The Annals of Applied Probability
1998, Vol. 8, No. 1, 10–44

WEAK CONVERGENCE RATES FOR STOCHASTIC APPROXIMATION WITH APPLICATION TO MULTIPLE TARGETS AND SIMULATED ANNEALING

BY MARIANE PELLETIER

University of Marne-la-Vallée

We study convergence rates of $\mathbb{R}^d$-valued algorithms, especially in the case of multiple targets and simulated annealing. We precise, for example, the convergence rate of simulated annealing algorithms, whose weak convergence to a distribution concentrated on the potential's minima had been established by Gelfand and Mitter or by Hwang and Sheu.

1. Introduction. Many well-known stochastic algorithms enable us to locate the zeros of a function h: $\mathbb{R}^d \to \mathbb{R}^d$ or the local (eventually global) minima of a function V: $\mathbb{R}^d \to \mathbb{R}$. Such algorithms may be written as

$$Z_{n+1} = Z_n + \gamma_n\big[h(Z_n) + \eta_{n+1}\big] + \sigma_n \xi_{n+1}, \tag{1.1}$$

and we take $h = -\nabla V$ in the search for the minima of V.

The scale factors or gains of the algorithm $(\gamma_n)_{n\geq 0}$ and $(\sigma_n)_{n\geq 0}$ are two strictly positive deterministic sequences, decreasing to zero, and are freely chosen. In many cases, $h(Z_n)$ is observable only up to a disturbance η_{n+1}, for instance, up to a Markovian disturbance [2] (if the function h is known, then $\eta_{n+1} = 0\ \forall n$). The random noise (ξ_n) might be a simulated sequence of independent identically distributed random vectors; such a noise prevents the algorithm from falling into useless "traps" such as saddle points or local maxima of V when $h = -\nabla V$.

Our contribution is related to the rate of weak convergence of (1.1). We first consider weakly disturbed algorithms in the presence of multiple targets, and then simulated annealing algorithms.

1.1. *Weak convergence rates in the presence of multiple targets.* When the gains (γ_n) and (σ_n) are chosen such that $v(n) = \gamma_n/\sigma_n^2$ is an increasing sequence with $v(n)/\ln n \to +\infty$, (1.1) is said to be a *weakly disturbed algorithm*.

Weakly disturbed algorithms and their relationship with the ordinary differential equation (ODE),

$$\frac{dz}{dt}(t) = h[z(t)], \tag{1.2}$$

have been widely studied, and we know many criteria that give almost sure convergence of (Z_n) to an attractive target z^*, that is a zero of h, which is asymptotically stable for (1.2). For such results, see [2], [6], [7] and [24].

Received May 1995; revised May 1997.
AMS 1991 *subject classifications*. 62L20, 60F05, 60H10.
Key words and phrases. Stochastic algorithms, simulated annealing, weak convergence.

Once the convergence of (1.1) has been established, it is interesting to study the rate of convergence of (Z_n). Several results on the rate of weak convergence have been proved, for instance, by Bouton [3], Kersting [21], Kushner-Huang [25] and Nevel'son and Ha'sminskiĭ [31], when it is known that the sequence (Z_n) converges with probability 1 to an attractive target of h. A more complete bibliography on that subject can be found in Walk's chapter of [28]. Kaniovski [19] and Kaniovski and Pflug [20] have considered the case of a finite number of attractive targets, under the assumption that the algorithm almost surely converges to one of them.

More recently, the study of high-dimensional algorithms has increased interest in almost sure asymptotic behaviors of algorithms, which are associated with more complicated ODE (see [1], [6] and [10]), and we know that some algorithms may have various limit sets (not necessarily reduced to a single point), which are linked to the ODE (1.2); this case is called the case of *multiple targets*. It has been seen by simulations that the probability the algorithm converges to an attractive target z^* is strictly positive, but the precise study of the link between the initial value given to the algorithm and the probability that (Z_n) converges to z^* remains an open problem.

