The Annals of Applied Probability
1997, Vol. 7, No. 4, 1118–1139

CONVERGENCE RATES FOR ANNEALING DIFFUSION PROCESSES[1]

BY DAVID MÁRQUEZ

Universidad Central de Venezuéla and Université de Marne-La-Vallée

We consider the annealing diffusion process and investigate convergence rates. Namely, for the diffusion $dX_t = -\nabla V(X_t)\,dt + \sigma(t)\,dB_t$, where $(B_t)_{t\geq 0}$ is the d-dimensional Brownian motion and $\sigma(t)$ decreases to zero, we prove a large deviation principle for $(V(X_t))$ and weak convergence of $(\sigma^{-2}(t)(V(X_t) - \inf V))$.

1. Introduction. Let V be a real-valued function defined on $\mathbb{R}^d$. Following the idea of simulated annealing to search for the global minima of V, several papers [3, 10, 11, 12, 14, 18, 20, 22] have considered the annealing diffusion process defined by

$$dX_t = -\nabla V(X_t)\,dt + \sigma(t)\,dB_t, \tag{1}$$

where X_0 is independent of the d-dimensional Brownian motion (B_t) and where $\frac{1}{2}\sigma^2(t)$ is the annealing rate (or *temperature*) which decreases to zero if $t \to \infty$. Under suitable conditions on V and $\sigma(\cdot)$, these works proved the convergence in probability of (X_t) to the set

$$\operatorname{Argmin} V \equiv \{x \in \mathbb{R}^d\colon V(x) = \inf V\} \tag{2}$$

with $\inf V = \inf_{y\in\mathbb{R}^d} V(y)$, and the weak convergence of (X_t) to some probability on $\operatorname{Argmin} V$. In this work we consider the annealing diffusion processes on $\mathbb{R}^d$ and obtain the following results on large deviations from the global minima and weak convergence rates, whose precise statements will be given in Section 1.2.

Large deviations. For $r > 0$ small enough, if $E[V(X_0)] < \infty$,

$$\lim_{t\to\infty} \sigma^2(t) \ln P(V(X_t) \geq \inf V + r) = -2r. \tag{3}$$

Weak convergence. Under some regularity conditions on V in a neighborhood of $\operatorname{Argmin} V$, $4\sigma^{-2}(t)[V(X_t) - \inf V]$ converges weakly to a chi-square random variable.

Throughout this work we consider a function V satisfying the following assumptions.

A1 (Assumptions about the function V). $V\colon \mathbb{R}^d \to \mathbb{R}$ is twice continuously

Received May 1995; revised May 1997

[1]Research supported in part by INTEVEP S. A. Research and technical support center, affiliated to PDVSA Venezuéla Oil Company.

AMS 1991 *subject classifications.* 62L20, 60F05, 60F10, 60J60.
Key words and phrases. Simulated annealing, diffusion, large deviations, weak convergence.

differentiable and:

(i) V tends to ∞ as $\|x\| \to \infty$;
(ii) $\|\nabla V\|^2 - \Delta V$ is bounded from below;
(iii) $\{\nabla V = 0\}$ has a finite number of connected components;
(iv) for positive constants A and B, $V \le A\|\nabla V\|^2 + B$.

REMARK 1. Part (iv) of assumption A1 is not necessary for the preliminary results stated below (where $\lim V(x) = \infty$ as $\|x\| \to \infty$ would be sufficient). Its introduction in [16] was given in order to prove a logarithmic Sobolev inequality; not surprisingly, it will also be helpful to obtain convergence rates (namely, for Proposition 1).

Symbols. We adopt the following symbols throughout the rest of the paper: The symbol $\otimes$ denotes the product between measures. We will use ∇V and ΔV, respectively, to denote the gradient and Laplacian of the potential V on $\mathbb{R}^d$. $\|\cdot\|_{\mathrm{var}}$ is the total variation of a measure and $\asymp$ denotes asymptotic equivalence. The symbol $\mathscr{K}$ stands for a generic strictly positive constant, whose value might change during a proof. Specifying the initial state $X_0 = x$, X_t will sometimes be denoted X_t^x.

Before precisely stating our results, we need some preliminaries, given in Section 1.1. Then, in Section 1.2 we shall state our theorems. The proofs are given in Sections 2 and 3.

1.1. *Previous results.*

1.1.1. *Large deviation principle for Gibbs distribution.* Under assumptions A1, for any *temperature* $\tau > 0$, the normalization constant

$$c(1/\tau) = \int \exp(-V(x)/\tau)\,dx \tag{4}$$

is finite (see [15], page 347). Let (G_τ) be the Gibbs distribution with density $[c(1/\tau)]^{-1}\exp(-V/\tau)$ with respect to the Lebesgue measure. From Bryc's inverse Varadhan lemma (see [6]), it follows that the family (G_τ) satisfies a large deviation principle with rate function $V - \inf V$ when $\tau \to 0$. Consequently, for any $r > 0$,

$$\lim_{\tau \to 0} \tau \ln G_\tau(V - \inf V \ge r) = -r. \tag{5}$$

1.1.2. *About the homogeneous gradient diffusion process.* The homogeneous diffusion process defined by (1) taking $\sigma(\cdot) = \sigma$ for a constant $\sigma > 0$ has been extensively studied. It is a recurrent process and its stationary distribution is $G_{\sigma^2/2}$. The infinitesimal generator

$$L_\sigma(\cdot) = [\sigma^2/2]\Delta(\cdot) - \nabla V \cdot \nabla(\cdot) \tag{6}$$

has a self-adjoint negative extension on $L^2(G_{V,\sigma^2/2})$ with discrete spectrum $0 = \lambda_1^\sigma \ge \lambda_2^\sigma \ge \cdots \ge \lambda_n^\sigma \ge \cdots$. Large deviation principles for this diffusion for small σ are linked to a constant Λ described in several works [14, 15]. Jacquot [15] gives the following characterization of Λ. For two points x, y in $\mathbb{R}^d$, we denote by Γ_{xy} the set of C^1 parametric curves joining x and y. Then,

$$(7)\qquad \Lambda = 2 \sup_{x,y\in\mathbb{R}^d} \inf_{\gamma\in\Gamma_{x,y}} \left\{ \sup_t \{V(\gamma(t))\} - V(x) - V(y) + \inf V \right\}.$$

REMARK 2. It is clear that $\Lambda \geq 0$. When $\operatorname{Argmin} V$ has several connected components, $\Lambda > 0$.

Based on large deviation properties [1, 9, 24], the following results are stated in [14].

Large deviations for the spectral gap.

$$(8)\qquad \lim_{\sigma\to 0} \sigma^2 \ln(-\lambda_2^\sigma) = -\Lambda.$$

Convergence rate to the stationary distribution. For all $\lambda > \Lambda$, there exists a $\beta > 0$ and a $C > 0$, such that for all compact sets K and $\sigma \leq \sigma_0$ small enough, taking $T(\sigma) = \exp(\lambda\sigma^{-2})$,

$$(9)\qquad \sup_{x\in K} \left\| P(X^x_{T(\sigma)} \in \cdot) - G_{\sigma^2/2} \right\|_{\mathrm{var}} \leq \exp(-\beta\sigma^{-2}),$$

$$(10)\qquad \sup_{x\in K} P\left(\sup_{t\leq T(\sigma)} \|X_t^x\| \geq C \right) \leq \exp(-\beta\sigma^{-2}).$$

1.1.3. *Simulated annealing.* Taking the annealing schedule $\sigma^2(t) = c/\ln t$, with $c > \Lambda$, for the critical constant described above, the annealing diffusion process (X_t^x) ruled by (1) satisfies,

$$\left\| P(X_t^x \in \cdot) - G_{\sigma^2(t)/2} \right\|_{\mathrm{var}} \to 0.$$

This is proved in [3] and [14] for $c > 3\Lambda/2$ and in [22] for $c > \Lambda$. See also [20]. Following [14], there also exists a constant $\Lambda_1 \leq \Lambda$ such that, for $c \in (\Lambda_1, \Lambda]$, the weaker statement applies: the distance of (X_t^x) to $\operatorname{Argmin} V$ converges to zero in probability.

1.1.4. *Time change.* (a) For most of our proofs, it will be convenient to consider the time-changed diffusion process defined as follows: For $a = \sigma^{-2}$ set $A(t) = \int_0^t \sigma^2(s)\,ds$ and $B_t^1 = \int_0^{A^{-1}(t)} \sigma(s)\,dB_s$. Then (B_t^1) is a Brownian motion and $Y_t = X_{A^{-1}(t)}$ is a diffusion with slowly increasing drift, driven by

$$(11)\qquad dY_t = -a(t)\nabla V(Y_t)\,dt + dB_t^1.$$

All results claimed for the annealing diffusion process have their translation for the time changed diffusion processes.

