Math. Proc. Camb. Phil. Soc. (1989), **105**, 177
Printed in Great Britain

On multi-dimensional annealing problems

BY TERENCE CHAN
Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB

(*Received* 20 *January*; *revised* 27 *April* 1988)

1. *Introduction*

In [1], Chan and Williams considered a one-dimensional diffusion of the form

$$dY_t = -F(Y_t)\,dt + \epsilon(t)\,dB_t,$$

where F is a strictly increasing continuous function with $F(0) = 0$ and ϵ is a decreasing deterministic function such that $\epsilon(0)$ is finite and $\epsilon(t) \downarrow 0$ as $t \uparrow \infty$, and gave necessary and sufficient conditions for $Y_t \to 0$ a.s. as $t \to \infty$.

The aim of the present paper is to extend the results of Chan and Williams [1] to multi-dimensional diffusions and obtain analogous necessary and sufficient conditions for almost sure convergence. Consider now an n-dimensional diffusion process $\tilde{Y}$ driven by a d-dimensional Brownian motion $\tilde{W}$, satisfying the following stochastic differential equation:

$$d\tilde{Y}_t = -F(\tilde{\epsilon}(t), \tilde{Y}_t)\,dt + \tilde{\epsilon}^T(t)\,d\tilde{W}_t, \tag{1·1}$$

where $\tilde{\epsilon}: [0, \infty) \to \mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$ (space of $d \times n$ matrices) and $\|\tilde{\epsilon}(\cdot)\|$ is a decreasing function on $[0, \infty)$ such that $\|\tilde{\epsilon}(t)\| \downarrow 0$ as $t \uparrow \infty$ and $\|\tilde{\epsilon}(0)\|$ is finite, and where $F: \mathscr{L}(\mathbb{R}^n, \mathbb{R}^d) \times \mathbb{R}^n \to \mathbb{R}^n$ is a function which satisfies the following properties:

(i) $F(0, 0) = 0$;

(ii) for all fixed $\epsilon \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$, there exists a positive constant $a \leqslant 1$ such that for all y

$$F(\epsilon, y).y \geqslant a|F(\epsilon, y)|\,|y|;$$

(iii) there exists a function $f: \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ such that $|F(\epsilon, y)| \geqslant f(\|\epsilon\|, |y|)$ and such that for all $e \geqslant 0$, $f(e, \cdot)$ is a strictly increasing function with $f(e, 0) = 0$ and for all y, $f(\cdot, |y|)$ is a non-increasing function.

These assumptions on F and $\tilde{\epsilon}$ will hold throughout the rest of this paper.

The last two conditions correspond to the assumption that F is strictly increasing in the one-dimensional case, where the drift is negative (resp. positive) whenever the process is positive (resp. negative). Condition (ii) implies that the angle between the drift vector and the current position of the process is bounded away from $\pm\pi/2$ so the drift would always tend to pull the process nearer towards the origin. The last condition just says that the size of the drift tends to increase with distance from the origin. (The condition that $f(\cdot, |y|)$ is non-increasing is a general 'good-behaviour' type condition.)

We are allowing the drift F to depend explicitly on the 'cooling schedule' $\tilde{\epsilon}$. This is motivated by the work of Durran and Truman on Nelsonian diffusions and stochastic mechanics which is due to appear shortly, and as there has been no attempt to consider such types of problems in the literature on annealing which have appeared to date, the present work should be of some interest.

In most applications one is interested in the case where $F = \nabla U$ for some potential

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

U which has several wells, one of which (the global minimum, say) is located at the origin. In this case, (i)–(iii) is only satisfied locally in some ball around the origin, and one can no longer use Theorem 1 below to say that $\tilde{Y}_t \to 0$ almost surely, but only that $\tilde{Y}_t \to 0$ with positive probability, since there is now a positive probability that the process will tunnel out of the well at the origin.

For a comprehensive survey of applications of annealing and the various results already known, we refer the reader to Hajek [3]. We note in particular that the results on convergence derived so far all concern convergence in probability, rather than almost sure convergence.