Our first aim is to extend the previous results on the rate of weak convergence of (1.1) to the case of multiple targets. More precisely, z^* being an attractive target, our goal is to establish the weak convergence rate of (Z_n) towards z^* given the event $\Gamma(z^*) = \{\omega; Z_n(\omega) \to z^*\}$, and whatever the behavior of (Z_n) outside of $\Gamma(z^*)$ may be. We shall assume that the probability of $\Gamma(z^*)$ is strictly positive, but not necessarily equal to 1. Moreover our assumptions on (1.1) will be local assumptions [that is, assumptions required only once (Z_n) is sufficiently close to z^*]. Thus, our assumptions are less restrictive than those found in [3], [18], [19], [20], [24], [30] and our results can be applied, of course, to the case of multiple targets, but also to algorithms obtained by truncation or projection.

Roughly speaking, assuming smooth regularity properties of h in a neighborhood of z^* (and whatever h might be elsewhere), we prove that, given $\Gamma(z^*)$,

$$\sqrt{v(n)}(Z_n - z^*) \Rightarrow \mathscr{N}(0, \Sigma),$$

with $v(n) = \gamma_n \sigma_n^{-2}$, $\Rightarrow$ denoting the weak convergence, $\mathscr{N}$ the Gaussian distribution and Σ a covariance matrix.

1.2. *Weak convergence rates of simulated annealing algorithms on* $\mathbb{R}^d$. Stochastic algorithms used in the search for the minima of a function $V: \mathbb{R}^d \to \mathbb{R}$ are gradient algorithms of the form

$$Z_{n+1} = Z_n - \gamma_n[\nabla V(Z_n) + \eta_{n+1}] + \sigma_n \xi_{n+1}, \tag{1.3}$$

(ξ_n) being a sequence of independent simulated random vectors with distribution $\mathscr{N}(0, I)$.

The gains (γ_n) and (σ_n) being chosen in the same way as in 1.1, (1.3) is an example of a weakly disturbed algorithm and our previous result applies: z^*

being a local minimum of V, given the event $\Gamma(z^*) = \{Z_n \to z^*\}$,

$$\sqrt{v(n)}(Z_n - z^*) \Rightarrow \mathscr{N}(0, \Sigma).$$

This implies that, given $\Gamma(z^*)$,

$$4v(n)[V(Z_n) - V(z^*)] \\ \Rightarrow Y^T[D^2V(z^*) + \zeta I]^{-1/2} D^2V(z^*)[D^2V(z^*) + \zeta I]^{-1/2} Y,$$

where $v(n) = \gamma_n/\sigma_n^2$, Y is a d-dimensional random vector with distribution $\mathscr{N}(0, I)$, and ζ depends on the choice of the gain γ_n. [In most cases, the asymptotic distribution is the chi-square distribution with d degrees of freedom, denoted by $\chi^2(d)$].

However, it is often necessary to find the global minima of V, thus avoiding its local minima. The basic idea is to increase the simulated disturbance $(\sigma_n \xi_{n+1})$. As a matter of fact, we know that if the gains (γ_n) and (σ_n) are chosen such that $v(n) = \gamma_n \sigma_n^{-2}$ is an increasing sequence, $(v(n)/\ln n)$ being suitably bounded, then (Z_n) converges weakly to a distribution concentrated on the global minima of V. This is the framework studied by Gelfand and Mitter [11], [12]; in this case, (1.3) is no longer a weakly disturbed algorithm, but a simulated annealing algorithm.

Although convergence rates of simulated annealing type algorithms on finite spaces have been studied intensively, few results have been established on the rate of simulated annealing on $\mathbb{R}^d$. This is our second aim and we prove in particular that if Argmin V is a finite set and if V is a three times continuously differentiable function such that $D^2V(z^*)$ is invertible for any global minimum z^* of Argmin V, then

$$4v(n)[V(Z_n) - \inf V] \Rightarrow \chi^2(d),$$

with $v(n) = \gamma_n/\sigma_n^2$ and $\inf V = \inf_{z \in \mathbb{R}^d} V(z)$. Thus the rate of weak convergence of simulated annealing algorithms cannot be better than $1/v(n) = \mathscr{C}/\ln n$ (where $\mathscr{C}$ is a strictly positive constant), whereas the optimal rate of weak convergence of weakly disturbed algorithm is known to be $1/v(n) = \mathscr{C}/n$.