(b) Taking $\sigma(\cdot) = \sigma$, if (P_t^σ) is the transition semigroup associated with the homogeneous gradient diffusion processes (1), then the semigroup associated with (Y_t) is (Π_t^a) where $\Pi_t^a(x, dy) = P_{t/a}^{a^{-1/2}}(x, dy)$.

1.2. *Statement of our results.* We make the following assumptions on the annealing schedule $\sigma(\cdot)$.

A2 (Assumptions on the "small annealing schedule" $\sigma(\cdot)$). We have $\sigma^2(t) = c/\ln(t+a)$ with $c > \Lambda$ and $\sigma^2(0) < 2$. Or, more generally, we have the following:

(i) $\sigma^2(t) \geq c/\ln t$ with $c > \Lambda$ for t large enough and $\sup_t \sigma^2(t) < 2$;

(ii) if $t \to \infty$, σ is a regular and slowly varying function decreasing to zero from $[0, \infty]$ to $(0, \infty)$, that is, $\sigma(t)$ decreases to zero and $\sigma(tx)/\sigma(t)$ tends to 1 for all $x > 0$;

(iii) $\sigma(t\sigma^2(t))/\sigma(t) \to 1$;

(iv) for large t, $\sigma^2(t)$ is continuously differentiable and convex.

REMARK 3. As pointed out in [14], Λ is the best constant in the sense that the weak convergence of the annealing process (X_t) to some probability concentrating on $\operatorname{Argmin} V$ fails to hold if $c < \Lambda$. For the convergence rates investigated in this work, we do not consider the case $c \in (\Lambda_1, \Lambda]$ mentioned in Section 1.1.3.

REMARK 4. For slow variation, see [8], pages 268–276. For $a = \sigma^{-2}$ and the function A defined in 1.1.4, we have, if $t \to \infty$,

$$A(t) \asymp \frac{a(t)}{t}, \qquad a'(t) = o\left(\frac{a(t)}{t}\right).$$

Hence, according to part (iii), $a(A(t)) \asymp a(t)$.

1.2.1. *Large deviation principles.*

PROPOSITION 1. *Under assumptions* A1 *and* A2, *the annealing diffusion process has the following tightness property: there exist real constants* R, $C > 0$ *such that, for any* $x \in \mathbb{R}^d$,

$$E\left[V(X_t^x)\mathbf{1}_{\{V(X_t^x) > R\}}\right] \leq C[1 + V(x)]\sigma^2(t). \tag{12}$$

THEOREM 1 (Large deviations). *For any compact set* Γ, *and for* $r > 0$ *small enough*,

$$\lim_{t \to \infty} \sup_{x \in \Gamma} \sigma^2(t) \ln P(V(X_t^x) \geq \inf V + r) = -2r. \tag{13}$$

Moreover, for $E[V(X_0)] < \infty$ *and* $r > 0$ *small enough*,

$$\lim_{t \to \infty} \sigma^2(t) \ln P(V(X_t) \geq \inf V + r) = -2r. \tag{14}$$

1.2.2. *Weak convergence rates.* Following Hwang [13], some regularity assumptions on V in a neighborhood of $\operatorname{Argmin} V$ ensure the weak convergence of the Gibbs distribution G_τ when $\tau \to 0$ to a probability G_0 concentrating on $\operatorname{Argmin} V$. Let us consider the three frameworks analyzed by Hwang.

A3 (Complementary assumptions on V).

A3.1. The set $\operatorname{Argmin} V$ has a strictly positive Lebesgue measure.

Then G_0 is the uniform distribution on $\operatorname{Argmin} V$.

A3.2. The function V is three times continuously differentiable and, for all $z \in \operatorname{Argmin} V$, $D^2V(z)$ is positive definite.

Then, as $V(x)$ tends to ∞ as $\|x\| \to \infty$, $\operatorname{Argmin} V$ is a finite set and, for all $y \in \operatorname{Argmin} V$,

$$G_0(y) = \left[\det D^2V(y)\right]^{-1/2}\left(\sum_{z \in \operatorname{Argmin} V} \left[\det D^2V(z)\right]^{-1/2}\right)^{-1}.$$

A3.3. The function V is three times continuously differentiable, $\operatorname{Argmin} V$ has a finite number of connected compact components and each component is a smooth manifold. Furthermore, for all points of any of these manifold with the highest dimension, the "second order partial differential of V with respect to smooth normal coordinates" is invertible.

Then G_0 concentrates on the highest dimensional components. We refer to [13] for a precise statement of assumptions A3.3 based on smooth local coordinates of V.

We now state some convergence rates under regularity assumptions linked to those of Hwang, as we will describe in Remark 5.

THEOREM 2. *Assuming that, on a neighborhood of* $\operatorname{Argmin} V$, $\|\nabla V\|^2 \geq \mathscr{C}(V - \inf V)$, *we have, for any compact set* Γ,

$$\sup_t \sup_{x \in \Gamma} \sigma^{-2}(t) E\left[V(X_t^x) - \inf V\right] < \infty.$$

Moreover, if $E[V(X_0)] < \infty$,

$$\sup_t \sigma^{-2}(t) E\left[V(X_t) - \inf V\right] < \infty.$$

We denote $\Rightarrow$ for "converges weakly."

THEOREM 3. *We assume that the function* $a \mapsto c(a) = \int_{\mathbb{R}^d} \exp(-aV(x))\,dx$ *varies regularly with exponent* $(-\delta/2)$, $\delta \in \mathbb{N}$. *Then, if* $E[V(X_0)] < \infty$,

$$4\sigma^{-2}(t)\left[V(X_t) - \inf V\right] \Rightarrow \chi^2(\delta),$$

where $\chi^2(0)$ *is the Dirac measure on* 0 *and, for* $\delta > 0$, $\chi^2(\delta)$ *is a chi-square random variable with* δ *degrees of freedom.*

REMARK 5. (a) Under assumption A3.1, the function $c(a)$ tends, as $a \to \infty$, to the Lebesgue measure of the set $\operatorname{Argmin} V$. Thus, theorem 3 holds with $\delta = 0$.

(b) Under assumption A3.2 in a neighborhood of $z_i \in \operatorname{Argmin} V$, $V(x) \asymp \mathscr{C}\|x - z_i\|^2$. Thus Theorem 2 and Theorem 3 apply with $\delta = d$.

(c) Let ν be the highest dimension of the regular components. Following the proof of Theorem 3.1 of [13], it is easy to check that, under assumption A3.3, Theorem 2 and Theorem 3 apply with $\delta = d - \nu$.

THEOREM 4. *Under assumption* A3.2, *if* $E[V(X_0)] < \infty$, *then, when* $t \to \infty$, *we have*

$$4\sigma^{-2}(t)[V(X_t) - \inf V] \Rightarrow \chi^2(\delta), \tag{15}$$

$$(X_t, \sigma^{-1}(t)\nabla V(X_t)) \Rightarrow \sum_{z \in \operatorname{Argmin} V} G_0(z) \otimes N(0, D^2V(z)), \tag{16}$$

denoting by $N(0, D^2V(z))$ *the Gaussian distribution with covariance* $D^2V(z)$.

Furthermore, the family of processes $(X_{t+u}, \sigma^{-1}(t+u)\nabla V(X_{t+u}))_{t>0}$ *converges weakly to* $Z^{(\infty)} = (Z^{(\infty,1)}, Z^{(\infty,2)})$, *where* $Z_0^{(\infty)}$ *has the distribution*

$$\sum_{z \in \operatorname{Argmin} V} G_0(z) \otimes N(0, D^2V(z)), \tag{17}$$

with $Z_t^{(\infty,1)} = Z_0^{(\infty,1)}$ *for all t and*

$$dZ_t^{(\infty,2)} = -D^2V(Z_0^{(\infty,1)})Z_t^{(\infty,2)} + D^2V(Z_0^{(\infty,1)})\,dB_t, \tag{18}$$

where $(B_t)_{t\leq 0}$ *is a Brownian motion independent of* $(Z_0^{(\infty)})$.

1.2.3. *Comments.* Convergence rates for simulated annealing on discrete spaces have been widely studied, mostly with large deviation methods. A large deviation principle similar to our Theorem 1 is proved for annealing diffusions on compact Riemannian manifolds by [11]. In the same framework, [4] gives bounds for the density of (X_t) with respect to $G_{\sigma^2(t)/2}$. As the best function $\sigma(\cdot)$ available for global optimization (as soon as Argmin V has several connected components) is $\sigma(t) = (c/\ln t)^{1/2}$ for $c > \Lambda$, such rates of weak convergence are, of course, disappointing for a practical simulated annealing purpose. However, they might be helpful in better understanding the mathematical structure of such nonstationary diffusions; further studies should focus on accelerating this optimization process.

A companion paper of Pelletier [21] investigates similar rates of convergence of discrete time annealing algorithms on $\mathbb{R}^d$.