Chiang *et al.* [2] have studied the equation (1·1) for the case where $F = \nabla U$ for some potential U and $\tilde{\epsilon}(t) = e(t)I$ for some (real-valued) function e, in connection with the problem of finding global minima of U by simulated annealing. They give a sufficient condition for weak convergence of the transition density of $\tilde{Y}$ to a measure concentrated on the global minima of U.

In this paper we shall prove the following theorem, giving necessary and sufficient conditions for almost sure convergence to 0 of the solution $\tilde{Y}$ of (1·1).

THEOREM 1. $\tilde{Y}_t \to 0$ *(a.s.) if and only if*

$$\mathrm{I}(c) := \int_0^\infty \exp\{-c\|\tilde{\epsilon}(u)\|^{-2}\}\,du < \infty \quad \textit{for all} \quad c > 0. \tag{1·2}$$

(We shall use the symbol ' := ' to mean 'is defined to be'.) As was proved in [1], the condition (1·2) is equivalent to

$$\|\tilde{\epsilon}(u)\|^2 \log u \to 0. \tag{1·3}$$

Here and elsewhere, $\|\cdot\|$ denotes the natural norm on $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$, but, of course, the theorem is valid for any norm since all norms are equivalent in a finite-dimensional vector space.

The condition (1·2) is identical to that obtained in Chan and Williams [1] for the one-dimensional case, with $\|\tilde{\epsilon}(u)\|$ replacing the ordinary scalar-valued function. This is similar to the condition that was derived in Hajek [4] for convergence in probability to the global minimum of some potential function in the finite state case: there it was proved that $\mathrm{I}(c_i) = \infty$ for all i, where c_i are the depths of all local (but not global) minimum wells, is a necessary and sufficient condition for escaping from all wells of the local (but non-global) minima of the potential function, and is hence a necessary and sufficient condition for convergence in probability to the global minimum. However, Theorem 1 shows that, irrespective of the depth of a well, the process will either converge pathwise (with positive probability) to the minimum of that well or it will not converge pathwise to the minimum of *any* well, according as (1·2) converges or diverges.

In the present problem, the question of 'tunnelling' (that is, escaping) out of wells is rather more complicated. As will become apparent in the proof of Theorem 1, the constant c in (1·2) is not the actual depth of the well, but some function of it. For the finite state case considered in Hajek [4], we can think of the drift function F as a piecewise linear function, so the constant c in (1·2) *is* indeed the depth of the well. Chan and Williams [1] considered an example where the drift is linear (so the corresponding potential is then of the form $U(x) = Ax^2$); the process was actually shown to oscillate *boundedly* about the well in the critical case in which the rate of

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

cooling is given by $\tilde{c}(t) = k(\log t)^{-\frac{1}{2}}$. The constant k in that example turns out to be proportional to the size of the oscillation:

$$\liminf \tilde{Y}_t = -(2A)^{-\frac{1}{2}}k, \quad \limsup \tilde{Y}_t = (2A)^{-\frac{1}{2}}k.$$

Because $\mathrm{I}(c) = \infty$ for $c \leqslant k^2$, by Theorem 1 the process does not converge to the bottom of the well, but if the depth of the well is greater than $k^2/2$, so that the potential barrier of the well is sufficiently high, it does not escape from the well either. Since in that example $\mathrm{I}(c) < \infty$ for $c > k^2$, we see that the condition (1·2) – modulo a constant factor of 2 due to a difference in normalization – exactly agrees with the result of Hajek [4] in so far as escape from wells is concerned. However, the question of what choice of constant c will give rise to tunnelling in the general case where the drift is non-linear is as yet unresolved and it is hoped that this problem will be addressed in a later paper. Here we simply make the following conjecture: that the critical choice of constant giving rise to tunnelling is just the depth of the well exactly as in the linear drift case, regardless of the form of the drift coefficient F.

