e_k be the unit vector along the k th coordinate,

$$\begin{aligned}
 p_{x_k}(s, \pi^s, t, x) &= \lim_{\varepsilon \downarrow 0} \sum_{i_1, \dots, i_n} P(i_k) \hat{p}_{x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x_1, \dots, x_{k-1}, \varepsilon, x_{k+1}, \dots, x_n)) \\
 &= \lim_{\varepsilon \downarrow 0} \sum_{i_j: j \neq k} \sum_{p=-\infty}^{\infty} \\
 &\quad \cdot \{ \hat{p}_{x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x_1, \dots, x_{k-1}, 0, x_{k+1}, \dots, x_n)) + (2p + \varepsilon) e_k \\
 &\quad - \hat{p}_{x_k}(s, \pi^s, t, G_{i_1, \dots, i_n}(x_1, \dots, x_{k-1}, 0, x_{k+1}, \dots, x_n)) + (2p - \varepsilon) e_k \} \\
 &= 0.
 \end{aligned}$$

Now combine the boundary conditions on p with our boundary assumptions on ∇U (set out in (A)):

$$\frac{\partial}{\partial x_k} [p(t, \pi^s, t, x) / \pi^t(x)] = 0$$

for all x such that $x_k = 0$ or 1 . Multiplying the equation in (3.6) by $p(s, \pi^s, t, x) / \pi^t(x)$

and integrating x over $[0, 1]^n$ gives:

$$\begin{aligned} \frac{1}{2}B(s, t) &= \int p_t(s, \pi^s, t, x)p(s, \pi^s, t, x)\left(\frac{1}{\pi^t(x)}\right) dx \\ &= T(t) \sum_{k=1}^n \int_{x_j; j \neq k} \left\{ \int_{x_k} (p(s, \pi^s, t, x)/\pi^t(x)) \right. \\ &\quad \cdot \left. \frac{\partial}{\partial x_k} \left\{ \pi^t(x) \frac{\partial}{\partial x_k} [p(s, \pi^s, t, x)/\pi^t(x)] \right\} dx_k \right\} \\ &\quad \cdot dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_n \\ &= (\text{integrating over } x_k \text{ by parts}) \\ &\quad - T(t) \sum_{k=1}^n \int \left\{ \frac{\partial}{\partial x_k} [p(s, \pi^s, t, x)/\pi^t(x)] \right\}^2 \pi^t(x) dx \\ &= -T(t) \int |\nabla [p(s, \pi^s, t, x)/\pi^t(x)]|^2 \pi^t(x) dx. \end{aligned}$$

It remains to show that

$$(3.7) \quad \int |\nabla [p(s, \pi^s, t, x)/\pi^t(x)]|^2 \pi^t(x) dx \geq e^{-2\Delta/T(t)} N(s, t).$$

This final step is a consequence of the following proposition.

PROPOSITION. *If $\theta: [0, 1]^n \rightarrow R$ is continuously differentiable, and if*

$$\int \theta(x) dx = 0,$$

then

$$\int \theta(x)^2 dx \leq \int |\nabla \theta(x)|^2 dx.$$

For a more general version of this, see Gilbarg and Trudinger [7, p. 157].

Finally, fix s and t , and let

$$\phi(x) = \frac{p(s, \pi^s, t, x)}{\pi^t(x)} - 1$$

and $\alpha = \int \phi(x) dx$. Since $\int \phi(x) \pi^t(x) dx = 0$,

$$N(s, t) = \int \phi(x)^2 \pi^t(x) dx \leq \int (\phi(x) - \alpha)^2 \pi^t(x) dx.$$

Notice that $e^{-\Delta/T(t)} \leq \pi^t(x) \leq e^{\Delta/T(t)}$, and therefore

$$\begin{aligned} N(s, t) &\leq e^{\Delta/T(t)} \int (\phi(x) - \alpha)^2 dx \\ &\leq e^{\Delta/T(t)} \int |\nabla \phi(x)|^2 dx \quad (\text{by the proposition}) \\ &\leq e^{2\Delta/T(t)} \int |\nabla \phi(x)|^2 \pi^t(x) dx, \end{aligned}$$

which is the same as (3.7).

REFERENCES

- [1] F. ALUFFI-PENTINI, V. PARISI AND F. ZIRILLI, *Global optimization and stochastic differential equations*, J. Optim. Theory Applications, 1985, to appear.
- [2] C. ČERNÝ, *A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm*, preprint, Institute of Physics and Biophysics, Comenius Univ., Bratislava, 1982.
- [3] E. B. DYNKIN, *Markov Processes—II*, Springer-Verlag, New York, 1965.
- [4] S. GEMAN AND D. GEMAN, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE-PAMI, 6 (1984), pp. 721–741.
- [5] B. GIDAS, *Non-stationary Markov chains and convergence of the annealing algorithm*, J. Statistical Physics, 39 (1985), pp. 73–131.
- [6] ———, *The Langevin equation as a global minimization algorithm*, to appear in *Disordered Systems and Biological Organizations*, E. Bienenstock, F. Flogeman and G. Weisbuch, eds., Springer-Verlag, Berlin.
- [7] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second-Order*, Springer-Verlag, Berlin, 1977.
- [8] U. GREENDER, *Tutorial in Pattern Theory*, Div. Applied Mathematics, Brown Univ., Providence, RI, 1984.
- [9] C.-R. HWANG, *Laplace's method revisited: weak convergence of probability measures*, Ann. Probab., 8 (1980), pp. 1177–1182.
- [10] S. KIRKPATRICK, C. D. GELATT, JR. AND M. P. VECCHI, *Optimization by simulated annealing*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1982.
- [11] H. J. KUSHNER, *Asymptotic global behavior for stochastic approximations and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo*, Lefschetz Center for Dynamical Systems report 85-7, Div. Applied Mathematics, Brown Univ., Providence, RI, 1985.
- [12] J. L. MARROQUIN, *Surface reconstruction preserving discontinuities*, preprint, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, 1984.
- [13] M. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER AND E. TELLER, *Equations of state calculations by fast computing machines*, J. Chem. Phys., 21 (1953), pp. 1087–1091.
- [14] D. W. STROOCK AND S. R. S. VARADHAN, *Multidimensional Diffusion Processes*, Springer-Verlag, New York, 1979.