Diffusion for Global Optimization in n R

Article *in* SIAM Journal on Control and Optimization · January 1987

Shuenn-Jyi Sheu
National Central University

SIAM J. CONTROL AND OPTIMIZATION
Vol. 25, No. 3, May 1987

015

DIFFUSION FOR GLOBAL OPTIMIZATION IN $\mathbb{R}^n$*

TZUU-SHUH CHIANG†, CHII-RUEY HWANG† AND SHUENN-JYI SHEU†

Abstract. We seek a global minimum of $U:\mathbb{R}^n \to \mathbb{R}$. The solution to $(*)(d/dt)X(t) = -\nabla U(X(t))$ will find local minima. Using the idea of simulated annealing, we consider the diffusion process, $dX(t) = -\nabla U(X(t))\,dt + \sigma(t)\,dW(t)$, $X(0) = x$, where $W(\cdot)$ is the n-dimensional standard Brownian motion and $\frac{1}{2}\sigma^2(t)$ is the annealing rate which decreases to zero as t goes to ∞. Under suitable condition on $U(x)$, we prove that $X(t)$ converges weakly to a probability measure π if for large t, $\sigma^2(t) = c/\log t$ with $c > c_0$, where c_0 has a simple expression involving the action function of the dynamical system $(*)$, π concentrates on the global minima of U and is the weak limit of the Gibbs densities $\pi_t(x) \propto \exp(-2U(x)/\sigma^2(t))$.

The above result can also be formulated as follows: consider the Fokker-Planck equation (forward equation)

$$\frac{\partial}{\partial t} V(t,y) = \frac{1}{2}\sigma^2(t)\Delta V(t,y) + \nabla \cdot (V(t,y)\nabla U(y))$$

with $V(0,y) = \delta_x(y)$.

If $\sigma^2(t) = c/\log t$ for large t and $c > c_0$, then $V(t,y) \to \pi$ weakly.

Key words. diffusion, global optimization, simulated annealing, perturbed dynamical system, large deviation, action functional

AMS(MOS) subject classifications. GOH10, GOJ70

1. Introduction. For a fixed $U:\mathbb{R}^n \to [0,\infty)$, we give suitable conditions on U such that by choosing

$$\sigma^2(t) = \frac{c}{\log t} \quad \text{for large } t \text{ with } c > c_0 \quad \text{as } t \to \infty$$

$p(s,x,t,\cdot)$ converges weakly to a probability measure π concentrating on the global minima of U, $p(s,x,t,\cdot)$ is the transition probability of the diffusion process defined by

$$dZ(t) = -\nabla U(Z(t))\,dt + \sigma(t)\,dW(t), \tag{1.1}$$

where $\frac{1}{2}\sigma^2(t)$ corresponding to the "temperature" is the annealing rate, $W(t)$ is a standard Brownian motion in $\mathbb{R}^n$. The probability π is the weak limit of the Gibbs density

$$\pi_t(x) \propto \exp\left(-\frac{2U(x)}{\sigma^2(t)}\right) \quad \text{as } t \to \infty. \tag{1.2}$$

The constant c_0, which will be defined in § 2, has a simple expression involving the action function of the dynamical system

$$\frac{dY(t)}{dt} = -\nabla U(Y(t)). \tag{1.3}$$

The idea of our approach is as follows: Heuristically if we hold the temperature at time s for a *fairly large* amount of time, then $Z(t)$ defined by (1.1) and the fixed temperature process behaves almost the same at the end of that time interval. Hence, instead of (1.1) we may consider

$$\begin{aligned} &dX(t) = -\nabla U(X(t))\,dt + \sigma(s)\,dW(t), \\ &X(0) = x. \end{aligned} \tag{1.4}$$

* Received by the editors October 23, 1985; accepted for publication (in revised form) April 16, 1986.
† Institute of Mathematics, Academia Sinica, Taipei, Taiwan.

Note that the weak limit π depends only on the local property of U near the minima [9]. If we modify U for large $|x|$, π remains unchanged. One may consider a modified version with $U(X)=|x|^4$ for large $|x|$. In this case $X(t)$ comes back from "infinity" to a fixed finite ball in a finite time which is independent of $\sigma(s)$. It is almost as in the compact situation. Some of the ideas used in [4], which dealt with a reflected version of (1.1), can be used again in here. Furthermore, results and ideas in [13], [14] are available when we consider (1.4).

Independently, Gidas and Kushner also consider (1.1) in their recent works [6], [11], respectively.

Our work was inspired by the "simulated annealing" [1], [10] which deals mainly with the discrete state space. A lot of research has been going on in this aspect, see e.g. [3], [5], [8].

The use of (1.1) as a global minimization algorithm is motivated by problems in imaging processing [4], [7] as well as in studying lattice gauge theory [12].

We think that the constant c_0 obtained here is not the best possible. One may argue heuristically as follows. For the fixed temperature process (1.4) with $\varepsilon=\sigma(s)$, Lemma 3 in § 3 describes a distance between p_t and π^ε. Let $L_\varepsilon=\frac{1}{2}\varepsilon^2\Delta-\nabla U\cdot\nabla$ and $\lambda_2(\varepsilon)$ denote the second eigenvalue of L_ε. Let $\| \ \|_{\pi^\varepsilon}$ denote the norm of $L^2(\pi^\varepsilon)$; then clearly

$$\|p_t^\varepsilon(x,f)-\pi^\varepsilon(f)\|_{\pi^\varepsilon}\leqq \exp(t\lambda_2(\varepsilon))\|f\|_{\pi^\varepsilon}.$$

If $\lim_{\varepsilon\to 0}\varepsilon^2\log(-\lambda_2(\varepsilon))=-c_1$, then for $c>c_1$ such that $c>c_1+a$ we have $-\lambda_2(\varepsilon)\geqq \exp(-(c_1+a)/\varepsilon^2)$ for small ε. For $\varepsilon^2\approx c/\log t$

$$\|p_t^\varepsilon(x,f)-\pi^\varepsilon(f)\|_{\pi^\varepsilon}\leqq \exp(-t^{1-((c_1+a)/c)})\to 0 \quad \text{as } t\to\infty.$$

One would expect c_1 here is the critical constant.

Another heuristic approach is to consider the function

$$N(t)=\int\left|\frac{p(0,x,t,y)}{\pi_t(y)}-1\right|^2\pi_t(y)\,dy, \qquad t>1,$$

which was discussed previously in [4]. If $N(t)\to 0$ as $t\to\infty$, then it is easy to see that $p(0,x,t,\cdot)\to\pi(\cdot)$ weakly. For simplicity, let us write $\sigma^2(t)=2T(t)$ and heuristically one has

$$\begin{aligned}\frac{dN(t)}{dt}&=\left(\frac{d}{dt}\frac{1}{T(t)}\right)\int\frac{1}{\pi_t(y)}(U(y)-\pi_t(U))p(0,x,t,y)^2\,dy\\ &\quad -2T(t)\int\left|\nabla_y\left(\frac{p(0,x,t,y)}{\pi_i(y)}\right)\right|^2\pi_t(y)\,dy\\ &\leqq\frac{c_2}{t}(N(t)+1)-2(-\lambda_2(\sigma(t)))N(t)\\ &=\frac{c_2}{t}+N(t)\left(\frac{c_2}{t}-2t^{-((c_1+a)/c)}\right)\end{aligned}$$

by

$$\int|f(y)-\pi_t(f)|^2\pi_t(y)\,dy\leqq\frac{T(t)}{-\lambda_2(\sigma(t))}\int|\nabla f(y)|^2\pi_t(y)\,dy.$$

Then one can establish $N(t)\to 0$ from this differential inequality.