There is thus a kind of asymmetry between known results on weak convergence and our result: on the one hand there is a critical rate of cooling (essentially $(\log t)^{-\frac{1}{2}}$) which will ensure escape from wells of local minima and convergence in probability to the global minimum well; on the other hand the same rate of cooling is critical for (strong) convergence to *any* well. For more on escape times from wells, see Hajek [3].

2. *The basic machinery: comparison processes*

Much of the proof of Theorem 1 will follow the method used already by Chan and Williams [1] for the one-dimensional case. Indeed, the method of proof will be to reduce the problem to a one-dimensional one. Here we give only a brief description of the so-called 'comparison process'. For a more detailed treatment of how these processes are constructed and proofs of results quoted below, we refer the reader to Chan and Williams [1].

Let b be a positive constant and let B be a Brownian motion on $\mathbb{R}$. Consider reflecting Brownian motion with drift $-b$ defined via the stochastic differential equation

$$dX_t = -b\,dt + dB_t + dL_t, \quad X_0 = 0,$$

where L is the local time at 0 for X:

$$L_t = \sup\{-(B_s - bs) : s \leqslant t\}.$$

Let V be an exponentially distributed random variable of rate λ which is independent of B. Define

$$T := \inf\{t : L_t > V\}.$$

The process $\{X_t : t < T\}$ is elastic Brownian motion with drift $-b$ and killing rate λ and will be denoted by $\mathrm{EBMD}(-b, \lambda; B)$.

Now take a sequence of positive constants (b_k) with $b_k \uparrow \infty$ and a family of independent random variables (V_k), each V_k being exponentially distributed with rate

$$\lambda_k := 1/b_k.$$

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

Let B be a Brownian motion independent of (V_k) and define

$$S_0 := 0;$$

$$\{X_t^{(k)} : t < T_k\} := \text{EBMD}(-b_k, \lambda_k; B(S_{k-1}+\cdot)) \quad (k \geqslant 1),$$

$$S_k := \sum_{i=1}^{k} T_i.$$

Finally we define

$$Z_t := X^{(k)}(t - S_{k-1}) \quad (t \in [S_{k-1}, S_k))$$

and we write for convenience $Z = \text{COMP}(-b_k, \lambda_k : k \in \mathbb{N}; B)$. The process Z satisfies the stochastic differential equation

$$dZ_t = -h(t)\,dt + dB_t + dL_t^Z, \tag{2·1}$$

where h is the function

$$h(t) = b_k \quad (t \in [S_{k-1}, S_k)) \tag{2·2}$$

and L^Z is the local time at 0 for Z.

It is easy to show that

$$S_k/k \to 1 \quad \text{(a.s.)},$$

so that for all $\eta > 0$

$$(1-\eta)\,k < S_k < (1+\eta)\,k \tag{2·3}$$

for all sufficiently large k. A simple argument using the Borel–Cantelli lemmas shows

$$Z_t \to 0 \quad \text{a.s.} \quad \text{if and only if} \quad \sum_k \exp(-cb_k) < \infty \quad \text{for all} \quad c > 0 \tag{2·4}$$

(see Chan and Williams [1]).

It is somewhat surprising that such an innocuous-looking method should prove to be so powerful and to have such wide applicability. Of course, this method of comparison processes has a subtlety that belies its apparent simplicity, for, in spite of the similarity between the proofs below of the 'if' and the 'only if' parts of Theorem 1, one must remember that a different underlying driving Brownian motion is being considered in each case and that a different Brownian motion is used each time to construct the necessary comparison process. That it is possible to change with such ease the basic 'parameters' used to construct these comparison processes is crucial to the derivation of *both* upper and lower bounds for the process $\tilde{Y}$. Also, in the following, notice the way in which the effects of terms involving the local time – which one might expect would cause a lot of trouble – all conveniently (and some would say magically!) vanish. Here again, it is the choice of the comparison process which is the key ingredient behind this piece of trickery.