2. Proof of large deviation principles. We first prove Proposition 1 in Section 2.1. In Section 2.2 we state some upper bounds for Chiang, Hwang and Sheu's proof, which leads in Section 2.3 to the proof of Theorem 1 when $\sigma^2(t) \geq c/\ln t$, $c > 3\Lambda/2$. In Section 2.4 we consider the general case $c > \Lambda$ and conclude the proof of Theorem 1, precising upper bounds in Royer's proof.

2.1. *Proof of Proposition* 1. *Step* 1. Let us first prove that

$$E[V(X_t^x)] \leq V(x) + \mathscr{C}t.$$

By Itô's formula and part (ii) of assumptions A1,

$$\begin{aligned} dV(X_t^x) &= -\|\nabla V(X_t^x)\|^2\,dt + (\sigma^2(t)/2)\Delta V(X_t^x) + \sigma(t)\langle \nabla V(X_t), dB_t\rangle \\ &\leq \mathscr{C}\,dt + \sigma(t)\langle \nabla V(X_t), dB_t\rangle, \end{aligned}$$

and the assertion follows.

STEP 2. By parts (ii) and (iv) of assumptions A1 there exist two constants $r > 0$ $D_1 > 0$, such that, for $V \geq r$,

$$V + \Delta V \leq D_1 \|\nabla V\|^2.$$

Let ϕ be an increasing C^2 function from $\mathbb{R}$ to $[0, 1]$, equal to 0 on $(-\infty, r]$ and equal to 1 on $[R, \infty)$. Since $\nabla(\phi \circ V) = (\phi' \circ V)\nabla V$ and $\Delta(\phi \circ V) = (\phi'' \circ V)\|\nabla V\|^2 + (\phi' \circ V)\Delta V$ are continuous functions with compact support, they are bounded.

Set $\Psi = (\phi \circ V)V$. Then, a short computation shows that

$$\nabla \Psi = (\phi' \circ V) V \nabla V + (\phi \circ V)\nabla V$$

and

$$\begin{aligned}\Delta \Psi &= (\phi \circ V)\Delta V + (\phi' \circ V)(2\|\nabla V\|^2 + V\Delta V) + (\phi'' \circ V) V \|\nabla V\|^2 \\ &= O\big((\phi \circ V)\|\nabla V\|^2 + 1\big).\end{aligned}$$

By Itô's formula,

$$\begin{aligned}d(\Psi(X_t^x)) = &-\phi(V(X_t^x))\|\nabla V(X_t^x)\|^2\, dt \\ &- V(X_t^x)\phi'(V(X_t^x))\|\nabla V(X_t^x)\|^2\, dt \\ &+ \tfrac{1}{2}\sigma^2(t)\Delta\Psi(X_t^x)\, dt + \sigma(t)\langle \nabla\Psi(X_t^x), dB_t\rangle.\end{aligned}$$

Set $\alpha(t) = E[\phi(V(X_t^x))V(X_t^x)]$; then, for $t \geq t_0$, t_0 large enough, $s > 0$,

$$\alpha(t+s) - \alpha(t) \leq -D\int_t^{t+s} \alpha(u)\, du + \int_t^{t+s} \sigma^2(u)\, du,$$

and, by Gronwall's lemma,

$$\alpha(t) \leq \sigma^2(t)\alpha(t_0).$$

Finally, combining the inequalities above we get, for all x,

$$E\big[V(X_t^x)\mathbf{1}_{\{V(X_t^x) \geq R\}}\big] \leq \mathscr{C}[1 + V(x)]\sigma^2(t),$$

which proves the proposition. □

2.2. *Upper bounds in Chiang, Hwang and Sheu's proof.* We return to the proof given in [3]. The basic idea is to consider, for any $\tau > 0$, the shifted annealing diffusion $X^{(\tau)} = (X_{t+\tau})$, and to check how long it "follows" the homogeneous diffusion with the same starting point and the constant schedule $\sigma \equiv \sigma(\tau)$. Until that time $\alpha(\tau)$, properties of the homogeneous diffusion can be transferred to the annealing diffusion. As the basic tool is Girsanov's theorem, it will be easier to handle the time-changed diffusions described in Section 1.1.4.

2.2.1. *Diffusion processes with an increasing drift.* Let h be a bounded and Lipschitz function from $\mathbb{R}^d$ to $\mathbb{R}^d$. For a and $\hat{a}$, nondecreasing right-continuous $\mathbb{R}^+$-valued functions with $a(0) = \hat{a}(0)$, let us consider the diffusion processes

$$(19)\qquad \begin{aligned} dY_t &= a(t)h(Y_t)\,dt + dB_t,\\ dZ_t &= \hat{a}(t)h(Z_t)\,dt + dB_t. \end{aligned}$$

We denote as always Y_t^x and Z_t^x if $Y_0 = Z_0 = x$.

(a) How does (Z_t^x) follow (Y_t^x) and how long? According to Girsanov's theorem and following [3] (see Pages 743–744), we set

$$K_t = \int_0^t [a(s) - \hat{a}(s)]^2\,ds.$$

Then for any bounded Borel function ϕ from $\mathbb{R}^d$ and $\mathbb{R}$ and any $A > 0$,

$$(20)\qquad |E[\phi(Y_t^x)] - E[\phi(Z_t^x)]| \le C_1\|\phi\|K_t^{1/2}\exp(C_2K_t),$$

$$(21)\qquad P\Big(\sup_{s\le t}\|Y_s^x\| \ge A\Big) - P\Big(\sup_{s\le t}\|Z_s^x\| \ge A\Big) \le C_1K_t^{1/2}\exp(C_2K_t);$$

C_1 and C_2 are two positive constants depending only on $\|h\| = \sup\|h(\cdot)\|$.

(b) Application to shifted diffusion processes. For $\tau > 0$, $B_t^{(\tau)} = (B_{t+\tau} - B_\tau)_{t\ge 0}$, the above result holds for the shifted diffusion processes ruled by

$$(22)\qquad \begin{aligned} dY_t^{(\tau)} &= a(t+\tau)h\big(Y_t^{(\tau)}\big)\,dt + B_t^{(\tau)},\\ dZ_t^{(\tau)} &= a(\tau)h\big(Z_t^{(\tau)}\big)\,dt + B_t^{(\tau)} \end{aligned}$$

with K_t replaced by

$$\begin{aligned} K_t^{(\tau)} &= \int_0^t [a(s+\tau) - a(\tau)]^2\,ds\\ &\le (a'(\tau))^2\frac{t^3}{3}\\ &= o\left(\left(\frac{a(\tau)}{\tau}\right)^2\frac{t^3}{3}\right), \qquad \tau\to\infty. \end{aligned}$$

For a $\mathbb{R}^+ =$ valued function $\alpha(\tau)$, increasing to infinity, such that $r^2(\tau) = K_{\alpha(\tau)}^{(\tau)}$ tends to 0 as $\tau\to\infty$, inequalities (20) and (21) can be written as

$$(23)\qquad \begin{aligned} &\big|E\big[\phi(Y_{\alpha(\tau)+\tau})/Y_\tau = x\big] - E\big[\phi(Z_{\alpha(\tau)+\tau})/Z_\tau = x\big]\big|\\ &\quad\le C_3\|\phi\|r^2(\tau), \end{aligned}$$

$$(24)\qquad \begin{aligned} &P\Big(\sup_{s\le\alpha(\tau)}\|Y_{s+\tau}\| \ge 2\|x\| + A/Y_\tau = x\Big)\\ &\quad\le P\Big(\sup_{s\le\alpha(\tau)}\|Z_{s+\tau}\| \ge AZ_\tau = x\Big) + C_3r^2(\tau), \end{aligned}$$

C_3 only depending on $\|h\|$.

2.2.2. *Accompanying the homogeneous gradient diffusion process.* Let us study the diffusion

$$dY_t = -a(t)\nabla V(Y_t)\,dt + dB_t. \tag{25}$$

For a real continuous function ϕ with compact support, taking R large enough in Proposition 1, there exists a compact set K containing the support of ϕ and the set $\{V \le R\}$, such that

$$P(Y_t^x \notin K) \le \mathscr{C}(1 + V(x))1/a(t). \tag{26}$$

Let us consider, for each $u \ge 0$, a homogeneous diffusion process driven by

$$dZ_t^{(u,x)} = -a(u)\nabla V\big(Z_t^{(u,x)}\big)\,dt + B_t^{(u)}, \tag{27}$$

with $Z_0^{(u,x)} = x$, $x \in K$. Then, the inequalities (9) and (10) of 1.1.2 can be translated as follows: for $\lambda > \Lambda$ and $u \ge u_0$ large enough, if $\alpha(u) = a(u)\exp(\lambda a(u))$, there exists a constant $\beta > 0$ such that

$$\sup_{x \in K} \|P\big(Z_{\alpha(u)}^{(u,x)} \in \cdot\big) - G_{1/2a(u)}\|_{\mathrm{var}} \le \exp(-\beta a(u)), \tag{28}$$

and there exists a constant A such that $K \subset \{\|X\| \le A\}$ and

$$\sup_{x \in K} P\Big(\sup_{t \le \alpha(u)} \|Z_t^{(u,x)}\| \ge A\Big) \le \exp(-\beta a(u)). \tag{29}$$

Unfortunately the function ∇V is not bounded. Therefore we cannot directly apply the results obtained in Section 2.2.1. However, as we shall see, for a fairly large amount of time, the diffusion remains bounded on an event of probability increasing to 1.