2. Statement of result. Let U be a twice continuously differentiable function from $\mathbb{R}^n$ to $[0,\infty)$ such that the following assumptions hold:

$$\min_{x\in\mathbb{R}^n} U(x)=0,$$

(A1) $$U(x)\to\infty \quad \text{and } |\nabla U(x)|\to\infty \quad \text{as } |x|\to\infty,$$

$$\varliminf_{|x|\to\infty} |\nabla U(x)|^2-\Delta U(x)>-\infty.$$

For $0<\varepsilon<1$,

(A2) $$\pi^{\varepsilon}(x):=\frac{1}{c(\varepsilon)}\exp\left(-\frac{2U(x)}{\varepsilon^2}\right),$$

where $c(\varepsilon)=\int_{\mathbb{R}^n}\exp\left(-\frac{2U(x)}{\varepsilon^2}\right)dx<\infty.$

(A3) π^{ε} has a unique weak limit π as $\varepsilon\downarrow 0$.

Clearly π concentrates on the global minima of U. The detailed discussion for the existence of π and its characterization in terms of the Hessian of U can be found in [9].

For simplicity we shall assume $\sigma^2(t)<1$, $\sigma^2(t)=c/\log t$ for large t and the process $Z(t)$ starts at $Z(0)=x$.

Let S denote the set of all stationary points of U, i.e., $S=\{x|\nabla U(x)=0\}$.

For any $\eta>0$, $\xi>0$, we define the following:

$S(\eta):=\{x|d(x,S)<\eta\}$,

$K(\eta):=$ the set containing all the solutions of the dynamical system (1.3) with starting points in $S(\eta)$,

$K(\eta,\xi):=\{x|d(x,K(\eta))\leqq\xi\}$,

$$I(t,x,y):=\inf_{\substack{\psi(0)=x\\ \psi(t)=y}}\frac{1}{2}\int_0^t|\dot\psi(s)+\nabla U(\psi(s))|^2\,ds,$$

$$J(t,\eta,\xi):=\sup_{x,y\in K(\eta,\xi)}(I(t,x,y)-2U(y)),$$

$$J(\eta,\xi):=\varlimsup_{t\to\infty} J(t,\eta,\xi),$$

$$c_0:=\frac{3}{2}\inf_{\eta}(\inf_{\xi} J(\eta,\xi)).$$

For a measure μ, $\mu(f):=\int f\,d\mu$.

THEOREM. *Assume* (A1), (A2) *and* (A3) *and* $c>c_0$; *then for any bounded continuous function* f

$$p(0,x,t,f)\to\pi(f) \quad \textit{as } t\to\infty$$

and the convergence is uniform for x *in a compact set.* $p(s,x,t,\cdot)$ *here is the transition probability of* (1.1).

Remark 1. Without going into detail, we note that $J(\eta, \xi)$ is independent of η, ξ and

$$c_* = J(\eta, \xi) = \sup_{x,y\in K(\eta,\xi)} (V(x,y) - 2U(y))$$
$$= \sup_{x,y\in S} (V(x,y) - 2U(y)),$$

where $V(x,y) = \lim_{t\to\infty} I(t,x,y)$. This $V(x,y)$ is the same function used by Freidlin and Wentzell for describing the long time behavior of perturbed dynamical systems $dX(t) = -\nabla U(X(t))\,dt + \varepsilon\, dw(t)$.

Remark 2. We suspect that $c > c_* = \frac{2}{3} c_0$ is enough for the result of the theorem to hold.

3. Proof of theorem. The proof of the main theorem is based on the following three lemmas.

LEMMA 1. $\lim_{t\to\infty} p(s,x,t,K(\eta,\xi)) = 1$. *The convergence is uniform for x in a compact set.*

LEMMA 2. *Consider a family of processes defined by*

$$\begin{aligned} dY(s,t) &= -\nabla U(Y(s,t))\,dt + \sigma(s)\,dW(t), \\ Y(s,0) &= y. \end{aligned} \tag{3.1}$$

Then for $h(s) \leqq s^{2/3}$ and $h(s)$ increasing to ∞,

$$\lim_{s\to\infty} E_{0,y}(f(Y(s,h(s)))) - E_{s,y}(f(Z(\beta(s)))) = 0,$$

where $\beta(\cdot)$ is defined by

$$\int_s^{\beta(s)} \frac{\log s}{\log u}\,du = h(s).$$

And the convergence is uniform for y in a compact set.

LEMMA 3. *Consider the following process*

$$\begin{aligned} dX(t) &= -\nabla U(X(t)) + \varepsilon\, dW(t), \\ X(0) &= x. \end{aligned} \tag{3.2}$$

Then there exist $T_0 > 0 \ni \forall M > 0, \forall T > 2T_0, \forall \alpha > 0$

$$\overline{\lim_{\varepsilon\to 0}} |E_x^\varepsilon f(X(mT)) - \pi^\varepsilon(f)| \leqq 4e^{-M}\|f\|,$$

where

$$m = M \exp\left(\frac{1}{\varepsilon^2}(J(t,\eta,\xi) + \alpha)\right), \qquad t = T - 2T_0,$$

α is an arbitrary fixed positive constant. The convergence is uniform for x in a compact set.

Assuming the validity of these, we establish the theorem as follows: For a fixed $c > c_0$, there exists an $\alpha > 0$ such that for sufficiently large time t, sufficiently small η and ξ,

$$c > \tfrac{3}{2}(J(t,\eta,\xi) + \alpha). \tag{3.3}$$

Choose a fixed large T such that (3.3) holds for time $T - 2T_0$, where T_0 is the constant in Lemma 3.

Choose $h(s)$ in Lemma 2 as

$$h(s) = MT \exp\left(\frac{1}{\sigma^2(s)}(J(T-2T_0, \eta, \xi)+\alpha)\right)$$
$$(3.4) \qquad = MTs^{(J(T-2T_0,\eta,\xi)+\alpha)/c}$$
$$< s^{2/3} \quad \text{for large } s.$$

Note that h and β are strictly increasing functions and $s+h(s) \leqq \beta(s) \leqq s+2h(s)$. Hence for $t \gg 1$, one can choose s such that $t=\beta(s)$. Clearly $s<t$ and $s \to \infty$.

$$p(0, x, t, f) - \pi_s(f) = \int p(0, x, s, y)p(s, y, t, f)\, dy - \pi_s(f)$$
$$= \int_{y \in K(\eta, \xi)} p(0, x, s, y)(p(s, y, t, f) - \pi_s(f))\, dy$$
$$+ \int_{y \notin K(\eta, \xi)} p(0, x, s, y)(p(s, y, t, f) - \pi_s(f))\, dy.$$

The second term is bounded by

$$2\|f\|(1-p(0, x, s, K(\eta, \xi))),$$

which goes to zero uniformly over x in a compact set as $s \to \infty$ by Lemma 1. Note that $\pi_s(f) \to \pi(f)$.