3. *Proof of the 'if' part of Theorem* 1

By Itô's formula,

$$d|\tilde{Y}_t| = -\frac{1}{2|\tilde{Y}_t|}\left\{2\tilde{Y}_t \cdot F(\tilde{\epsilon}(t), \tilde{Y}_t) - \text{tr}(\tilde{\epsilon}(t)\,\tilde{\epsilon}^T(t)) + \frac{|\tilde{\epsilon}(t)\,\tilde{Y}_t|^2}{|\tilde{Y}_t|^2}\right\}dt + \frac{|\tilde{\epsilon}(t)\,\tilde{Y}_t|^2}{|\tilde{Y}_t|^2}\,d\tilde{B}_t, \tag{3·1}$$

where $d\tilde{B}_t = |\tilde{\epsilon}(t)\tilde{Y}_t|^{-1}\{(\tilde{\epsilon}(t)\tilde{Y}_t) \cdot d\tilde{W}_t\}$ is a one-dimensional Brownian motion. Define

$$\gamma_t := \inf\left\{s : \int_0^s \frac{|\tilde{\epsilon}(u)\,\tilde{Y}_u|^2}{|\tilde{Y}_u|^2}\,du > t\right\}.$$

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

(We are assuming that

$$\int_0^\infty \frac{|\tilde{\epsilon}(u)\tilde{Y}_u|^2}{|\tilde{Y}_u|^2}du = \infty$$

as usual; the case when this last integral is finite can easily be handled independently as a separate case.) Set

$$\epsilon(t) = \tilde{\epsilon}(\gamma_t),\ Y_t = \tilde{Y}_{\gamma(t)} \quad \text{and} \quad B_t = \int_0^{\gamma(t)} \frac{|\tilde{\epsilon}(u)\,\tilde{Y}_u|}{|\tilde{Y}_u|}d\tilde{B}$$

(so B is a Brownian motion). Then (3·1) takes the form

$$\begin{aligned} d|Y_t| &= -\frac{|Y_t|}{2|\epsilon(t)Y_t|^2}\left\{2Y_t.F(\epsilon(t), Y_t) - \mathrm{tr}(\epsilon(t)\,\epsilon^T(t)) + \frac{|\epsilon(t)Y_t|^2}{|Y_t|^2}\right\}dt + dB_t \\ &= -g(t, Y_t)\,dt + dB_t \quad \text{(say)}. \end{aligned} \tag{3·2}$$

(We leave the reader to check that the singularity in $g(t, y)$ when $y = 0$ does not cause any problems.)

It is sufficient to prove that, for all $\delta > 0$,

$$\limsup |Y_t| \leqslant \delta. \tag{3·3}$$

Now $\sqrt{\mathrm{tr}\,(\epsilon\epsilon^T)}$ is a norm on $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$ which must be equivalent to $\|\cdot\|$. Hence there exist positive constants b, b' such that

$$b'\|\epsilon\|^2 \leqslant \mathrm{tr}(\epsilon\epsilon^T) \leqslant b\|\epsilon\|^2$$

for all $\epsilon \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$. Since

$$Y_t.F(\epsilon(t), Y_t) \geqslant a|Y_t|\,|F(\epsilon(t), Y_t)| \geqslant a|Y_t|f(\|\epsilon(t)\|, |Y_t|),$$

and since $\|\epsilon(t)\| \downarrow 0$ as $t \to \infty$, we have, for $|Y_t| > \delta$,

$$a|Y_t|\,|F(\epsilon(t), Y_t)| - \mathrm{tr}\,(\epsilon(t)\,\epsilon^T(t)) \geqslant 0$$

for all t sufficiently large. Hence, for t sufficiently large,

$$\begin{aligned} g(t, Y_t) &\geqslant \frac{|Y_t|}{2|\epsilon(t)Y_t|^2}\{2a|Y_t|\,|F(\epsilon(t), Y_t)| - \mathrm{tr}(\epsilon(t)\,\epsilon^T(t))\} + \frac{2}{|Y_t|} \\ &\geqslant \frac{|Y_t|}{2\|\epsilon(t)\|^2\,|Y_t|^2}\{2a|Y_t|f(\|\epsilon(t)\|, |Y_t|) - \mathrm{tr}(\epsilon(t)\,\epsilon^T(t))\} \\ &\geqslant \frac{af(\|\epsilon(t)\|, |Y_t|)}{\|\epsilon(t)\|^2} - \frac{b}{2|Y_t|}. \end{aligned} \tag{3·4}$$