For $A_1 = A + 2\sup_{x \in K}\|x\|$, let $\tilde{V}$ be a twice continuously differentiable function, such that the following holds:

1. the restriction of $\tilde{V}$ to the ball of center 0 and radius A_1 is V;
2. $\tilde{V}$ and its first- and second-order derivatives are bounded.

Then, 2.2.1(b) applies to compare $\tilde{Y}^{(u)}$ and $\tilde{Z}^{(u)}$ governed by

$$\begin{aligned} d\tilde{Y}_t^{(u)} &= -a(t+u)\nabla\tilde{V}\big(\tilde{Y}_t^{(u)}\big)\,dt + B_t^{(u)},\\ d\tilde{Z}_t^{(u)} &= -a(u)\nabla\tilde{V}\big(\tilde{Z}_t^{(u)}\big)\,dt + B_t^{(u)} \end{aligned} \tag{30}$$

with $\tilde{Y}_0^{(u)} = \tilde{Z}_0^{(u)} = x$, $x \in K$. For the function r defined in 2.2.1(b) we have

$$r(u) = o\big([\alpha(u)]^{3/2}a(u)/u\big) = o\big([a(u)]^{5/2}u^{-1}\exp(3\lambda a(u)/2)\big).$$

As we assumed that $a(u) \le \ln u/c$, it follows that

$$r(u) = o\big([\ln u]^{5/2}u^{-1+3\lambda/2c}\big). \tag{31}$$

2.3. *Proof of Theorem* 1 *when* $c > 3\Lambda/2$. We take the constant $\lambda > \Lambda$ such that $c > 3\lambda/2$. By (31), we have

$$r(u) \le \mathscr{C}\exp(-sa(u))$$

for $0 < s < 1 - 3\lambda/2c$. Thus, we have by 2.2.1(b),

$$P\Big(\sup_{s \le \alpha(u)} \|\tilde{Y}_s^{(u)}\| \ge A_1\Big) \le P\Big(\sup_{s \le \alpha(u)} \|\tilde{Z}_s^{(u)}\| \ge A\Big) + \mathscr{C}\exp(-sa(u))$$

with a constant $\mathscr{C}$ independent of $x \in K$.

Hence, if $x \in K$,

$$P\Big(\sup_{s \le \alpha(u)} \|Y_s^{(u)} - \tilde{Y}_s^{(u)}\| > 0 / Y_u = \tilde{Y}_0^{(u)} = x\Big) \le \mathscr{C}\exp(-sa(u)),$$

$$P\Big(\sup_{s \le \alpha(u)} \|Z_s^{(u)} - \tilde{Z}_s^{(u)}\| > 0 / Z_u = \tilde{Z}_0^{(u)} = x\Big) \le \mathscr{C}\exp(-sa(u)).$$

Moreover, we have, by 2.2.1(b),

$$\Big|E\Big[\phi\big(\tilde{Z}_{\alpha(u)}^{(u)}\big)/Z_0^{(u)} = x\Big] - E\Big[\phi\big(\tilde{Y}_{\alpha(u)}^{(u)}\big)/Y_0^{(u)} = x\Big]\Big|$$
$$\le \mathscr{C}\|\phi\|\exp(-\rho a(u)),$$

with $\rho = \inf(s, \beta)$. Thus, thanks to inequality (9) applied to $Z^{(u)}$, we get, uniformly if $x \in K$,

$$\begin{aligned} &\Big|E\Big[\phi\big(\tilde{Y}_{\alpha(u)}^{(u)}\big)/\tilde{Y}_0^{(u)} = x\Big] - G_{1/2a(u)}(\phi)\Big| \le \mathscr{C}\|\phi\|\exp(-\rho a(u)),\\ &\Big|E\big[\phi(Y_{\alpha(u)+u})/Y_u = x\big] - G_{1/2a(u)}(\phi)\Big| \le \mathscr{C}\|\phi\|\exp(-\rho a(u)). \end{aligned} \tag{32}$$

Hence, by Proposition 1,

$$\Big|E\big[\phi(Y_{\alpha(u)+u}^x)\big] - G_{1/2a(u)}(\phi)\Big|$$
$$\le \mathscr{C}\|\phi\|(\exp(-\rho a(u)) + (1 + V(x))1/a(u)).$$

As $a(u + \alpha(u)) \asymp a(u)$, we also have

$$\begin{aligned} &\Big|E\big[\phi(Y_u^x)\big] - G_{1/2a(u)}(\phi)\Big|\\ &\quad\le \mathscr{C}\|\phi\|(\exp(-\rho a(u)) + (1 + V(x))1/a(u)). \end{aligned} \tag{33}$$

The constant $\mathscr{C}$ is uniform over all continuous functions whose supports are included in the same compact. For $0 < r_1 < r < R$, R large enough such that Proposition 1 holds, let us take the function ϕ in the last result such that

$$\mathbf{1}_{\{r + \inf V < V < R\}} \le \phi \le \mathbf{1}_{\{r_1 + \inf V \le V < 2R\}}.$$

Then, from (33), we obtain

$$P(V(Y_u^x) > \inf V + r) \le E\big[\phi(Y_u^x)\big] + \mathscr{C}(1 + V(x))1/a(u)$$

and

$$P(V(Y_u^x) \ge \inf V + r_1) \ge E\big[\phi(Y_u^x)\big],$$

which implies

$$\begin{aligned} P(V(Y_u^x) \ge \inf V + r) \le{}& G_{1/2a(u)}(V \ge r_1 + \inf V)\\ &+ \mathscr{C}(\exp(-\rho a(u)) + (1 + V(x))1/a(u)),\\ P(V(Y_u^x) \ge \inf V + r_1) \ge{}& G_{1/2a(u)}(V \ge r_1 + \inf V)\\ &- \mathscr{C}(\exp(-\rho a(u)) + (1 + V(x))1/a(u)). \end{aligned}$$

Hence, by the large deviation principle in Section 1.1.1, for $2r < \rho$,

$$\lim_{t\to\infty} \sup_{x\in\Gamma} \frac{1}{a(t)} \ln P(V(Y_t^x) \geq \inf V + r) = -2r.$$

By part (iii) of assumptions A2, $a(A(t))/a(t) \to 1$, hence (13) follows, taking $X_t^x = Y_{A(t)}^x$, with $A(t) = \int_0^t a(s)\,ds$. For $E[V(X_0)] < \infty$ we find (14) by the same arguments. □

REMARK 6. For $\Lambda < c \leq 3\Lambda/2$, the previous proof tells us that, until the time $\hat{\alpha}(u) = O(u^{\nu})$, $\nu < 2/3$, the annealing diffusion follows the homogeneous diffusion, with $r(u) = o(u^{\delta/c})$. Unfortunately, this time is not sufficiently long to guarantee the convergence to the Gibbs measure. However, we still have

$$\sup_{x\in K} P\Big(\sup_{t\leq\hat{\alpha}(u)} \|Y_t^{(u)}\| \geq A_1 / Y_u = x\Big) \leq \mathscr{C} \exp(-\rho a(u)). \tag{34}$$

The aim of the next section is to prove Theorem 1 for $c > \Lambda$.

2.4. *Proof of Theorem* 1 *for* $c > \Lambda$.

2.4.1. *Accompanying the stepwise diffusion process.* For $0 < \alpha < 1/3$, define $t_n = \sum_{k=1}^n k^{-\alpha} \asymp (1/(1-\alpha))n^{1-\alpha}$, and a stepwise function $\hat{a}(t) = a(t_n)$ on the interval $[t_n, t_{n-1})$. We return to the framework of Section 2.2.1.

How does $(\hat{Y}_t)$ follow (Y_t) and how long? Let A be a function from $\mathbb{N}$ to $\mathbb{N}$ increasing to infinity with $A(N) = o(N)$. Let us consider the shifted diffusion processes $(\hat{Y}_{t+t_N})$ and (Y_{t+t_N}). For $t_{N-1} \leq u < t_N$, we set $\hat{\alpha}(u) = t_{N+A(N)} - t_{N-1}$, then

$$\begin{aligned}
K_{\hat{\alpha}(u)} &= \int_{t_N}^{t_{N+A(N)}} [a(s) - \hat{a}(s)]^2\,ds \\
&= o\Bigg(\sum_{j=N}^{N+A(N)} \big[a'(t_j)\big]^2 (t_{j+1} - t_j)^3\Bigg) \\
&= o\Bigg(\sum_{j=N}^{N+A(N)} j^{-3\alpha}[\ln j]^2 j^{-2+2\alpha}\Bigg) \\
&= o\Big((\ln N)^2\big[N^{-1-\alpha} - (N + A(N))^{-1-\alpha}\big]\Big) \\
&= o\Big((\ln N)^2 N^{-2-\alpha} A(N)\Big).
\end{aligned}$$

If we take $A(N) = N^{\tau}$ with $\alpha < \tau < \alpha + 2(1-\alpha)/3$, then

$$\begin{aligned}
\hat{\alpha}(u) &\asymp \frac{1}{1-\alpha}\big\{(N + N^{\tau})^{1-\alpha} - N^{1-\alpha}\big\} \\
&\asymp N^{\tau-\alpha} \\
&= O\big(u^{(\tau-\alpha)/(1-\alpha)}\big),
\end{aligned}$$

with $(\tau-\alpha)/(1-\alpha) < 2/3$. Hence, $K_{\hat{\alpha}(u)} = o((\ln N)^2 N^{-2-\alpha+\alpha+2(1-\alpha)/3})$.