By Lemma 2,

$$E_{0,y}(f(Y(s, h(s)))) - p(s, y, \beta(s), f) \to 0,$$
$$E_{0,y}(f(Y(s, h(s)))) = E_y^{\sigma(s)} f(X(h(s)))$$
$$= E_y^{\sigma(s)} f(X(mT))$$

by identifying $h(s)$ with mT and $\sigma(s)$ with ε.

Now by Lemma 3, we have the theorem.

4. Proof of Lemma 1. Let us first assume the validity of the following two lemmas.

LEMMA 4.1. *For any compact set K in $\mathbb{R}^n$, the family of probability measures*

$$\{p(s, x, t, \cdot)|s<t, x \in K\}$$

is tight.

LEMMA 4.2. *For any compact set K, there exists T such that for any $t>T$, $Y(t) \in K(\eta)$, where*

$$\frac{dY(t)}{dt} = -\nabla U(Y(t)), \qquad Y(0) = y \in K.$$

The proof of Lemma 1 is as follows: By Lemma 4.1, for any $\delta>0$ and for any given compact set J, there exists a compact set K such that

$$p(s, x, t, K) > 1-\delta/2 \quad \text{for all } s<t, x \in J.$$

Choose T as in Lemma 4.2, then

$$p(s, x, t, K(\eta, \xi)) = \int p(s, x, t-T, dy)p(t-T, y, t, K(\eta, \xi))$$
$$> \int_K p(s, x, t-T, dy)p(t-T, y, t, K(\eta, \xi)).$$

It remains to show that there exists t_0 such that

$$p(t-T, y, t, K(\eta, \xi)) > 1-\delta/2, \quad y \in K, \quad t > t_0.$$

Let $Y(\cdot)$ be the solution of (4.1) with $Y(t-T)=y$. Then by Lemma 4.2,

$$\begin{aligned} p(t-T, y, t, K(\eta, \xi)) &= E_{t-T,y}\{Z(t) \in K(\eta, \xi)\} \\ &= E_{t-T,y}\{|Z(t)-Y(t)| \leqq \xi\} \\ &\quad + E_{t-T,y}\{|Z(t)-Y(t)| > \xi, Z(t) \in K(\eta, \xi)\} \\ &\geqq E_{t-T,y}\{|Z(t)-Y(t)| \leqq \xi\} \\ &\geqq 1 - E_{t-T,y}\{\tau \leqq t\}, \end{aligned}$$

where $\tau := \inf\{s > t-T, |Z(s)-Y(s)| > \xi\}$.

Now consider the process $Z(t)$ starting at $Z(t-T)=y$. Compare $Z(t)$ and $Y(t)$ up to τ. For $u \leqq \tau$,

$$Z(u)-Y(u) = \int_{t-T}^{u} (-\nabla U(Z(s)) + \nabla U(Y(s)))\, ds + H(u),$$

where $H(u) = \int_{t-T}^{u} \sigma(s)\, dW(s)$. Note that for $t-T \leqq s \leqq \tau$, $Z(s)$ and $Y(s)$ are in a compact set in which U is Lipschitz with constant d, and we have

$$|Z(u)-Y(u)| \leqq d \int_{t-T}^{u} |Z(s)-Y(s)|\, ds + |H(u)|.$$

By Gronwall inequality,

$$|Z(u)-Y(u)| \leqq \exp(d(u-(t-T))) \sup_{t-T \leqq s \leqq u} |H(s)|.$$

For $\tau \leqq t$,

$$\xi = |Z(\tau)-Y(\tau)| \leqq e^{dT} \sup_{t-T \leqq s \leqq t} |H(s)|,$$

$$\begin{aligned} p\{\tau \leqq t\} &\leqq p\left\{ \sup_{t-T \leqq s \leqq t} |H(s)| \geqq e^{-dT}\xi \right\} \\ &\leqq 2n \exp\left\{ \frac{-\xi^2 \log t}{2cnT} e^{-2dT} \right\} \\ &\leqq \frac{\delta}{2} \quad \text{if } t \geqq t_0 \text{ for a fixed large } t_0 \end{aligned}$$

[15, p. 87]. Hence,

$$p(t-T, y, t, K(\eta, \xi)) \geqq 1 - \frac{\delta}{2}.$$

This completes the proof.

Proof of Lemma 4.1.

$$\begin{aligned} de^{U(Z(t))} e^{\lambda t} &= \left(\frac{\sigma^2(t)}{2} \Delta U(Z(t)) - \left(1 - \frac{\sigma^2(t)}{2}\right) |\nabla U(Z(t))|^2 + \lambda \right) e^{\lambda t} e^{U(Z(t))}\, dt \\ &\quad + e^{\lambda t}\, dM(t), \end{aligned}$$

where $M(t) = \int_0^t \sigma(s) \nabla U(Z(s))\, e^{U(Z(s))}\, dW(s)$ is a local martingale.

For any $\lambda > 0$, there exists constant $A = A(\lambda) > 0$ such that

$$\left(\frac{\sigma^2(t)}{2}\Delta U(z) - \left(1 - \frac{\sigma^2(t)}{2}\right)|\nabla U(z)|^2 + \lambda\right) e^{U(z)}$$

$$= \left[\frac{\sigma^2(t)}{2}(\Delta U(z) - |\nabla U(z)|^2) - (1 - \sigma^2(t))|\nabla U(z)|^2 + \lambda\right] e^{U(z)}$$

$$\leqq A \quad \forall t \text{ and } z \in \mathbb{R}^n,$$

since for large $|z|$, the term in the bracket parentheses is negative for all t.

Let $\tau_m := \inf\{t; |Z(t)| > m\}$ and $\tau = \lim_{m\to\infty} \tau_m$ is the explosion time. Then

$$E_{s,x}\{e^{U(Z(t\wedge\tau_m))} e^{\lambda(t\wedge\tau_m)}\} \leqq A E_{s,x}\left\{\int_s^{t\wedge\tau_m} e^{\lambda u}\, du + e^{U(x)} e^{\lambda s}\right\}.$$

Let $m \to \infty$,

$$E_{s,x}\{e^{U(Z(t\wedge\tau))} e^{\lambda(t\wedge\tau)}\} \leqq \frac{A}{\lambda}(e^{\lambda t} - e^{\lambda s}) + e^{U(x)} e^{\lambda s}.$$

If $p\{\tau \leqq \infty\} > 0$, then there exists t such that $E_{s,x}\, e^{U(Z(t\wedge\tau))} e^{\lambda(t\wedge\tau)} = \infty$. Hence we conclude that $p\{\tau = \infty\} = 1$.

Now we have

$$E_{s,x}\, e^{U(Z(t))} \leqq \frac{A}{\lambda} + e^{U(x)} e^{-\lambda(t-s)}$$

$$\leqq \frac{A}{\lambda} + e^{U(x)}.$$

From this, it is easy to show that $\{p(s, x, t, \cdot), s < t, x \in K\}$ is tight.

Proof of Lemma 4.2.

$$U(Y(t)) - U(y) = -\int_0^t |\nabla U(Y(s))|^2\, ds. \tag{4.1}$$

For $z \notin S(\eta)$, there exists $\upsilon > 0$ independent of z such that $U(z) > \upsilon$ and $|\nabla U(z)| > \upsilon$. Hence by (4.1) and the compactness of K, there exists T such that $Y(t) \in S(\eta)$ for some $t \leqq T$. But by the definition of $K(\eta)$, once $Y(t) \in S(\eta) \subseteq K(\eta)$, then $Y(t') \in K(\eta)$ if $t' > t$. Therefore, $Y(t) \in K(\eta)$ if $t \geqq T$.