Choose $\alpha > 0$ and $K < 1$. Since $\|\epsilon(t)\| \downarrow 0$, for t sufficiently large, $\|\epsilon(t)\| < \alpha$. Set

$$b_k := \left(\frac{af(\alpha, \delta)}{\|\epsilon(K(k-1))\|^2} - \frac{b}{2\delta}\right) \vee \alpha.$$

Clearly for k sufficiently large

$$b_k = \frac{af(\alpha, \delta)}{\|\epsilon(K(k-1))\|^2} - \frac{b}{2\delta}. \tag{3·5}$$

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

Set $\lambda_k := 1/b_k$ and introduce $Z = \mathrm{COMP}(-b_k, \lambda_k : k \in \mathbb{N}; B)$. Since $\mathrm{I}(c) < \infty$ for all $c > 0$ (where $\mathrm{I}(c)$ is the integral at (1·2)) and $\|\tilde{\epsilon}(t)\| \downarrow 0$, we have

$$\begin{aligned}\int_0^\infty \exp\{-c\|\epsilon(u)\|^{-2}\}\,du &= \int_0^\infty \exp\{-c\|\tilde{\epsilon}(\gamma_u)\|^{-2}\}\,du \\ &\leqslant \int_0^\infty \|\tilde{\epsilon}(s)\|^2 \exp\{-c\|\tilde{\epsilon}(s)\|^{-2}\}\,ds \\ &\leqslant A\int_0^\infty \exp\{-c\|\tilde{\epsilon}(s)\|^{-2}\}\,ds < \infty\end{aligned}$$

for all $c > 0$. Hence $\Sigma \exp\{-cb_k\} < \infty$ for all $c > 0$ and so by (2·4) $Z_t \to 0$ as $t \to \infty$. The rest of the argument from now on is identical to that used by Chan and Williams [1].

We have the following situation:

$$dZ_t = -h(t)\,dt + dB_t + dL_t^Z \quad \text{(from (2·1))} \tag{3·6}$$

$$dY_t = -g(t, Y_t)\,dt + dB_t \tag{3·7}$$

where it is almost surely true that

(i) $Y_t \leqslant \delta$ infinitely often;

(ii) because of (3·4), (3·5) and (2·3), there exists some random time ρ_1 such that $h(t) \leqslant g(t, Y_t)$ for all $t \geqslant \rho_1$ whenever $|Y_t| > \delta$.

We can therefore find a random time $\rho > \rho_1$ at which

$$Y_\rho \leqslant Z_\rho + \delta. \tag{3·8}$$

We now show that

$$Y_t \leqslant Z_t + \delta \quad \text{for all} \quad t \geqslant \rho. \tag{3·9}$$

We employ a pathwise argument as follows, so we regard the sample path ω as fixed.

Assume that for some $\eta > 0$,

$$\tau := \inf\{t > \rho : Y_t \geqslant Z_t + \delta + 2\eta\} < \infty$$

and set

$$\sigma := \sup\{t < \tau : Y_t \leqslant Z_t + \delta + \eta\}.$$

By (3·8), we have $\sigma \geqslant \rho$. For $\sigma \leqslant t \leqslant \tau$ we have $Y_t > \delta + \eta$ so that $g(t, Y_t) \geqslant h(t)$ for all $t \in [\sigma, \tau]$. On subtracting (3·6) from (3·7) and integrating from σ to τ, we see that

$$(Y_\tau - Z_\tau) - (Y_\sigma - Z_\sigma) \leqslant 0,$$

while our assumption that $\tau < \infty$ implies that

$$(Y_\tau - Z_\tau) - (Y_\sigma - Z_\sigma) = \eta > 0,$$

whence a contradiction. Result (3·3) now follows from (3·9) and the fact that $Z_t \to 0$. ∎