Then, if we take up the notations of Section 2.3, we obtain by (34),

$$\sup_{x\in K} P\Big(\sup_{u\le t\le u+\hat{\alpha}(u)} \|Y_t\| > A_1/Y_u = x\Big) \le \mathscr{C}\exp(-\rho a(u)). \tag{35}$$

Thus, an easy adaptation of the proof in Section 2.3 gives, for $A_2 = A_1 + 2\sup_{x\in K}\|x\|$,

$$P\Big(\sup_{u\le t\le u+\hat{\alpha}(u)} \|\hat{Y}_t\| > A_2/\hat{Y}_u = x\Big) \le \mathscr{C}\exp(-\rho a(u)), \tag{36}$$

$$\begin{aligned}&\Big|E\big[\phi(Y_{u+\hat{\alpha}(u)})/Y_u = x\big] - E\Big[\phi\big(\hat{Y}_{u+\hat{\alpha}(u)}\big)/\hat{Y}_u = x\Big]\Big|\\ &\qquad\le \mathscr{C}\|\phi\|\exp(-\rho a(u)).\end{aligned} \tag{37}$$

The constant $\mathscr{C}$ depends only on the support of ϕ. In Section 2.4.3 we shall prove that, uniformly for $x \in K$, for $t_{N-1} \le u < t_N$,

$$\Big|E\Big[\phi\big(\hat{Y}_{u+\hat{\alpha}(u)}\big)/\hat{Y}_u = x\Big] - G_{1/2a(t_N+A(N))}(\phi)\Big| \le \mathscr{C}\|\phi\|\exp(-\rho a(u)). \tag{38}$$

Then the end of the proof in Section 2.3 remains valid and we again obtain (32) with α replaced by $\hat{\alpha}$, thus (33) and Theorem 1 hold with $c > \Lambda$.

2.4.2. *Convergence rates for stepwise annealing diffusion process.* We now consider the diffusion process ruled by

$$d\hat{Y}_t = -\hat{a}(t)\nabla V\big(\hat{Y}_t\big)\,dt + dB_t. \tag{39}$$

Let us state Royer's results [22], precising some upper bounds. In this section we assume that $V(x) = \|x\|^4$ for large $\|x\|$ (super normal case).

(a) Hypercontractivity in "the supernormal case." Based on Log-Sobolev inequalities for the Schrödinger operators [2], [5], Royer [22] obtains the following hypercontractive estimates for the transition semigroup (P_t^{σ}) of the homogeneous gradient diffusion process (1).

For all $\delta > 0$ if $t = \delta\ln(2)$, then, for $f \in L^2(G_{\sigma^2/2})$,

$$\|P_t^{\sigma}(f)\|_{L^4(G_{\sigma^2/2})} \le e^M\|f\|_{L^2(G_{\sigma^2/2})} \tag{40}$$

with

$$2M = \mathscr{C}(1+\sigma^{-2}) - \frac{d\ln\delta}{4} + \frac{\sigma^2}{4\delta^2}. \tag{41}$$

The above result could be written for the transition of the time-changed diffusion (Y_t), $\Pi_t^a(x,dy) = P_{t/a}^{a^{-1/2}}(x,dy)$, as

$$\|\Pi_t^a(t)\|_{L^4(G_{1/2a})} \le e^M\|f\|_{L^2(G_{1/2a})} \tag{42}$$

with

$$(43) \qquad 2M = \mathscr{C}(1+a) - \frac{d}{4}\ln\left(\frac{t}{a}\right) - \ln(\ln 2) + \frac{(\ln 2)^2}{4t^2}a.$$

(b) Hypercontractivity for stepwise transition semigroup. The bound (42) gives a constant M_n such that

$$(44) \qquad \|\Pi^{a(t_n)}_{t_{n+1}-t_n}(f)\|_{L^4(G_{1/2a(t_n)})} \le e^{M_n}\|f\|_{L^2(G_{1/2a(t_n)})}$$

and

$$(45) \qquad 2M_n \le \mathscr{C} n^{2\alpha}\ln n.$$

(c) Spectral gap. By (8), for $\Lambda < \lambda < c$ and $n \ge N$, N large enough, it follows from result (8) in Section 1.1.2, for $f \in L^2(G_{1/2a(t_n)})$, $\int f\,dG_{1/2a(t_n)} = 0$,

$$\begin{aligned} &\|\Pi^{a(t_n)}_{t_{n+1}-t_n}(f)\|_{L^2(G_{1/2a(t_n)})} \\ &\quad\le \exp\left(-\frac{(t_{n+1}-t_n)}{a(t_n)}\exp(+\lambda a(t_n))\right)\|f\|_{L^2(G1/2a(t_n))} \\ &\quad\le \exp\left(-\frac{cn^{-\alpha}}{(1-\alpha)\ln n}\exp\left(-\frac{\lambda}{c}(1-\alpha)\ln n\right)\right)\|f\|_{L^2(G_{1/2a(t_n)})} \end{aligned}$$

by the asymptotic behavior of t_n. Hence,

$$\begin{aligned} 1 - r_n &= \sup\left\{\frac{\|\Pi^{a(t_n)}_{t_{n+1}-t_n}(f)\|_{L^2(G_{1/2a(t_n)})}}{\|f\|_{L^2(G1/2a(t_n))}}; \int f\,dG_{1/2a(t_n)} = 0\right\} \\ &\le \exp\left(-\frac{cn^{-\alpha}}{(1-\alpha)\ln n}\exp\left(-\frac{\lambda}{c}(1-\alpha)\ln n\right)\right) \\ &= \exp\left(\frac{-cn^{-\alpha-(\lambda/c)(1-\alpha)}}{(1-\alpha)\ln n}\right). \end{aligned}$$

(d) Variations of Gibbs distributions. Set $\nu_\beta = c_\beta \exp(-\beta V)$ the density of G_β. On $\{V(x) \ge n^{3\alpha}\}$, the following inequalities are proved in [22], for n large enough:

$$\nu_{2a(t_{n+1})}(x) \le \nu_{2a(t_n)}(x)$$

and

$$\nu_{2a(t_n)}(x) - \nu_{2a(t_{n+1})}(x) \le \phi_n(x)\nu_{2a(t_{n+1})}(x)$$

for a function ϕ_n defined by

$$\phi_n(x) = \left(\nu_{2a(t_n)}(x)/\nu_{2a(t_{n+1})}(x) - 1\right)\mathbf{1}_{\{V(x)\ge n^{3\alpha}\}}$$

satisfying

$$\ln\left(\int \phi_n^2(x)\nu_{2a(t_n)}(x)\,dx\right)^{1/2} \asymp -a(t_n)n^{3\alpha}.$$

Hence, for all $\alpha < 1$,

$$\ln\left(\frac{e^{2M_n}\|\phi_n\|_{L^2(\nu_{2a(t_n)})}}{r_n}\right) \asymp -a(t_n)n^{3\alpha}.$$

On the other hand, for $V(x) \le n^{3\alpha}$ and for n large enough, [22] obtains

$$|\nu_{2a(t_{n-1})}(x) - \nu_{2a(t_n)}(x)| \le \gamma_n \nu_{2a(t_n)}(x)$$

with

$$\gamma_n = n^{3\alpha}(a(t_{n+1}) - a(t_n)) = o\left(\frac{n^{2\alpha}\ln t_n}{t_n}\right) = o(n^{-1+3\alpha}\ln n).$$

(e) Convergence rates for the stepwise annealing diffusion. Following the proof of Lemma 2.1 in [22], we pointed out that, for $n \ge N$, if $\hat{Y}_{t_n}$ admits a density g_N with respect to $G_{1/2a(t_N)}$, then, for $n \ge N$, $\hat{Y}_{t_n}$ admits also a density g_n with respect to $G_{1/2a(t_n)}$ and

$$y_n = \int [1 - g_n]^2\, dG_{1/2a(t_n)}$$

satisfies

$$y_{n+1} \le a_n y_n + b_n,$$

where

$$a_n(1-\gamma_n) = (1-r_n)^2 + e^{2M_n}\|\phi_n\|_{L^2(G_{1/2a(t_n)})}$$

and

$$b_n(1-\gamma_n) = \gamma_n + e^{2M_n}\|\phi_n\|_{L^2(G_{1/2a(t_n)})}.$$

Then, for $0 < k < 2$, $n \ge N$, N large enough, we have $r_n < 1/2$ and $R_n = r_N + \cdots + r_n$,

$$a_n \le 1 - kr_n \quad \text{and} \quad b_n \le r_n R_n^{-s}$$

with $s > 0$. Hence,

$$y_{n+1} \le (1 - kr_n)y_n + r_n R_n^{-s}.$$

Thus (see, e.g., [7], Lemma 4.I.1) we obtain

$$y_n \le \sup\{\exp(-kR_n)y_N, \mathscr{C}R_n^{-s}\},$$

that is,

$$\|P_{\hat{Y}_{t_n}} - G_{1/2a(t_n)}\|_{\mathrm{var}}^2 \le \sup\{\exp(-kR_n)y_N, \mathscr{C}R_n^{-s}\}. \tag{46}$$

2.4.3. *Proof of formula* (39).