A companion paper from Marquez [29] investigates convergence rates for annealing diffusion processes.

2. Main results.

2.1. *Weak convergence rates under local assumptions.* We will look at the d-dimensional stochastic algorithm (1.1) defined on a probability space $(\Omega, \mathscr{A}, \mathscr{P})$ with a filtration $\mathscr{F} = (\mathscr{F}_n)_{n \geq 0}$.

Such a formula includes several types of weakly disturbed algorithms. Refering to regression problems, let us call the "noise" a sequence of increments of a d-dimensional square integrable martingale, adapted to $\mathscr{F}$, for instance a sequence of independent identically distributed random vectors. The Robbins–Monro algorithm is obtained with $\eta = 0$, (ξ_n) a noise and $\gamma_n = \sigma_n$. The Kiefer–Wolkowitz algorithm [22] introduces $h = -\nabla V$, (ξ_n)

being a noise and (η_n) a small residual disturbance. Finally, we may write η_{n+1} as $\eta_{n+1} = e_{n+1} + r_{n+1}$, where (e_n) is a noise, and the residual term (r_n) allows Markovian disturbances in the framework of [2]. In order to study unitarily all these situations, we set $\varepsilon_{n+1} = \xi_{n+1} + (\gamma_n/\sigma_n)e_{n+1}$ and rewrite (1.1) as

$$Z_{n+1} = Z_n + \gamma_n[h(Z_n) + r_{n+1}] + \sigma_n \varepsilon_{n+1}, \tag{2.1}$$

where (ε_n) is a noise and (r_n) a residual term.

(A1.1) Assumptions about the function $h\colon \mathbb{R}^d \to \mathbb{R}^d$.

(i) $h(z^*) = 0$.

(ii) On a neighborhood of z^*, $h(z) = \rho(z - z^*)H.(z - z^*) + O(\|z - z^*\|^2)$ with $\rho\colon \mathbb{R}^d \to [1, \overline{\rho}]$ $(\overline{\rho} > 1)$ Lipschitz on $\{z; \|z\| = 1\}$ and such that $\rho(tz) = \rho(z)$ for all $t > 0$.

(iii) H is a stable $d\times d$ matrix; that is, the largest real part of its eigenvalues is $(-L)$ with $L > 0$.

(A1.2) Assumptions about the disturbances (r_n) and (ε_n). For two constants $M > 0$ and $b > 2$, almost surely,

$$E(\varepsilon_{n+1}|\mathscr{F}_n)1_{\{\|Z_n - z^*\| \le M\}} = 0,$$

$$\sup_{n\ge 0} E\big(\|\varepsilon_{n+1}\|^b|\mathscr{F}_n\big)1_{\{\|Z_n - z^*\| \le M\}} < \infty,$$

$$E\big(v(n)\|r_{n+1}\|^2 1_{\{\|Z_n - z^*\| \le M\}}\big) \to 0.$$

Almost surely on $\Gamma(z^*) = \{Z_n \to z^*\}$, $E(\varepsilon_{n+1}\varepsilon_{n+1}^T|\mathscr{F}_n) \to \Gamma$, where Γ is a positive definite deterministic matrix.

(A1.3) Assumptions about the gains (γ_n) and (σ_n). Let γ and σ be two positive functions defined on $[0, +\infty[$, which decrease to zero. Let $\gamma_n = \gamma(n)$, $\sigma_n = \sigma(n)$.

Let $v = \gamma/\sigma^2$; we assume that v is a function increasing to infinity, differentiable and such that its differentiate v' varies regularly with exponent $\beta - 1 \ge -1$ (that is, for any $x > 0$, $v'(tx)/v'(t) \to x^{\beta-1}$ as $t \to +\infty$; cf. [9], [33]).

Moreover we assume either (A1.3.1) or (A1.3.2):

(A1.3.1) γ varies regularly with exponent $(-\alpha)$, $0 \le \alpha < 1$;

(A1.3.2) For $t \ge 1$, $\gamma(t) = \gamma_0/t$ with $2L\gamma_0 > \beta$.

COMMENTS ON THE ASSUMPTIONS. (a) Assumptions (A1.1) and (A1.2) are local. Given the convergence to z^*, they easily apply to the case of multiple targets as well as to projected or truncated algorithms (see, e.g., [4]).