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

4. *Proof of the 'only if' part of Theorem* 1

We consider each component $\tilde{Y}_t^i$ of $\tilde{Y}_t$ separately, so we regard i as fixed in what follows. We have

$$\begin{aligned} d\tilde{Y}_t^i &= -F^i(\tilde{\epsilon}(t), \tilde{Y}_t)\,dt + \sum_j \tilde{\epsilon}_j^i(t)\,d\tilde{W}_t^j \\ &= -F^i(\tilde{\epsilon}(t), \tilde{Y}_t)\,dt + |\tilde{\epsilon}(t)e_i|\,d\tilde{B}_t, \end{aligned} \tag{4·1}$$

where e_i is the ith standard basis vector in $\mathbb{R}^n$ and $d\tilde{B}_t = |\tilde{\epsilon}(t)\,e_i|^{-1}\,\Sigma_j\,\tilde{\epsilon}_j^i(t)\,d\tilde{W}_t^j$ is a one-dimensional Brownian motion. (We are using $\tilde{\epsilon}_j^i(t)$ to denote the (j,i)-entry of $\tilde{\epsilon}(t)$.) Define

$$\gamma_t := \inf\left\{s : \int_0^s |\tilde{\epsilon}(u)\,e_i|^2\,du > t\right\}$$

and set

$$Y_t = \tilde{Y}_{\gamma(t)}, \quad \epsilon(t) = \tilde{\epsilon}(\gamma_t) \quad \text{and} \quad B_t = \int_0^{\gamma(t)} |\tilde{\epsilon}(s)e_i|\,d\tilde{B}_s$$

(so B is also a Brownian motion on $\mathbb{R}$). Then (4·1) becomes

$$dY_t^i = -|\epsilon(t)\,e_i|^{-2}\,F^i(\epsilon(t), Y_t)\,dt + dB_t,$$

and by Tanaka's formula

$$d|Y_t^i| = -\operatorname{sgn}(Y_t^i)\,|\epsilon(t)\,e_i|^{-2}\,F^i(\epsilon(t), Y_t)\,dt + dW_t + dL_t^Y, \tag{4·2}$$

where W is the Brownian motion defined by

$$dW_t = \operatorname{sgn}(Y_t^i)\,dB_t$$

and L^Y is the local time at 0 for Y.

Choose $K > 1$ and set

$$b_k := |\epsilon(Kk)\,e_i|^{-2}, \quad \lambda_k := 1/b_k. \tag{4·3}$$

Let $Z = \mathrm{COMP}(-b_k, \lambda_k : k \in \mathbb{N}; W)$; so from (2·1)

$$dZ_t = -h(t)\,dt + dW_t + dL_t^Z. \tag{4·4}$$

Since $Y_t \to 0$, $\|\epsilon(t)\| \downarrow 0$ and $F(0,0) = 0$, there exists a random time ρ_1 such that $F^i(\epsilon(t), Y_t) \leqslant |F^i(\epsilon(t), Y_t)| \leqslant 1$ for all $t \geqslant \rho_1$. From (4·3) and (2·3), we can find a random time $\rho_2 \geqslant \rho_1$ such that

$$h(t) \geqslant |\epsilon(t)\,e_i|^{-2} \quad \text{for all} \quad t \geqslant \rho_2, \tag{4·5}$$

whence

$$\operatorname{sgn}(Y_t^i)\,|\epsilon(t)\,e_i|^{-2}\,F^i(\epsilon(t), Y_t) \leqslant h(t) \quad \text{for all} \quad t \geqslant \rho_2.$$

Let $\rho = \inf\{S_k : S_k \geqslant \rho_2\}$, so $Z_\rho = 0$. Clearly

$$0 = Z_\rho \leqslant |Y_\rho^i|. \tag{4·6}$$

We now show that

$$Z_t \leqslant |Y_t^i| \quad \text{for all} \quad t \geqslant \rho. \tag{4·7}$$

As before, we assume that for some $\eta > 0$

$$\tau := \inf\{t > \rho : Z_t \geqslant |Y_t^i| + 2\eta\} < \infty,$$

and set

$$\sigma := \sup\{t < \tau : Z_t \leqslant |Y_t^i| + \eta\}.$$

From (4·6), $\sigma > \rho$. Between the times σ and τ we have $Z_t \geqslant \eta$, so that $L_\tau^Z - L_\sigma^Z = 0$.