5. Proof of Lemma 2. For simplicity, we shall write $b = -\nabla U$. Define $\beta(s, t)$ by

$$\int_s^{\beta(s,t)} \frac{\sigma^2(u)}{\sigma^2(s)}\, du = t.$$

Note that $\beta(s)$ defined in the statement is $\beta(s, h(s))$. For any fixed s, define $\tilde{Z}(s, t) = Z(\beta(s, t))$; then

$$\tilde{Z}(s, t) = x + \int_0^t b(\tilde{Z}(s, u)) \frac{\log \beta(s, u)}{\log s}\, du + \sigma(s) W(t).^1$$

[1] The Wiener process $W(t)$ may not be the same at each occurrence. This does not matter because we are only interested in the probability distributions.

Now compare $\tilde{Z}(s,\cdot)$ with $Y(s,\cdot)$,

$$Y(s,t)=x+\int_0^t b(Y(s,u))\,du+\sigma(s)W(t).$$

Let us first consider $|b(x)|\leqq M<\infty$. By the Girsanov theorem,

$$Ef(\tilde{Z}(s,t))=E(f(Y(s,t))\exp(A(t)-\tfrac{1}{2}B(t))),$$

where

$$A(t)=\int_0^t b(Y(s,u))\left(\frac{\log\beta(s,u)}{\log s}-1\right)\frac{1}{\sigma(s)}\,dW(u),$$

$$B(t)=\int_0^t |b(Y(s,u))|^2\left(\frac{\log\beta(s,u)}{\log s}-1\right)^2\frac{1}{\sigma^2(s)}\,du,$$

$$Ef(\tilde{Z}(s,t))=Ef(Y(s,t))+E\left\{f(Y(s,t))\left(\exp\left(A(t)-\frac{1}{2}B(t)\right)-1\right)\right\}.$$

We shall show that the second term tends to zero for $t=h(s)\leqq s^{2/3}$ as $s\to\infty$.

$$\begin{aligned}E(\exp(A(t)-\tfrac{1}{2}B(t))-1)^2&=E(\exp(2A(t)-B(t)))-1\\ &=E(\exp(2A(t)-2B(t))(\exp B(t)-1)),\end{aligned}\tag{5.1}$$

since $\exp(A(t)-\frac{1}{2}B(t))$ and $\exp(2A(t)-2B(t))$ are martingales with expectation 1.

$$\begin{aligned}B(t)&=\int_0^t |b(Y(s,u))|^2\left(\frac{\log\beta(s,u)}{\log s}-1\right)^2\frac{1}{\sigma^2(s)}\,du\\
&\leqq\frac{M^2}{c}\log s\int_0^t\left(\frac{\log\beta(s,u)}{\log s}-1\right)^2du\\
&=\frac{M^2}{c}\log s\int_s^{\beta(s,t)}\left(\frac{\log u}{\log s}-1\right)^2\frac{\log s}{\log u}\,du\\
&\leqq\text{constant}\,\frac{1}{\log s}\int_s^{\beta(s,t)}\left(\frac{u}{s}-1\right)^2du\\
&=\text{constant}\,\frac{1}{\log s}\,\frac{(\beta(s,t)-s)^3}{s^2}\\
&\leqq\text{constant}\,\frac{1}{\log s}\to 0,\end{aligned}$$

since $s+2t\geqq\beta(s,t)\geqq s+t$ and we choose $t=h(s)\leqq s^{2/3}$. Then (5.1) is bounded by

$$\text{constant}\,\frac{1}{\log s}\,E(\exp(2A(t)-2B(t)))=\text{constant}\,\frac{1}{\log s}\to 0.$$

Therefore for bounded $b(x)$, we have proved

$$E_{s,x}f(Z(\beta(s)))-E_{0,x}f(Y(s,h(s)))\to 0.\tag{5.2}$$

Now let us prove the lemma for the general case. Let

$$\tau_r=\inf\{t\colon U(Z(t))>r\},$$

$$\tau_r(s)=\inf\{t\colon U(Y(s,t))>r\}.$$

Using the same argument as before by taking f an indicator function and noticing that b is bounded on the compact set $\{U(x) \leqq r\}$, we can show that as $s \to \infty$.

$$E_{s,x}\{\tau_r > \beta(s)\} - E_{0,x}\{\tau_r(s) > h(s)\} \to 0. \tag{5.3}$$

If there exists r such that

(5.4) $\quad E_{0,x}\{\tau_r(s) > h(s)\} \to 1$ uniformly over x in a compact set,

then by combining (5.2) for bounded b and (5.3), one gets Lemma 2.

As for (5.4), it is an easy consequence of Lemma 6.4.

6. Proof of Lemma 3.

Super normal case. Let us first prove Lemma 3 for the following particular super normal case: there is a large fixed R_0, such that

$$\begin{aligned} &U(x) = |x|^4 \quad \text{for } |x| > R_0; \quad \text{then} \\ &|\nabla U(x)| = 4|x|^3, \qquad \Delta U(x) = (4n+8)|x|^2, \end{aligned} \tag{6.1}$$

and $K(\eta, \xi) \subseteq \{|x| < R_0\}$.

$$\begin{aligned} &dX(t) = -\nabla U(X(t))\,dt + \varepsilon\, dW(t), \\ &X(0) = x. \end{aligned} \tag{6.2}$$

Let $\tau = \inf\{t \mid |X(t)| = 2R_0\}$.

CLAIM. *There exists a constant c_1 such that for any $|x| > 2R_0$, for any $0 < \varepsilon < 1$, $E_x^\varepsilon(\tau) \leqq c_1$.*

Proof. For $|x| > 2R_0$, $\tau_0 := \inf\{t \mid |X(t)| = \frac{1}{2}|x|\}$, then

$$E_x^\varepsilon U(X(\tau_0)) - U(x) = E_x^\varepsilon \int_0^{\tau_0} \left(-|\nabla U(X(s))|^2 + \frac{\varepsilon^2}{2} \Delta U(X(s)) \right) ds.$$

$$\begin{aligned} \left|\frac{1}{2}x\right|^4 - |x|^4 &= E_x^\varepsilon \int_0^{\tau_0} \left(-16|X(s)|^6 + \frac{\varepsilon^2}{2n+4}|X(s)|^2 \right) ds \\ &\leqq -c_3|x|^6 E_x^\varepsilon \tau_0. \end{aligned}$$

Therefore,

$$E_x^\varepsilon \tau_0 \leqq c_4|x|^{-2}.$$

Now let us define the following stopping times:

$$\begin{aligned} &\tau_1 = \inf\{t \mid |X(t)| = \tfrac{1}{2}|x|\}, \\ &\tau_2 = \inf\{t > \tau_1 \mid |X(t)| = \tfrac{1}{2}|X(\tau_1)|\}, \\ &\vdots \\ &\tau_{i+1} = \inf\{t > \tau_i \mid |X(t)| = \tfrac{1}{2}|X(\tau_i)|\}. \end{aligned}$$