(b) The introduction of the function ρ in (A1.1) allows directional derivatives, as in [20].

(c) Assumptions (A1.2) on the noise (ε_n) and the residual term (r_n) are satisfied in the most usual cases, for example, for Robbins–Monro and Kiefer–Wolfowitz algorithms, as well as for algorithms with Markovian disturbances.

(d) Assumption (A1.3) seems quite general, and Theorem 1 can be applied to the usual gains $\gamma_n = \sigma_n = \gamma_0 n^{-\alpha}$, $0 < \alpha \le 1$, as well as to slower gains considered in [7], for instance to $\gamma_n = \sigma_n = (c_n/\ln n)$, with $c_n \to 0$ and $v(n) = (\ln n)/c_n$ satisfying assumptions (A1.3).

(e) Since v' varies regularly with exponent $\beta - 1 \ge -1$, $nv'(n)/v(n) \to \beta$. We deduce that

$$\left[\frac{v(n)}{v(n-1)}\right]^{1/2} = 1 + \frac{\beta}{2n} + o\left(\frac{1}{n}\right)$$

and, in view of the definition of γ_n, under (A1.3.1), $[v(n)/v(n-1)]^{1/2} = 1 + o(\gamma_n)$; under (A1.3.2),

$$\left[\frac{v(n)}{v(n-1)}\right]^{1/2} = 1 + \frac{\beta}{2\gamma_0}\gamma_n + o(\gamma_n).$$

Set $\zeta = 0$ under (A1.3.1) and $\zeta = \beta/2\gamma_0$ under (A1.3.2). In both cases,

$$\left[\frac{v(n)}{v(n-1)}\right]^{1/2} = 1 + \zeta\gamma_n + o(\gamma_n). \tag{2.2}$$

THEOREM 1 (Convergence rate under local assumptions). *We assume* (A1.1) *to* (A1.3). *Given* $\Gamma(z^*)$,

$$\sqrt{v(n)}(Z_n - z^*) \Rightarrow \mu,$$

where μ is the stationary distribution of the diffusion

$$dX_t = [\rho(X_t)H + \zeta I]X_t\,dt + \Gamma^{1/2}\,dB_t \tag{2.3}$$

Moreover, this convergence is stable, that is, for any $\mathscr{F}_\infty$-measurable random variable ξ, $(\xi, \sqrt{v(n)}(Z_n - z^*))$ converges weakly, given $\Gamma(z^*)$, to the product measure of the distribution of ξ and of μ.

REMARKS. (a) Generally $\rho = 1$, and then $\mu = \mathscr{N}(0, \Sigma)$ where $\mathscr{N}$ is the Gaussian distribution and Σ is the solution of Lyapunov's equation

$$(H + \zeta I)\Sigma + \Sigma(H^T + \zeta I) = -\Gamma.$$

(b) The stability of this weak convergence is new in this framework; it takes its root from Feigin [8] or Touati [34].

2.2. *Convergence rate of simulated annealing algorithms.* Simulated annealing algorithms are used to find the global minima of a function $V: \mathbb{R}^d \to \mathbb{R}$.

We split the problem into two different parts, the gradient ∇V being known or observable only together with a disturbance.

2.2.1. *When the function ∇V is known.* Here we consider the gradient algorithm

$$Z_{n+1} = Z_n - \gamma_n \nabla V(Z_n) + \sigma_n \xi_{n+1} \tag{2.4}$$

under the following assumptions.

(A2.1) Assumptions about the function V. $V\colon \mathbb{R}^d \to \mathbb{R}$ is twice continuously differentiable and we have the following:

(i) $V(z) \to +\infty$ when $\|z\| \to +\infty$;
(ii) $\|\nabla V\|^2 - \Delta V$ is bounded from below;
(iii) $0 < \liminf \|\nabla V(z)\|^2/V(z) \le \limsup \|\nabla V(z)\|^2/V(z) < +\infty$ when $\|z\| \to +\infty$;
(iv) ∇V is globally Lipschitz;
(v) $\{\nabla V = 0\}$ has a finite number of connected components;
(vi) $E(V(Z_0)) < +\infty$.