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press

On subtracting (4·2) from (4·4) and integrating from σ to τ, we see that, because of (4·5),

$$(Z_\tau - |Y_\tau^i|) - (Z_\sigma - |Y_\sigma^i|) \leqslant 0,$$

while our assumption that $\tau < \infty$ implies

$$(Z_\tau - |Y_\tau^i|) - (Z_\sigma - |Y_\sigma^i|) = \eta > 0,$$

giving a contradiction.

This establishes (4·7), and hence $Z_t \to 0$ a.s. and by (2·4)

$$\Sigma \exp\{-cb_k\} < \infty \quad \text{for all} \quad c > 0.$$

We thus have for each $i = 1, 2, \ldots, n$

$$\int_0^\infty \exp\{-c|\epsilon(u)\, e_i|^{-2}\}\, du = \int_0^\infty |\tilde{\epsilon}(u)\, e_i|^2 \exp\{-c|\tilde{\epsilon}(u)\, e_i|^{-2}\}\, du < \infty \quad \text{for all} \quad c > 0. \tag{4·8}$$

But for $0 < c_0 < c_1$, the ratio

$$\frac{\exp\{-c_1|\tilde{\epsilon}(u)\, e_i|^{-2}\}}{|\tilde{\epsilon}(u)\, e_i|^2 \exp\{-c_0|\tilde{\epsilon}(u)\, e_i|^{-2}\}} = |\tilde{\epsilon}(u)\, e_i|^{-2} \exp\{-c_2|\tilde{\epsilon}(u)\, e_i|^{-2}\}$$

(where $c_2 = c_1 - c_0 > 0$) is bounded, so that (4·8) is equivalent to

$$\int_0^\infty \exp\{-c|\tilde{\epsilon}(u)\, e_i|^{-2}\}\, du < \infty \quad \text{for all} \quad c > 0, \quad i = 1, \ldots, n.$$

As was explained in Chan and Williams [1], this last condition is equivalent to

$$|\tilde{\epsilon}(u)\, e_i|^2 \log u \to 0, \quad \text{for each} \quad i \leqslant n.$$

Hence

$$(\max_{1 \leqslant i \leqslant n} |\tilde{\epsilon}(u)\, e_i|)^2 \log u \to 0. \tag{4·9}$$

But $(\max_i |\tilde{\epsilon}(u)\, e_i|)$ is a norm on $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^d)$, and so (4·9) is equivalent to (1·2). ∎

5. *Some concluding remarks*

There is no reason why we should confine ourselves to considering only deterministic diffusion coefficients $\tilde{\epsilon}$. If we allow $\tilde{\epsilon}$ to depend on $\tilde{Y}_t$ as well as on t, then, provided that $\tilde{\epsilon}$ satisfies some additional 'good-behaviour' condition (for example if $\tilde{\epsilon}(t, \cdot)$ is a bounded function), a similar condition to (1·2) can clearly be derived with very few modifications to the proof just given.

The author thanks David Williams and James Norris, and he thanks Aubrey Truman for pointing out some references in the literature.

REFERENCES

[1] T. CHAN and D. WILLIAMS. An excursion approach to an annealing problem. *Math. Proc. Cambridge Philos. Soc.* **105** (1989), 169–176.

[2] T.-S. CHIANG, C.-R. HWANG and S.-J. SHEU. Diffusion for global optimization in $\mathbb{R}^n$. *SIAM J. Control Optim.* **25** (1987), 737–753.

[3] B. HAJEK. A tutorial survey of theory and applications of simulated annealing. (Lecture presented at the IEEE Conference on Decision and Control, December 1985.)

[4] B. HAJEK. Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** (1988), 311–329.

https://doi.org/10.1017/S0305004100001511 Published online by Cambridge University Press