Let m be a positive integer such that

$$2^m R_0 < |x| \leqq 2^{m+1} R_0.$$

Then, $\tau \leqq \tau_m$ and

$$\begin{aligned} E_x^\varepsilon(\tau) &\leqq \sum_{k=2}^{m} E_x^\varepsilon(\tau_k - \tau_{k-1}) + E_x^\varepsilon(\tau_1) \\ &= \sum_{k=2}^{m} E_x^\varepsilon E_{x(\tau_{k-1})}^\varepsilon(\tau_0) + E_x^\varepsilon(\tau_1) \\ &\leqq c_4 \sum_{k=2}^{m} E_x^\varepsilon |X(\tau_{k-1})|^{-2} + c_4 |x|^{-2} \\ &\leqq c_4 R_0^{-2} \sum_{k=1}^{m} (2^{m-k+1})^{-2} \leqq \frac{1}{3} c_4 R_0^{-2} = c_1. \end{aligned}$$

CLAIM. *For any $\delta > 0$ there exist T_0 and ε_0 such that*

$$E_x^\varepsilon\{X(T_0) \in K(\eta, \xi)\} \geqq 1 - \delta \quad \textit{for all } x \in \mathbb{R}^n \textit{ and } \varepsilon \leqq \varepsilon_0. \tag{6.3}$$

Proof. First choose T_2 such that $c_1/T_2 < \delta/2$.

$$B(2R_0) = \{|x| \leqq 2R_0\} \supset K(\eta, \xi).$$

T_1 is the time in Lemma 4.2 such that with initial point in $B(2R_0)$ the solution of the dynamic system will be contained in $K(\eta)$ after time T_1. Now let $T_0 = T_1 + T_2$.

As in the proof of Lemma 1, we can choose an ε_0 such that

$$E_x^\varepsilon\{X(t) \in K(\eta, \xi)\} > 1 - (\delta/2) \quad \forall x \in B(2R_0), \quad \forall T_1 \leqq t \leqq T_0, \quad \forall \varepsilon \leqq \varepsilon_0.$$

Now for any $x \in \mathbb{R}^n$,

$$\begin{aligned} E_x^\varepsilon\{X(T_0) \in K(\eta, \xi)\} &\geqq E_x^\varepsilon\{E_{X(\tau)}^\varepsilon\{X(T_0 - \tau) \in K(\eta, \xi)\}, \tau \leqq T_2\} \\ &\qquad (\text{for } \tau \leqq T_2,\ T_1 \leqq T_0 - \tau \leqq T_0, \text{ and } X(\tau) \in B(2R_0)) \\ &\geqq \left(1 - \frac{\delta}{2}\right) E_x^\varepsilon\{\tau \leqq T_2\} \\ &\geqq \left(1 - \frac{\delta}{2}\right)\left(1 - \frac{c_1}{T_2}\right) > 1 - \delta. \end{aligned}$$

LEMMA 6.1. *Let $p_t^\varepsilon(x, y)$ denote the transition density of* (6.2) *and define*

$$q_t^\varepsilon(x, y)\pi^\varepsilon(y) = p_t^\varepsilon(x, y).$$

Then for any x_0, y_0 in $\mathbb{R}^n$, $\varepsilon \leqq \varepsilon_0$, $t > 0$,

$$q_{t+2T_0}^\varepsilon(x_0, y_0) \geqq \inf_{x, y \in K(\eta, \xi)} q_t^\varepsilon(x, y)(1 - \delta)^2,$$

the relation between δ, T_0, ε_0 is the same as in (6.3). *And one may take any fixed δ, say $\delta = \frac{1}{2}$.*

Proof. For $\varepsilon < 1$, by a similar argument as in Lemma 4.1, $X(t)$ has no explosion. By the Girsanov theorem it is obvious that $X(t)$ has transition densities.

Since the infinitesimal generator $(\varepsilon^2/2)\Delta - \nabla U \cdot \nabla$ is self-adjoint in the weighted space $L^2(\mathbb{R}^n, \pi^\varepsilon)$, it is not hard to show that

$$q_t^\varepsilon(x, y) = q_t^\varepsilon(y, x), \tag{6.4}$$

$$
\begin{aligned}
q^\varepsilon_{t+2T_0}(x_0, y_0) &= \int p^\varepsilon_{T_0}(x_0, x)p^\varepsilon_t(x, y)q^\varepsilon_{T_0}(y, y_0)\,dx\,dy \\
&\geqq \int_{x,y\in K(\eta,\xi)} p^\varepsilon_{T_0}(x_0, x)q^\varepsilon_t(x, y)\pi^\varepsilon(y)q^\varepsilon_{T_0}(y, y_0)\,dx\,dy \\
&\geqq \inf_{x,y\in K(\eta,\xi)} q^\varepsilon_t(x, y)p^\varepsilon_{T_0}(x_0, K(\eta, \xi)) \\
&\quad\cdot \int_{K(\eta,\xi)} q_{T_0}(y, y_0)\pi^\varepsilon(y)\,dy \quad \text{(by 6.4)} \\
&= \inf_{x,y\in K(\eta,\xi)} q^\varepsilon_t(x, y)p^\varepsilon_{T_0}(x_0, K(\eta, \xi))p^\varepsilon_{T_0}(y_0, K(\eta, \xi)) \\
&\geqq (1-\delta)^2 \inf_{x,y\in K(\eta,\xi)} q^\varepsilon_t(x, y), \qquad \varepsilon \leqq \varepsilon_0.
\end{aligned}
$$

This completes the proof.

LEMMA 6.2 (Sheu [13, Cor. 2.5]).

$$
\lim_{\varepsilon\to 0} \varepsilon^2 \log p^\varepsilon_t(x, y) \geqq -I(t, x, y)
$$

uniformly for x, y in a compact set.

COROLLARY 6.1. *For any $t>0$, $\alpha>0$, there is $\varepsilon_0>0$ such that for $\varepsilon\leqq\varepsilon_0$, $x_0, y_0\in\mathbb{R}^n$*

$$
q^\varepsilon_{t+2T_0}(x_0, y_0) \geqq \exp\left(-\frac{1}{\varepsilon^2}(J(t, \eta, \xi)+\alpha)\right).
$$

LEMMA 6.3 (Super normal case.) *For a fixed $t>0$, let $T=t+2T_0$. Then $\forall\alpha>0$, $\forall M>0$ there is $\varepsilon_0>0$ such that for $\varepsilon\leqq\varepsilon_0$*

$$
|p^\varepsilon_{mT}(x, f) - \pi^\varepsilon(f)| < 4\|f\|\exp(-M),
$$

where

$$
m = M\exp\left(\frac{1}{\varepsilon^2}(J(t, \eta, \xi)+\alpha)\right).
$$

Proof. Let $\beta = \exp(-1/\varepsilon^2(J(t, \eta, \xi)+\alpha))$.