(A2.2) Assumptions about the simulated noise. (ξ_n) is a simulated sequence of independent random vectors with normal distribution $\mathscr{N}(0, I)$, and independent of Z_0.

(A2.3) Assumptions about the gains (γ_n) and (σ_n). We add one of the following assumptions where Λ is a strictly positive constant, depending on V [see comment (b)].

(A2.3.1) (i) γ is decreasing and varies regularly with exponent $(-\alpha)$ where $\frac{1}{2} < \alpha < 1$.
(ii) v is increasing to infinity, continuously differentiable, concave, and varies slowly.
(iii) For t large enough, $v(t) \le [c \cdot (1-\alpha)] \ln t$ with $c < 1/\Lambda$.

(A2.3.2) For $t > 1$: (i) $\gamma(t) = \gamma_0 [t(\ln t)^m]^{-1/2}$ with $m \ge 2$.
(ii) $\sigma(t) = \sigma_0 [t^{1/4} (\ln t)^{(1/2)+(m/4)}]^{-1}$.
(iii) $\gamma_0/\sigma_0^2 < 1/\Lambda$.

(A2.3.3) For $t > \exp(1)$: (i) $\gamma(t) = \gamma_0/t$.
(ii) $\sigma(t) = \sigma_0 [t \ln(\ln t)]^{-1/2}$.
(iii) $\gamma_0/\sigma_0^2 < 1/\Lambda$.

COMMENTS ON THE ASSUMPTIONS. (a) Assumption (A2.1) implies that the Gibbs probability measure G_τ, whose density with respect to Lebesgue measure is $g_{V,\tau} = c_\tau \exp[-V/\tau]$, exists for any $\tau > 0$ (cf. [18]).

(b) The constant Λ, depending on V, is accurately described in [16] and [18].

THEOREM 2.

(i) *Rate of convergence in probability*: *under assumptions* (A2.1)–(A2.3), *for all* $r > 0$,

$$\sup_n E\big(v(n)[V(Z_n) - \inf V]1_{\{[V(Z_n)-\inf V]>r\}}\big) < \infty.$$

(ii) *Rate of convergence in expectation*: *if we add the assumption that there exists a constant* $\rho_V > 0$ *such that*, *on a neighborhood of* Argmin V, $\rho_V(V(z) - \inf V) \leq \|\nabla V(z)\|^2$, *then*

$$\sup_n E(v(n)[V(Z_n) - \inf V]) < \infty.$$

REMARKS. (a) The convergence of simulated annealing algorithms only require assumptions (i), (ii), (v), (vi) of (A2.1), $\|\nabla V(z)\| \to \infty$ as $\|z\| \to \infty$, (A2.2) and (A2.3). Thus our assumptions on the function V seem quite restrictive (and are in particular stronger than those of [11], [12] or [29]). However it proves that annealing algorithms converge very slowly since their convergence rates cannot be better than $1/v(n) = \mathscr{C}/\ln n$. Such rates of weak convergence are disappointing for practical purpose; further studies should focus on accelerating this optimization process.

(b) If Argmin V is known to be included in a compact set, we can weaken the assumptions on V. To this end, we replace V by a function that equals V on the compact set and satisfies the strong assumption (A2.1) for large $\|z\|$. Such a procedure is easier than a projection on a suitable compact.

(c) The condition $\rho_V(V - \inf V) \leq \|\nabla V\|^2$ is fulfilled if D^2V is positive definite on Argmin V (hence Argmin V being a finite set). It is also fulfilled if V is regular enough on a neighborhood of the connected components of Argmin V.

THEOREM 3 (Weak convergence rate and small deviations). *Under assumptions* (A2.1)–(A2.3), *and if* $a \mapsto g(a) = \int \exp[-aV(x)]\,dx$ *varies regularly with exponent* $(-\eta)$, $\eta \geq 0$, *we have the following*.

(i) Weak convergence:

$$4v(n)[V(Z_n) - \inf V] \Rightarrow \gamma\Big(\eta, \tfrac{1}{2}\Big),$$

$\gamma(\eta, \frac{1}{2})$ *denoting the Gamma distribution whose density with respect to the Lebesgue measure is proportional to* $e^{-x/2}x^{\eta-1}1_{\{x>0\}}$.