STEP 1. By (36), it is enough to prove, for $t_{N-1} \le u < t_N$ and $x \in K$,

$$\begin{aligned}\left|E\left[\phi\left(\hat{Y}_{u+\hat{\alpha}(u)}\right)\mathbf{1}_{[\sup_{u<s\le u+\hat{\alpha}(u)}\|\hat{Y}_s\|\le A_2]}/\hat{Y}_u = x\right] - G_{1/2a(t_{N+A(N)})}(\phi)\right| \\ \le \mathscr{C}\exp(-\rho a(u)).\end{aligned} \tag{47}$$

For any Borel set Γ of $\mathbb{R}^d$, set

$$F_N(x,\Gamma) = P\left(\left[\hat{Y}_{t_N} \in \Gamma\right] \cap \left[\sup_{u \le s \le t_N} \|\hat{Y}_s\| \le A_2\right] / \hat{Y}_u = x\right).$$

Define $\tilde{V}$ as in Section 2.3, with bounded derivatives of order 1 or 2 and with $V(y) = \hat{V}(y)$ for $\|y\| \le A_2$. Set $\check{Y}_0 = x$, and for $t > 0$,

$$d\check{Y}_t = -\hat{a}(u)\nabla\tilde{V}(\check{Y}_t)\,dt + dB_t.$$

Applying Lemma 1.1 of [23] (see also [17]), the transition density $(\check{p}_t^u)$ of $(\check{Y}_t)$ satisfies the inequality

$$\check{p}_t^u(x,y) \le (2\pi t)^{-d/2} \exp\left(\left[-\|x-y\|^2/2t + \mathscr{C}\, t\hat{a}(u)\right]\right).$$

Thus, $F_N(x,\cdot)$ has a density $f_N(x,\cdot)$ with respect to the Lebesgue measure which satisfies

$$f_N(x,y) \le \mathscr{C} N^{\alpha d/2} \mathbf{1}_{\{\|y\| \le A_2\}};$$

and, with respect to $G_{1/2a(t_N)}$, the normalized distribution $F_N(x,\cdot)/F_N(x,\mathbb{R}^d)$ has a density $g_N(x,\cdot)$ satisfying

$$g_N(x,y) \le \mathscr{C} N^{\alpha d/2} \mathbf{1}_{\{\|y\| \le A_2\}} \exp\left(2\hat{a}(u) \sup_{\|y\| \le A_2} V(y)\right) = \mathscr{C} N^{\beta},$$

with β positive constant.

Thus, in order to prove (38), it is sufficient to prove that, if Y_{t_N} has a distribution with a density respect to $G_{1/2a(t_N)}$, bounded above by $\mathscr{C} N^{\beta}$,

$$\tag{48} \begin{aligned} &\left|E\left[\phi\left(\hat{Y}_{t_{N+A(N)}}\right)\mathbf{1}_{\{\sup_{t_N \le s \le t_{N+A(N)}} \|\hat{Y}_s\| \le A_2\}}\right] - G_{1/2a(t_{N+A(N)})}(\phi)\right| \\ &\quad \le \mathscr{C} \exp(-\rho a(t_N)). \end{aligned}$$

STEP 2. In order to prove (48), we may modify V and take $V(y) = \|y\|^4$ for $\|y\| \ge 2A_2$. Thus, applying again the inequality (36), it is sufficient to prove, in the supernormal case,

$$\tag{49} \left\| E\left[\phi\left(\hat{Y}_{t_{N+A(N)}}\right)\right] - G_{1/2a(t_{N+A(N)})}(\phi)\right\| \le \mathscr{C} \exp(-\rho a(t_N)).$$

In Section 2.4.1 we have taken $A(N) = N^{\tau}$ with $\alpha < \tau < \alpha + 2(1-\alpha)/3$. Then for any $\delta < \alpha + \lambda(1-\alpha)/c$ and N large enough, we get by (46) for $R_{N+A(N)} = r_N + \cdots + r_{N+A(N)}$,

$$\begin{aligned} R_{N+A(N)} &\ge \sum_{j=N}^{N+A(N)} j^{\delta} \\ &\ge \mathscr{C}\left((N+A(N))^{1-\delta} - N^{1-\delta}\right) \\ &\ge \mathscr{C} N^{\tau-\delta} \quad \text{if } \tau < \delta. \end{aligned}$$

Thus, taking $\alpha < \tau < \alpha + (1-\alpha)\inf[2/3, \lambda/c]$,

$$\left|E\left[\phi\left(\hat{Y}_{t_{N+A(N)}}\right)\right] - G_{1/2a(t_{N+A(N)})}(\phi)\right|^2$$
$$\leq \mathscr{C} \sup\left\{N^{2\beta}\exp\left(-kR_{N+a(N)}\right), R_{N+a(N)}^{-s}\right\}$$

and (49) is proved. This completes the proof of Theorem 1. □

3. Proof of weak convergence rates. For simplicity, let us set $\inf V = 0$ throughout this section.

3.1. *Proof of Theorem* 2.

STEP 1. Large deviation principles for the shifted diffusion process.

It follows from the assumptions A2 that

$$\lim_{u\to\infty} \sup_{t\leq T} \left|\int_0^t \sigma^2(u+s)\,ds - t\sigma^2(u)\right| = 0.$$

Hence, Freidlin and Wentzell's results concerning the homogeneous diffusion process with small perturbations can be transcribed to the family of diffusion processes $X^{(u)} = (X_{u+t})_{0\leq t\leq T}$ (see [19, 24] for more details).

STEP 2. Let W be a neighborhood of $\operatorname{Argmin} V$ where $\|\nabla V\|^2 \leq \mathscr{C}V$; for $r > 0$ small enough, $\{V < 2r\} \subseteq W$ and $\{V < 2r\}$ is a region of attraction for the ordinary differential equation $\dot{z}(t) = -\nabla V(z(t))$. Thus, by Step 1, for any $T > 0$,

$$\limsup_{u\to\infty} \sigma^2(u) \ln P\left(\sup_{t\leq T} V(X_{t+u}) > 2r/V(X_u) \leq r\right) < 0.$$

STEP 3. Applying Theorem 1 and Proposition 1 with $r > 0$ defined in Step 2 and $R > r$, we get, for u large enough, $0 \leq t \leq T$ and constant $\rho > 0$,

$$\begin{aligned}E\left[V(X_{t+u}^x)\right] &\leq E\left[V(X_{t+u}^x)\mathbf{1}_{\{V(X_{t+u}^x)\leq r\}}\right] + \exp(-\rho/\sigma^2(u))\\ &\leq rP(V(X_u^x) \geq r) + E\left[V(X_{t+u}^x)\mathbf{1}_{\{V(X_u^x)\leq r\}}\right]\\ &\quad + \exp(-\rho/\sigma^2(u)).\end{aligned}$$

By Itô' formula and part (ii) of assumption A1, we have for all u,

$$\begin{aligned}&E\left[V(X_{t+u}^x)\mathbf{1}_{\{V(X_u^x)\leq r\}}\right]\\ &\quad\leq rP(V(X_u^x) \leq r)\\ &\qquad - \int_0^t E\left[\left(\|\nabla V(X_{s+u}^x)\|^2 + \tfrac{1}{2}\Delta V(X_{s+u}^x)\sigma^2(s+u)\right)\mathbf{1}_{\{V(X_u^x)\leq r\}}\right] ds\\ &\qquad + \mathscr{C}\int_0^t \sigma^2(s+u)\,ds\\ &\quad\leq E\left[V(X_u^x)\right] + \exp(-\rho/\sigma^2(u)) + \mathscr{C}\left[\int_0^t \sigma^2(s+u)\,ds\right]\\ &\qquad - \mathscr{C}\left[\int_0^t E\left[V(X_{s+u}^x)\mathbf{1}_{\{V(X_u^x)\leq r\}}\right] ds\right].\end{aligned}$$