$$
\begin{aligned}
&p^\varepsilon_{mT}(x_1, f) - p^\varepsilon_{mT}(x_2, f) \\
&\quad= \int p^\varepsilon_T(x_1, z)p^\varepsilon_{(m-1)T}(z, f)\,dz - \int p^\varepsilon_T(x_2, z)p^\varepsilon_{(m-1)T}(z, f)\,dz \\
&\quad= \int q_T(x_1, z)\pi^\varepsilon(z)p^\varepsilon_{(m-1)T}(z, f)\,dz \\
&\qquad\qquad - \int q^\varepsilon_T(x_2, z)\pi^\varepsilon(z)p^\varepsilon_{(m-1)T}(z, f)\,dz \\
&\quad= \int (q^\varepsilon_T(x_1, z)-\beta)\pi^\varepsilon(z)p^\varepsilon_{(m-1)T}(z, f)\,dz \\
&\qquad\qquad - \int (q^\varepsilon_T(x_2, z)-\beta)\pi^\varepsilon(z)p_{(m-1)T}(z, f)\,dz \\
&\quad\leqq (1-\beta)\,(\max_z p^\varepsilon_{(m-1)T}(z, f) - \min_x p^\varepsilon_{(m-1)T}(x, f)) \\
&\quad= (1-\beta)\sup_{x_1, x_2\in\mathbb{R}^n} |p^\varepsilon_{(m-1)T}(x_1, f) - p^\varepsilon_{(m-1)T}(x_2, f)|.
\end{aligned}
$$

By induction,

$$\sup_{x_1, x_2 \in \mathbb{R}^n} |p^\varepsilon_{mT}(x_1, f) - p^\varepsilon_{mT}(x_2, f)| \leqq 2\|f\|(1-\beta)^{[m]}.$$

Since π^ε is the invariant measure of $p^\varepsilon_t(x, y)$[16, p. 243],

$$|\pi^\varepsilon(f) - p^\varepsilon_{mT}(x, f)| \leqq \left| \int \pi^\varepsilon(z)(p^\varepsilon_{mT}(z, f) - p^\varepsilon_{mT}(x, f))\, dz \right|$$
$$\leqq 2(1-\beta)^{[m]}\|f\|.$$

General case. In order to compare the general case with the super normal case, we need the following lemma.

LEMMA 6.4. *Let* $B(r) = \{x | U(x) \leqq r\}$ *and* $\tau_r = \inf\{t | X(t) \notin B(r)\}$. *Then there exists* $c(r)$ *for large* r

(i) $\quad c(r) \to \infty \quad$ *as* $r \to \infty$,

(ii) $\quad \lim p^\varepsilon_x\left\{\tau_r > \exp\left(\frac{1}{\varepsilon^2} c(r)\right)\right\} = 1 \quad$ *uniformly for* $x \in K(\eta, \xi) \subseteq B(r)$.

Suppose that Lemma 6.4 holds. Choose r large enough such that

$$c(r) > J(t, \eta, \xi) + 1, \qquad K(\eta, \xi) \subset B(r).$$

Let $\hat{U}$ satisfy (6.1) for $R_0 > r$ and $\hat{U} = U$ on $B(r)$. Let $\hat{}$ denote the modified version.

$$|p^\varepsilon_{mT}(x, f) - \pi^\varepsilon(f)| \leqq |p^\varepsilon_{mT}(x, f) - \hat{p}^\varepsilon_{mT}(x, f)|$$
$$+ |\hat{p}^\varepsilon_{mT}(x, f) - \hat{\pi}^\varepsilon(f)| + |\hat{\pi}^\varepsilon(f) - \pi^\varepsilon(f)|.$$

The second term goes to zero by Lemma 6.3. Since $\hat{\pi}^\varepsilon$ and π^ε have the same weak limit, the third term also tends to zero.

$$|p^\varepsilon_{mT}(x, f) - \hat{p}^\varepsilon_{mT}(x, f)| \leqq 2\|f\| E^\varepsilon_x\{\tau_r \leqq mT\} \to 0 \quad \text{as } \varepsilon \to 0$$

by Lemma 6.4.

Proof of Lemma 6.4. Choose r_0 such that $K(\eta, \xi) \subseteq B(r_0) =: \Omega_1$ and $\Omega_2 :=$ $B(r_0 + 1) \subseteq B(r) =: \Omega_3$. Define

$$\sigma_1 = \inf\{t | X(t) \in \Omega_1\},$$
$$\theta_1 = \inf\{t > \sigma_1 | X(t) \notin \Omega_2\},$$
$$\vdots$$
$$\sigma_m = \inf\{t > \theta_{m-1} | X(t) \in \Omega_1\},$$
$$\theta_m = \inf\{t > \sigma_m | X(t) \notin \Omega_2\}.$$

If one can prove that before exit from Ω_3, the path spends a lot of time jumping between Ω_1 and Ω_2, then τ_r will have a good lower estimate.

Let $U(x) = r_0 + 1$ and Q^ε_x denote the measure of the zero drift process, then

$$p^\varepsilon_x\{\tau_r < \sigma_1\} = Q^\varepsilon_x\left\{\tau_r < \sigma_1, \exp\left(\frac{1}{\varepsilon^2}\int_0^{\tau_r} (-\nabla U(X(s))\, dX(s) - \frac{1}{2\varepsilon^2}\int_0^{\tau_r} |\nabla U(X(s))|^2\, ds)\right)\right\}$$
$$= Q^\varepsilon_x\left\{\tau_r < \sigma_1, \exp\left(-\frac{1}{\varepsilon^2}\{U(X(\tau_r)) - U(x)\} - \frac{1}{2\varepsilon^2}\int_0^{\tau_r} (|\nabla U(X(s))|^2 - \varepsilon^2 \Delta U(X(s)))\, ds\right)\right\}.$$

For $s \leqq \tau_r < \sigma_1$, $U(X(s)) > r_0$, then there exist $M_1 > 0$ and $M_2 > 0$ such that

$$\begin{aligned}
&|\nabla U(X(s))|^2 - \varepsilon^2 \Delta U(X(s)) \\
&\quad = \varepsilon^2(|\nabla U(X(s))|^2 - \Delta U(X(s))) + (1-\varepsilon^2)|\nabla U(X(s))|^2 \\
&\quad \geqq -\varepsilon^2 M_1 + (1-\varepsilon^2) M_2 > 0 \quad \text{for small } \varepsilon.
\end{aligned}$$

Hence,

$$\begin{aligned}
p_x^\varepsilon\{\tau_r < \sigma_1\} &\leqq Q_x^\varepsilon\left\{\tau_r < \sigma_1, \exp\left(-\frac{1}{\varepsilon^2}(r - r_0 - 1)\right)\right\} \\
&\leqq \exp\left(-\frac{1}{\varepsilon^2}(r - r_0 - 1)\right),
\end{aligned}$$

$$\begin{aligned}
p_x^\varepsilon(\tau_r < \sigma_m) &= \sum_{k=1}^m p_x^\varepsilon(\tau_r < \sigma_k, \tau_r \geqq \sigma_{k-1}) \\
&= \sum_{k=1}^m p_x^\varepsilon\left\{E_{X(\theta_{k-1})}\{\tau_r < \sigma_1\}, \tau_r \geqq \sigma_{k-1}\right\} \\
&\leqq m \exp\left(-\frac{1}{\varepsilon^2}(r - r_0 - 1)\right).
\end{aligned}$$

Now we shall show that σ_1 is not too small. Let

$$T^* = \inf_{U(x) = r_0 + 1} \inf\{t \mid Y(0) = x, Y(t) \in \Omega_1, Y(s) \text{ satisfies (1.3) for } 0 \leqq s \leqq t\}.$$

Let

$$0 < \delta_0 < d\left(\Omega_1, \left\{\begin{matrix} Y(T^*/2) \mid Y(0) = x, U(x) = r_0 + 1, \\ Y(s) \text{ satisfies (1.3) for } 0 \leqq s \leqq T^*/2 \end{matrix}\right\}\right).$$

Let $T_0 \leqq T^*/2$; then by a similar method as in the end of the proof of Lemma 1,

$$\begin{aligned}
p_x^\varepsilon\{\sigma_1 < T_0\} &\leqq p\{\tau < T_0\} \leqq (2n) \exp(-e^{-2dT_0}\delta_0^2/(2nT_0\varepsilon^2)) \\
&\leqq 2n \exp(-e^{-2dT_0}\delta/(T_0\varepsilon^2))
\end{aligned}$$

(n is the dimension and $\delta = \delta_0^2/2n$, $\tau = \inf\{t \mid |X(t) - Y(t)| > \delta_0\}$, d is the corresponding Lipschitz constant of ∇U in a compact set).