(ii) Small deviations: *for any real function* f *increasing to infinity*,

$$\lim_{n\to\infty} \frac{1}{f(v(n))\ln v(n)} \ln\bigg[P\bigg(V(Z_n) - \inf V \geq \frac{rf(v(n))\ln v(n)}{v(n)}\bigg)\bigg] = -2r.$$

REMARKS. (a) The condition that $a \mapsto g(a) = \int \exp[-aV(x)]\,dx$ varies regularly with exponent $(-\eta)$ is a technical assumption, which is fulfilled in the three cases considered by Hwang [15]. These cases, which ensure the weak convergence as $\tau \to 0$ of the Gibbs distribution G_τ to a probability G_0 concentrated on Argmin V, are the following.

Case 1. Argmin V has a strictly positive Lebesgue measure. In this case G_0 is the uniform distribution on Argmin V, $\eta = 0$ and $\gamma(\eta, 1/2)$ is the Dirac measure in zero.

Case 2. V is a three times continuously differentiable function and $D^2V(z^*)$ is invertible for any $z^* \in$ Argmin V. Then Argmin V is a finite set and G_0 is proportional to the measure

$$\sum_{z^* \in \text{Argmin } V} \left[\det D^2V(z^*)\right]^{-1/2} \delta_{z^*},$$

δ_{z^*} denoting the Dirac measure in z^*. Here $\eta = d/2$ and $\gamma(\eta, 1/2)$ is the chi-square distribution with d degrees of freedom.

Case 3. V is a three times continuously differentiable function, Argmin V has a finite number of connected components, each component being a smooth manifold. Moreover, for all points of these manifolds with the highest dimension, the "second order partial differential of V with respect to smooth normal coordinates" is invertible (see [15] for a precise statement based on regular local coordinates of V). In this case, G_0 concentrates on the highest dimensional components, and $\eta = (d - \nu)/2$, ν being the highest dimension of the regular components.

(b) Assumption (A2.3.2) gives the optimal convergence rate $1/v(t) = 1/(c \ln t)$ for any $c < 1/\Lambda$, whereas (A2.3.1) leads to $1/v(t) > 2/[c(\ln t)]$ and (A2.3.3), which corresponds to the framework studied by Gelfand and Mitter [11], gives a very slow rate ($1/v(t) = 1/[c \ln(\ln t)]$).

(c) For the simulated annealing diffusion, Marquez [29] establishes large deviation principles. For the discrete time algorithm, residual terms prevent us from obtaining such results. See [23] for large deviations results on the escape time from a neighborhood of Argmin V.

(d) Part (i) of Theorem 3 can be used to compute confidence regions with a given error, whereas part (ii) can be used to compute larger confidence regions with an error converging to zero with a rate close to $[v(n)]^{-2r}$.

For our complementary result on the weak convergence rate, we add the following assumption, which corresponds to the second case considered by Hwang [15], thus to $\eta = d/2$ in Theorem 3.

(A2.4) Additional assumptions on the function V.

(i) For any z^* of Argmin V, $D^2V(z^*)$ is positive definite (hence Argmin V is a finite set).

(ii) V is a three times continuously differentiable function.

THEOREM 4. *Under assumptions* (A2.1) *to* (A2.4),

$$\left(Z_n, \sqrt{v(n)}\nabla V(Z_n)\right) \Rightarrow \sum_{z \in \text{Argmin } V} G_0(z)\delta_z \otimes \mathcal{N}\left(0, \tfrac{1}{2}D^2V(z)\right),$$

$\mathcal{N}$ denoting the Gaussian distribution and $\otimes$ the product of measures.

2.2.2. *When the function ∇V is observable only together with a disturbance.* In this case, we consider the gradient algorithm

$$Z_{n+1} = Z_n - \gamma_n[\nabla V(Z_n) + r_{n+1} + \varepsilon_{n+1}] + \sigma_n \xi_{n+1}$$

defined on $(\Omega, \mathscr{A}, \mathscr{P})$ equipped with a filtration $\mathscr{F} = (\mathscr{F}_n)_{n\geq 0}$ with the following assumptions.