Hence

$$E\big[V(X_{t+u}^x)\big] \le E\big[V(X_u^x)\big] - \mathscr{C}\int_0^t E\big[V(X_{s+u}^x)\big]\,ds + \mathscr{C}\sigma^2(u).$$

By Gronwall's inequality we obtain

$$E\big[V(X_{T+u}^x)\big] \le \big(E\big[V(X_u^x)\big] + \mathscr{C}\sigma^2(u)\big)\exp(-\mathscr{C}T)$$

and, for k large enough,

$$\begin{aligned} E\big[V(X_{kT}^x)\big] &\le V(x)\exp(-k\mathscr{C}T) + \mathscr{C}\left(\sum_{i=1}^{k}\exp(-i\mathscr{C}T)\sigma^2((k-i)T)\right) \\ &\le V(x)\exp(-k\mathscr{C}T) + \mathscr{C}\exp(-(k/2)\mathscr{C}T) + \sigma^2((K/2)T) \\ &\le \mathscr{C}\sigma^2(kT), \end{aligned}$$

and, for $t = kT + h$, $h < k$ and k large enough,

$$\begin{aligned} E\big[V(X_t^x)\big] &\le E\big[V(X_{kT}^x)\big] + \sigma^2(kT) - \mathscr{C}\int_0^h E\big[V(X_{kT+s})\big]\,ds \\ &\le \mathscr{C}\sigma^2(kt) \le \mathscr{C}\sigma^2(t). \end{aligned}$$

This completes the proof of Theorem 2. □

3.2. *Proof of Theorem* 3.

STEP 1. For $c(a) = \int_{\mathbb{R}^d}\exp(-aV(x))\,dx$, with $a = 1/\tau$, and U_a a random variable with distribution $G_{1/2a}$, the Laplace transform of $4aV(U_a)$ is the function $\lambda \mapsto c(2a(2\lambda+1))/c(2a)$, which converges, as $a \to \infty$, to the function defined by $(2\lambda+1)^{-\delta/2}$, that is, to the Laplace transform of the distribution $\chi^2(\delta)$. Hence, for any $r > 0$,

$$G_{1/2a}(4aV \ge r) \to \chi^2(\delta)((r,\infty)).$$

STEP 2. For any $r > 0$, let us take $R > r$ satisfying Proposition 1 and $r_1 \in (0, r)$. For any $t > 0$, we apply (33) to a continuous function ϕ_t, such that

$$\mathbf{1}_{\{r/4a(t) \le V < R\}} \le \phi_t \le \mathbf{1}_{\{r_1/4a(t) \le V < 2R\}}.$$

Then

$$P(4a(t)V(Y_t^x) \ge r) \le E\big[\phi_t(Y_t^x)\big] + \mathscr{C}(1 + V(x))1/a(t)$$

and

$$P(4a(t)V(Y_t^x) \ge r_1) \ge E\big[\phi_t(Y_t^x)\big],$$

which implies, for any $r_1 < r$,

$$\begin{aligned} P(4a(t)V(Y_t^x) \ge r) &\le G_{1/2a(t)}(4a(t)V \ge r_1) \\ &\qquad + \mathscr{C}(\exp(-\rho a(t)) + (1 + V(x))1/a(t)), \\ P(4a(t)V(Y_t^x) \ge r_1) &\ge G_{1/2a(t)}(4a(t)V \ge r) \\ &\qquad - \mathscr{C}(\exp(-\rho a(t)) + (1 + V(x))1/a(t)), \end{aligned}$$

and Theorem 3 follows from Step 1. □

3.3. *Proof of Theorem* 4. Set $Z = \{Z_t\}_{t>0} = (X_t, W_t)$ where $W_t = \sigma^{-1}(t)\nabla V(X_t)$ and, for $u \geq 0$, let us define the family of shifted diffusion processes, $Z_t^{(u)} = Z_{t+u} = (X_t^{(u)}, W_t^{(u)})$.

STEP 1. For all function φ: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ continuous and bounded with compact support, we have to prove, denoting $\nu^z = N(0, (1/2)D^2V(z))$,

$$E[\varphi(X_t, W_t)] \to \sum_{z \in \operatorname{Argmin} V} G_0(z) \int \varphi(z, y)\, d\nu^z(y)$$

or, equivalently,

$$E\Big[\varphi\big(Y_t, [a(t)]^{1/2}\nabla V(Y_t)\big)\Big] \to \sum_{z \in \operatorname{Argmin} V} G_0(z) \int \varphi(z, y)\, d\nu^z(y).$$

Thanks to formula (34), this is equivalent to showing

$$E\Big[\varphi\big(Y_{t+\hat{\alpha}(t)}, [a(t+\hat{\alpha}(t))]^{1/2}\nabla V(Y_{t+\hat{\alpha}(t)})\big)\Big]$$
$$\to \sum_{z \in \operatorname{Argmin} V} G_0(z) \int \varphi(z, y)\, d\nu^z(y)$$

or

$$E\Big[\varphi\Big(Y_{t+\hat{\alpha}(t)}, [a(t+\hat{\alpha}(t))]^{1/2}\nabla V\big(Y_{t+\hat{\alpha}(t)}\big)\mathbf{1}_{[Y_t \in K, \sup_{t \le s \le t+\hat{\alpha}(t)}\|Y_s\| \le A_1]}\Big)\Big]$$
$$\to \sum_{z \in \operatorname{Argmin} V} G_0(z) \int \varphi(z, y)\, d\nu^z(y).$$

Consequently, in the proof of the above formula, we may modify V for large $\|x\|$ and, for technical reasons, we shall, in addition to assumption A1, assume that

$$\|\nabla V\|^2 \leq \mathscr{C} V$$

outside a suitable compact as Γ and that the partial derivatives of order 2 and 3 of V are bounded. Thus,

$$E\big[\sigma^{-2}(t)\|\nabla V(X_t)\|^2\big] \leq \mathscr{C}\sigma^{-2}(t)E\big[V(X_t)\big] + \mathscr{C}\sigma^{-2}(t)P(X_t \notin \Gamma)$$

with the right-hand side bounded, thanks to Theorem 2 and Proposition 1.

STEP 2 (Tightness). By Proposition 1, $(Z_0^{(u)})$ is tight. Hence, from Itô's formula,

$$d\big(\sigma^{-1}(t)\nabla V(X_t)\big) = -\sigma'(t)\sigma^{-2}(t)\nabla V(X_t)\, dt - \sigma^{-1}(t)D^2V(X_t)\nabla V(X_t)\, dt$$
$$+ \frac{\sigma(t)}{2}\Delta[\nabla V](X_t)\, dt + D^2V(X_t)\, dB_t,$$

where $\Delta[\nabla V]$ denotes the vector of Laplacians of the components of ∇V. Thus $Z = (X, W)$ is a solution of the stochastic differential system

$$\begin{aligned} dX_t &= -\nabla V(X_t)\,dt + \sigma(t)\,dB_t, \\ dW_t &= -D^2V(X_t)W_t\,dt + R_t\,dt + D^2V(X_t)\,dB_t, \end{aligned} \tag{50}$$

where

$$R_t = \frac{\sigma(t)}{2}\Delta[\nabla V](X_t) - \sigma'(t)\sigma^{-1}(t)B_t.$$

By Kolmogorov's inequality, we get

$$E\left[\sup_{t\le T}\left\|\int_u^{t+u}\sigma(r)\,dB_r\right\|^2\right] \le \int_u^{T+u}\sigma^2(r)\,dr \asymp 2T\sigma^2(u).$$

Hence, the family of processes $\{\int_u^{t+u}\sigma(r)\,dB_r\}_{t\ge 0}$ converges weakly to the process identical to zero. On the other hand, by Step 1,

$$E\left[\left\|\int_s^t \nabla V(X_{r+u})\,dr\right\|^2\right] \le (t-s)\int_{s+u}^{t+u} E\left[\|\nabla V(X_r)\|^2\right]dr \le \mathscr{C}(t-s)^2;$$

therefore, the family of processes $(X^{(u)})$ is tight.

Similarly, by Step 1,

$$\begin{aligned} E\left[\left\|\int_{s+u}^{t+u} D^2V(X_r)W_r\,dr\right\|^2\right] &\le (t-s)\int_{s+u}^{t+u} E\left[\|D^2V(X_r)W_r\|^2\right]dr \\ &\le \mathscr{C}(t-s)^2, \end{aligned}$$

and the family of processes $\{\int_u^{u-t} D^2V(X_s)W_s\,ds\}_{t\ge 0}$ is also tight.