Then, it is obvious that

$$p_x^\varepsilon\{\sigma_m < mT_0\} \leqq 2nm \exp\left(-e^{-2dT_0}\frac{\delta}{T_0\varepsilon^2}\right),$$

$$\begin{aligned}
p_x^\varepsilon\{\tau_r < mT_0\} &\leqq p_x^\varepsilon\{\tau_r < \sigma_m\} + p_x^\varepsilon\{\sigma_m < mT_0\} \\
&\leqq m \exp\left(-\frac{1}{\varepsilon^2}(r - r_0 - 1)\right) + 2nm \exp\left(-\frac{1}{\varepsilon^2}\frac{e^{-2dT_0}\delta}{T_0}\right).
\end{aligned}$$

Choose T_0 such that

$$\frac{e^{-2dT_0}\delta}{T_0} > (r - r_0 - 1).$$

And choose $m-1=[\exp(1/\varepsilon^2(r-r_0-1-v))$, where v is an arbitrary fixed small positive number

$$p_x^\varepsilon\left\{\tau_r \geqq \exp\left(\frac{1}{\varepsilon^2}(r-r_0-1-v)\right)T_0\right\}$$

$$>1-(2n+3)\exp\left(-\frac{v}{\varepsilon^2}\right)\to 1 \quad \text{as } \varepsilon\to 0.$$

Hence we may choose $c(r)=r-r_0-1-v$ for any fixed $v>0$.

For $x\in K(\eta,\xi)$,

$$p_x^\varepsilon\{\tau_r<\exp(c(r)/\varepsilon^2)\}=p_x^\varepsilon\{E_{X(\theta)}\{\tau_r<\exp(c(r)/\varepsilon^2)\},\theta<\exp(c(r)/\varepsilon^2)\}.$$

$$\to 0 \quad \text{as} \quad \varepsilon\to 0,$$

where $\theta=\inf\{t \mid U(X(t))=r_0+1\}$.

7. Appendix.

(1) Properties of $I(t,x,y)$ can be found in [2] and [14].

(2) $c_0<\infty$ is obvious. In fact if we consider

$$\begin{aligned}\phi(u)&=x+u(z-x) \quad \text{for } 0\leqq u\leqq 1,\\ &=z \quad \text{for } 1\leqq u\leqq t-1,\\ &=z+(u-(t-1))(y-z) \quad \text{for } t-1\leqq u\leqq t,\end{aligned}$$

where z is a stationary point, then it is easy to see why $c_0<\infty$.

(3) Suppose that the set S of all stationary points has only finitely many, say l, connected components. We also assume that points in each component can be connected by smooth curves. Choose ε small enough such that $S(\varepsilon)$ has l disjoint components $S_1,\cdots,S_l$. For x,y belong to the same S_i,

$$I(t,x,y)=O(1/t)+O(\varepsilon).$$

Indeed, connect x to a stationary point v in S_i with $|x-v|<2\varepsilon$ by straight line for time interval $[0,1]$. Connect y to a stationary point w in S_i with $|y-w|<2\varepsilon$ by straight line for time interval $[t-1,t]$. Connect v,w by a fixed smooth curve ψ with all the curve stationary points. Rescale the parameter of the curve to the interval $[1,t-1]$ and denote it by ϕ, then

$$\int_1^{t-1}|\dot\phi(u)+\nabla U(\phi(u))|^2\,du=\int_1^{t-1}|\dot\phi(u)|^2\,du$$

$$=\frac{1}{t-2}\int_0^1|\dot\psi(u)|^2\,du.$$

If $\phi(t-u)$, $0\leqq u\leqq t$, is a solution, then

$$\int_0^t|\dot\phi(u)+\nabla U(\phi(u))|^2\,du=0.$$

For any starting point x, and end point y, since U is a Lyapunov function for the dynamical system (1.4), within a fixed finite time x will reach some S_i via a solution of (1.4) and y some S_j via a solution. If ϕ is a solution, then

$$\frac{1}{4}\int_0^t |\dot{\phi}(u)+\nabla U(\phi(u))|^2\,du = U(\phi(t)) - U(\phi(0)).$$

(4) A good upper bound for $I(t, x, y)$ is a curve $\phi(0)=x$, $\phi(t)=y$ and ϕ spends most of its time at a stationary point. From (3), we only have to count the contribution from connecting different components. Consider $\{S_i\}_{i=1,\cdots,l}$ as nodes of a graph, and define S_i and S_j as neighboring nodes if there is a trajectory of (1.4) connecting S_i and S_j. Suppose $S_1, S_2, \cdots, S_m$, $m \leqq l$, are in the same connected component, and assume there exist points $x_1, \bar{x}_2, x_2, \bar{x}_3, x_3, \cdots, \bar{x}_{m-1}, x_{m-1}, \bar{x}_m$ in $S_1, \cdots, S_m$, respectively, such that a trajectory connects x_i to $\bar{x}_{i+1}$ (see Fig. 1). If $U(\bar{x}_{i+1}) > U(x_i)$, then the contribution

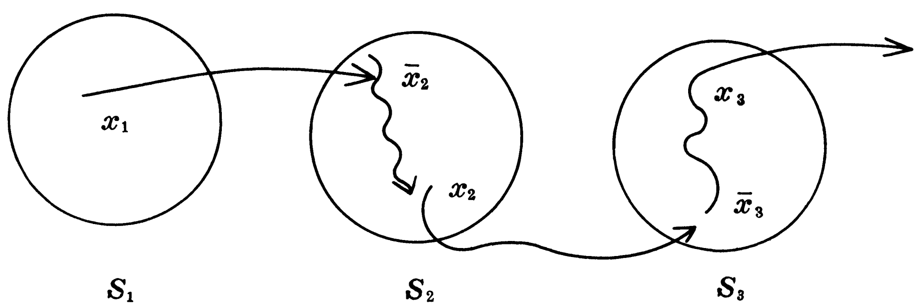

FIG. 1

is $2(U(\bar{x}_{i+1}) - U(x_i))$. Otherwise, it is a free ride. Of course, there are other paths to connect $S_1, \cdots, S_m$. Note that $U(x_j)$, $U(\bar{x}_j)$ and $U(z_j)$ are almost of the same value, where z_i is a stationary point in S_j. Two consecutive increasing trajectories contribute

$$2(U(\bar{x}_{i+2}) - U(x_{i+1}) + U(\bar{x}_{i+1}) - U(x_i)) \sim 2(U(\bar{z}_{i+2}) - U(z_i).$$

Hence, $2([m/2]\max_{1\leqq j\leqq m} U(z_j))$ is a bound.