(A3.1) Assumptions about the function V. We take up assumption (A2.1).

(A3.2) Assumptions about the gains (γ_n) and (σ_n). We assume either (A3.3.1) or (A3.3.2).

(A3.3.1) We take up (A2.3.1) with $\frac{2}{3} < \alpha < 1$;

(A3.3.2) We take up (A2.3.3).

(A3.3) Assumptions about the disturbances (ε_n) and (r_n). Disturbances (ε_n) and (r_n) are adapted to $\mathscr{F}$; $E(\varepsilon_{n+1}|\mathscr{F}_n) = 0$; $\sup_n E(\|\varepsilon_{n+1}\|^2|\mathscr{F}_n) < +\infty$; $\sup_n n^\delta E(\|r_{n+1}\|^2|\mathscr{F}_n) < \infty$, with $\delta > 1 - \alpha$.

(A3.4) Assumptions about the simulated noise (ξ_n). The random variables r_{n+1} and ξ_{n+1} are independent given $\mathscr{F}_n$; ξ_{n+1} is independent of $\mathscr{F}_n$, with distribution $\mathscr{N}(0, I)$.

(A3.5) We take up (A2.4).

THEOREM 5. *Under assumptions* (A3.1)–(A3.4), *Theorems* 2 *and* 3 *can be applied to* (Z_n). *Under the additional assumption* (A3.5), *Theorem* 4 *can be applied to* (Z_n).

REMARK. If we search for the *local* minima of V, we use the weakly disturbed algorithm $Z_{n+1} = Z_n - \gamma_n[\nabla V(Z_n) + r_{n+1} + \varepsilon_{n+1}]$ and, in view of Theorem 1, the optimal weak convergence rate is obtained when the gain $\gamma_n = \gamma_0/n$ is chosen. But if we search for the *global* minima of V, we use the simulated annealing algorithm $Z_{n+1} = Z_n - \gamma_n[\nabla V(Z_n) + r_{n+1} + \varepsilon_{n+1}] + \sigma_n \xi_{n+1}$, for which the optimal gain is no longer $\gamma_n = \gamma_0/n$. As a matter of fact, a slower gain $\gamma_n = \gamma_0/n^\alpha$, $2/3 < \alpha < 1$, and the corresponding gain σ_n give a better convergence rate for the simulated annealing algorithm.

3. Proofs. We first give two preliminary results in Section 3.1, which are proved in 3.2 and 3.3. Theorems 1, 2, 3, 4 and 5 are proved in Sections 3.4, 3.5, 3.6, 3.7 and 3.8, respectively.

3.1. *Preliminary results.* To investigate the weak convergence of $(\sqrt{v(n)}(Z_n - z^*))$ in the framework of Theorem 1 or of $\sqrt{v(n)}\nabla V(Z_n)$ in the framework of Theorem 3, we first need to establish the tightness of both sequences. In order to unitarily study these two situations, we give a result of tightness in 3.1.1 (Proposition 6) based on an auxiliary Lyapunov function, that is, a function $V\colon \mathbb{R}^d \to \mathbb{R}$, twice continuously differentiable and such that $V(z) \to +\infty$ when $\|z\| \to +\infty$. According to the considered problem, we shall apply this result, either letting V be defined by $V(z) = \|z\|^2$ or letting V be the function whose global minima are searched.

Let $Y_{n+1} = \sqrt{v(n)}(Z_{n+1} - z^*)$ in the framework of Theorem 1. We have then

$$Y_{n+1} = \sqrt{v(n)}(Z_n - z^*) + \gamma_n\Big[\sqrt{v(n)}h(Z_n) + \sqrt{v(n)}r_{n+1}\Big] + \sqrt{\gamma_n}\varepsilon_{n+1}.$$

Since $h(Y_n) = \sqrt{v(n-1)}h(Z_n) + r_{n+1}$ on the neighborhood of z^*, (r_n) being a "small" generic disturbance, we deduce that

$$Y_{n+1} = \sqrt{\frac{v(n)}{v(n-1)}}Y_n + \gamma_n\left[\sqrt{\frac{v(n)}{v(n-1)}}h(Y_n) + \sqrt{v(n)}r_{n+1}\right] + \sqrt{\gamma_n}\varepsilon_{n+1}.$$