Furthermore, the family of processes $\{\int_0^t R_{s+u}\,ds\}_{t\ge 0}$ converges weakly to the process identical to zero when $u\to\infty$. This results from

$$\begin{aligned} E\left[\sup_{t\le T}\left\|\int_u^{t+u} R_s\,ds\right\|\right] &\le \int_u^{T+u}\frac{\sigma(t)}{2}E[\|\Delta[\nabla V](X_t)\|]\,dt \\ &\quad + \int_u^{T+u}\sigma'(t)\sigma^{-2}(t)E[\|\nabla V(X_t)\|]\,dt \\ &\le \mathscr{C}\int_u^{T+u}\sigma(t)\,dt + \mathscr{C}\left(\frac{1}{\sigma(u+T)} - \frac{1}{\sigma(u)}\right) \end{aligned}$$

with the last expression decreasing to zero as $u\to\infty$.

Finally, by Burkholder's inequality,

$$\begin{aligned} E\left[\left\|\int_{s+u}^{t+u} D^2V(X_r)\,dB_r\right\|^4\right] &\le \mathscr{C}E\left[\left(\int_{s+u}^{t+u}\|D^2V(X_r)\|^2\,dr\right)^2\right] \\ &\le \mathscr{C}(t-s)\int_{s+u}^{t+u}E\left[\|D^2V(X_r)\|^4\right]dr \le \mathscr{C}(t-s)^2; \end{aligned}$$

thus, the family of processes $\{\int_u^{u+t} D^2V(X_s)\,dBs\}_{t\ge 0}$ is tight.

Hence the families of processes $(X^{(u)})$, $(W^{(u)})$ and $(Z^{(u)})$ are tight.

STEP 3 (Convergence). As we proved the convergence to the process identical to zero of $(\int_u^{t+u} \sigma(r)\, dB_r)_{t>0}$ and $(\int_0^t R_{s+u}\, ds)_{t>0}$ when $u \to \infty$, any closure point of $(Z^{(u)})$ is a solution of the stochastic differential system

$$
\begin{aligned}
dZ_t^{(\infty,1)} &= -\nabla V\big(Z_t^{(\infty,1)}\big)\, dt, \\
dZ_t^{(\infty,2)} &= -D^2 V\big(Z_t^{(\infty,1)}\big) Z_t^{(\infty,2)} + D^2 V\big(Z_t^{(\infty,1)}\big)\, dB_t.
\end{aligned}
\tag{51}
$$

Since (X_u) converges weakly to G_0, $Z_0^{(\infty,1)}$ has the distribution G_0. Moreover, the first equation is an ordinary differential equation whose initial value is a stable point for the gradient, hence $Z_t^{(\infty,1)} = Z_0^{(\infty,1)}$ for all t and

$$dZ_t^{(\infty,2)} = -D^2 V\big(Z_0^{(\infty,1)}\big) Z_t^{(\infty,2)} + D^2 V\big(Z_0^{(\infty,1)}\big)\, dB_t,$$

where $\{B_t\}_{t\geq 0}$ is a Brownian motion independent of $(Z_0^{(\infty,1)}, Z_0^{(\infty,2)})$.

For $H = D^2 V(Z_0^{(\infty,1)})$, we have

$$Z_t^{(\infty,2)} = \exp(-Ht)\left(Z_0^{(\infty,2)} + \int_0^t \exp(Hs)\, dB_s\right).$$

Thus, given a function ϕ Lipschitz and bounded, we obtain

$$
\begin{aligned}
&\bigg| E\big[\varphi\big(Z_t^{(\infty,1)}, Z_t^{(\infty,2)}\big)\big] \\
&\quad - \int G_0(dz)\, \varphi\left(z, \exp(-D^2 V(z) t) \int_0^t \exp(-D^2 V(z) s)\, dB_s\right)\bigg| \\
&\quad \leq \|\phi\| \exp(-\bar{\lambda} t),
\end{aligned}
$$

with $\bar{\lambda} = \inf_{z \in \operatorname{Argmin} V} \lambda_{\min} D^2 V(z)$.

Let μ be a probability on $\mathbb{R}^{2d}$, closure point of (X_u, W_u) for the weak convergence. The first marginal law of μ is G_0 and, by Step 1, the second marginal law has a second-order moment bounded above by $\sup_u E[\|W_u\|^2] < \infty$.

Let us consider now a sequence $\{u(n)\}_{n\geq 0}$ increasing to infinity such that $(X_{u(n)}, W_{u(n)})$ converges weakly to μ. By the tightness of the process $(X^{(u)}, W^{(u)})$, for all $t > 0$, there exists a subsequence of $\{u(n)\}_{n\geq 0}$, denoted by $\{v(n)\}_{n\geq 0}$, such that $(X^{(v(n)-t)}, W^{(v(n)-t)})$ converges weakly to the process $Z^{(\infty)}$, solution of (51).

Hence, for $\nu^z = N(0, (1/2) D^2 V(z))$,

$$
\begin{aligned}
&\bigg| E\big[\varphi\big(Z_t^{(\infty,1)}, Z_t^{(\infty,2)}\big)\big] - \sum_{z \in \operatorname{Argmin} V} G_0(z) \int \varphi(z, y)\, d\nu^z(y) \bigg| \\
&\quad \leq \mathscr{C} \|\phi\| \exp(-\bar{\lambda} t),
\end{aligned}
$$

and for all $t > 0$,

$$\left| \int \phi\, d\mu - \int \varphi(z, y)\, d\nu^z(y) \right| \leq \mathscr{C} \|\phi\| \exp(-\bar{\lambda} t),$$

thus

$$\mu = \sum_{z \in \operatorname{Argmin} V} G_0(z)\delta_z \otimes \nu^z$$

and Theorem 4 is established. □

Acknowledgments. I specially thank Professor Marie Duflo for her constant guidance and encouragement. I am also grateful to the referees and editors for a careful reading of the manuscript and several very useful suggestions.

REFERENCES

[1] AZENCOTT, R. (1980). *Grandes Déviations et Applications. Lecture Notes in Math.* **744**. Springer, Berlin.

[2] CARMONA, R. (1979). Regularity properties of Schrödinger and Dirchlet semigroups. *J. Funct. Anal.* **33** 259–296.

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

[4] CONCORDET, D. (1994). Estimation de la densité du recuit simulé. *Ann. Inst. H. Poincaré* **30** 265–302.

[5] DAVIES, E. B. and SIMON, B. (1984). Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *J. Funct. Anal.* **59** 335–395.

[6] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Barlett, Boston.

[7] DUFLO, M. (1996). Algorithmes stochastiques. *Math. Appl.* **23**. Springer, Paris.

[8] FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **2**, 3rd ed. Wiley, New York.

[9] FREIDLIN, W. and WENZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, Berlin.

[10] GEMAN, S. and HWANG, C. R. (1986). Diffusions for global optimization. *SIAM J. Control Optim.* **25** 1031–1043.

[11] HOLLEY, R., KUSUOKA, S. and STROOK, D. (1989). Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** 333–347.

[12] HOLLEY, R. and STROOK, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 533–569.

[13] HWANG, C. R. (1980). Laplace's method revisited. *Ann. Probab.* **8** 1177–1182.

[14] HWANG, C. R. and SHEU, S. J. (1990). Large time behavior of perturbed diffusion Markov processes with applications to second eigenvalue problem for Fokker–Plank operator and simulated annealing. *Acta Appl. Math.* **19** 253–295.

[15] JACQUOT, S. (1992). Comportement asymptotique de la seconde valeur propre des processus de Kolmogorov. *J. Multivariate Anal.* **40** 335–347.

[16] JACQUOT, S. and ROYER, G. A. (1995). Ergodicity of stochastic plates. *Probab. Theory Related Fields* **102** 19–44.

[17] KIFER, Y. (1974). Certain results concerning small random perturbations of dynamical system. *Theory Probab. Appl.* **3** 487–505.

[18] KUSHNER, H. J. (1987). Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via Monte Carlo. *SIAM J. Appl. Math.* **47** 169–185.

[19] LIPTSER, R. SH. and PUKHALSKII, A. A. (1992). Limit theorem on large derivations for semimartingales. *Stochastics Stochastics Rep.* **38** 201–249.

[20] MICLO, L. (1992). Recuit simulé sur $\mathbb{R}^n$. Etude de l'évolution de l'énergie libre. *Ann. Inst. H. Poincaré* **28** 235–266.

[21] PELLETIER, M. (1996). Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. Preprint.

[22] ROYER, G. A. (1989). A remark on simulated anealing of diffusion processes. *SIAM J. Control Optim.* **27** 1403–1408.

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

[24] WENTZELL, A. D. (1990). *Limit Theorems on Large Derivations for Markov Stochastic Processes.* Kluwer, Dordrecht.

EQUIPE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES
UNIVERSITÉ DE MARNE-LA-VALLÉE
2 RUE DE LA BUTTE VERTE
93166 NOISY-LE-GRAND CEDEX
FRANCE
E-MAIL: marquez@math.univ-mlv.fr

DEPARTAMENTO DE MATEMÁTICA
FACULTAD DE CIENCIAS
UNIVERSIDAD CENTRAL DE VENEZUÉLA
AV. LOS ILUSTRES. LOS CHAGUARAMOS
A.P. 20513 CARACAS 1020-A
VENEZUELA