(5) Suppose the graph $\{S_j\}_{1\leqq j\leqq l}$ has, say, $G_1, \cdots, G_k$ components. Let K be any bounded connected set containing all S_j's. For any $x \in K$ either x in some G_i or there exists a unique G_i such that x is connected to G_i by a trajectory. Now we have partitioned K into $K_1, \cdots, K_k$ disjoint sets. Define K_i, K_j are neighbors if $d(K_i, K_j) = 0$. If we regard $\{K_j\}_{j=1,\cdots,k}$ as nodes of a graph, then we can show it is a connected graph since K is connected. In other words we can connect G_i to G_j via trajectories of (1.4) and at most $k-1$ line segments of arbitrary small length. The same argument as in 4 yields that the contribution to connect different components is less than

$$2\left(\left[\frac{k}{2}\right]\max_{1\leqq i\leqq l} U(z_i)\right).$$

(6) One may use $3([l/2]+[k/2])\max_{1\leqq i\leqq l} U(z_i)$ as a rough bound for c_0.

(7) We give some examples (see Figs. 2-5) to calculate c_* for the one-dimensional case.

(i)

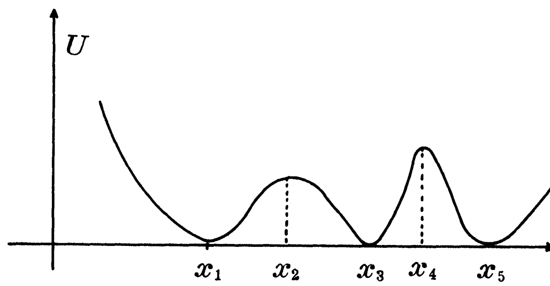

FIG. 2. $c_* = (U(x_2) - U(x_1)) + (U(x) - U(x_3)) = (U(x_2) - U(x_3)) + (U(x_4) - U(x_5))$.

(ii)

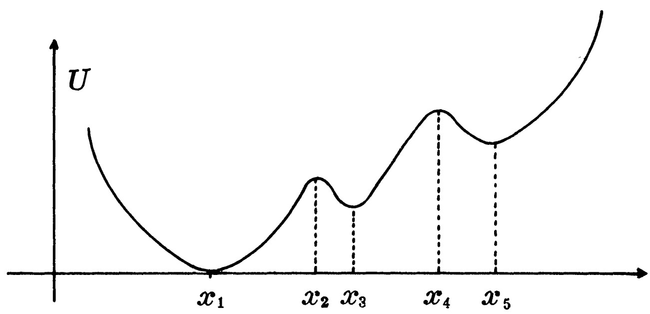

FIG. 3. $c_* = (U(x_2) - U(x_1)) + (U(x_4) - U(x_3)) - U(x_5) = (U(x_4) - U(x_5)) + (U(x_2) - U(x_3)), (U(x_1) = 0)$.

(iii)

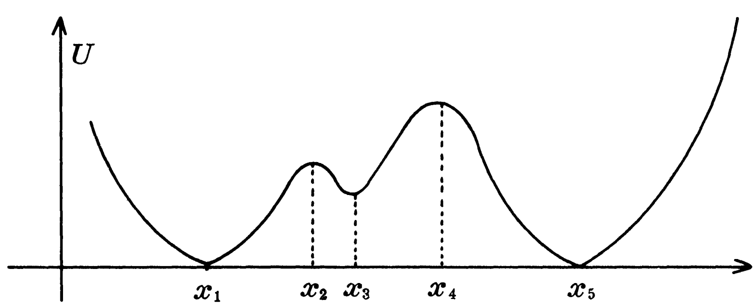

FIG. 4. $c_* = (U(x_2) - U(x_1)) + (U(x_4) - U(x_3)) = (U(x_4) - U(x_5)) + (U(x_2) - U(x_3))$.

(iv)

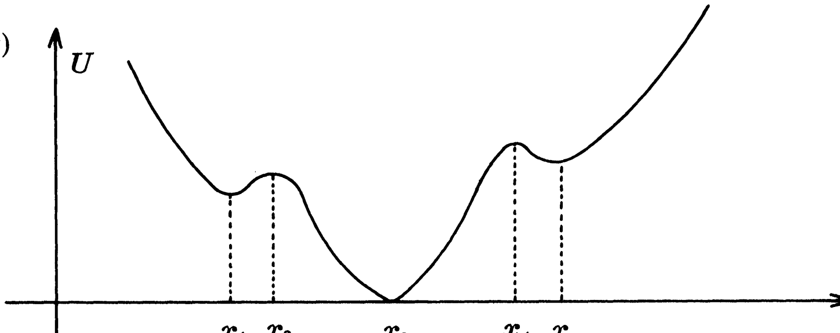

FIG. 5. $c_* = (U(x_2) - U(x_1)) + (U(x_4) - U(x_3)) - U(x_5) = (U(x_4) - U(x_5)) + (U(x_2) - U(x_3)) - U(x_1)$.

Acknowledgment. The authors would like to thank Professor Daniel Stroock for very helpful suggestions and discussions.

REFERENCES

[1] V. ČERNÝ, *A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm*, preprint, Inst. of Physics and Biophysics, Comenius Univ., Bratislava, 1982.

[2] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer-Verlag, Berlin, New York, 1984.

[3] S. GEMAN AND D. D. GEMAN, *Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. and Machine Intelligence, 6 (1984), pp. 721–741.

[4] S. GEMAN AND C.-R. HWANG, *Diffusion for global optimization*, this Journal, to appear.

[5] B. GIDAS, *Non-Stationary Markov chains and convergence of the annealing algorithm*, J. Statist. Phys., 39 (1985), pp. 73–131.

[6] ———, *Global minimization via the Langevin equation*, in preparation.

[7] U. GRENANDER, *Tutorial in Pattern Theory*, Brown Univ., Providence, RI, 1983.

[8] B. HAJEK, *Cooling schedules for optimal annealing*, Dept. of Electrical & Computer Engineering and the Coordinated Science Lab., Univ. of Illinois, Champaign-Urbana, IL, 1985.

[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*, Science, 220 (1983), pp. 621–680.

[11] H. J. KUSHNER, *Asymptotic global behavior for stochastic approximations and diffusion with slowly decreasing noise effects: global minimization via Monte Carlo*, preprint, Div. of Applied Mathematics, Brown Univ., Providence, RI, 1985.

[12] G. PARISI, *Prolegomena to any further computer evaluation of the QCD mass spectrum*, in Progress in Gauge Field Theory, Cargese, 1983.

[13] S.-J. SHEU, *Asymptotic behavior of transition density of diffusion Markov process with small diffusion*, Stochastics, 13 (1984), pp. 131–163.

[14] ———, *Asymptotic behavior of invariant density of diffusion Markov process with small diffusion*, SIAM J. Math Anal., 12 (1985), pp. 451–460.

[15] D. W. STROOCK AND S. R. S. VARADHAN, *Multidimensional Diffusion Processes*, Springer-Verlag, Berlin, New York, 1979.

[16] S. R. S. VARADHAN, *Lectures on Diffusion Problems and Partial Differential Equations*, Tata Inst., Bombay, 1980.