But $\sqrt{v(n)/v(n-1)} = 1 + \zeta\gamma_n + o(\gamma_n)$, thus Y_n is given by a recursive relation, which looks like

$$Y_{n+1} = Y_n + \gamma_n[\zeta Y_n + h(Y_n) + r_{n+1}] + \sqrt{\gamma_n}\varepsilon_{n+1}. \tag{3.1}$$

We say that (3.1) is a *strongly disturbed algorithm*. The weak convergence of such an algorithm is studied in Section 3.1.2 (Theorem 7) by taking up the method of the differential stochastic equation, employed by Bouton [3] and Kushner and Huang [25], [26]. But applying this method directly to the strongly disturbed algorithm will also be helpful in studying the weak convergence of $Y_{n+1} = \sqrt{v(n)}\nabla V(Z_{n+1})$ in the framework of Theorem 3.

We introduce some provisional assumptions for those preliminaries (Propsition 6 and Theorem 7), more restrictive than our basic assumptions given in Section 2. Then to prove the results stated in Section 2, we shall use some trucation arguments in order to be able to apply the preliminary results.

3.1.1. *Tightness.* We take up the algorithm (2.1). The following assumptions ensure the tightness of (Z_n).

(AT.1) Assumptions about the disturbances (r_n) and (ε_n).

(i) Z_0 is $\mathscr{F}_0$-measurable and the sequences (r_n) and (ε_n) are adapted to $\mathscr{F}$;
(ii) $E(\varepsilon_{n+1}|\mathscr{F}_n) = 0,\ \forall\ n;\ \sup_{n\geq 0} E(\|\varepsilon_{n+1}\|^2) < \infty$;
(iii) $\sup_{n\geq 0} E(\|r_{n+1}\|^2) < \infty$.

(AT.2) Assumptions about an auxiliary Lyapunov function. There exists a function $V\colon \mathbb{R}^d \mapsto \mathbb{R}^+$, differentiable and such that $V(z) \to +\infty$ when $\|z\| \to +\infty$, with the following conditions:

(i) ∇V is globally Lipschitz;
(ii) There exist two constants $a > 0$ and $C > 0$ such that $\forall\ z \in \mathbb{R}^d$ $\langle \nabla V(z), h(z)\rangle \leq -aV(z) + C$ and $\|\nabla V(z)\|^2 + \|h(z)\|^2 \leq C(1 + V(z))$;
(iii) $E(V(Z_0)) < \infty$.

(AT.3) Assumptions about the gains (γ_n) and (σ_n). We assume either (AT.3.1) or (AT.3.2).

(AT.3.1) (i) (γ_n) is a regular gain with exponent $(-\alpha)$, $0 \le \alpha < 1$.

(ii) $\sigma_n^2 = \gamma_n / v(n)$ where v increases to infinity and varies regularly with exponent $\beta \ge 0$.

(AT.3.2) (i) For $t \ge 1$, $\gamma(t) = \gamma_0/t$ with $\gamma_0 > 0$.

(ii) $\sigma_n^2 = \gamma_n / v(n)$ where v increases to infinity and varies regularly with exponent β, $0 \le \beta < a\gamma_0$, a being defined in (AT.2).

PROPOSITION 6 (Tightness). (i) *Under assumptions* (AT.1) *and* (AT.2),

$$\sup_{n\ge 0} E(V(Z_n)) < \infty.$$

(ii) *Under assumptions* (AT.1)–(AT.3), *and if* $\sup_{n\ge 0} E(v(n)\|r_{n+1}\|^2) < \infty$, *then there exists* R, $0 \le R < \infty$, *such that*

$$\sup_n v(n) E\big(V(Z_n) 1_{\{V(Z_n)\ge R\}}\big) < \infty.$$

3.1.2. *Weak convergence of strongly disturbed algorithms.* We consider the strongly disturbed algorithm

$$Z_{n+1} = Z_n + \gamma_n[h(Z_n) + r_{n+1}] + \sqrt{\gamma_n}\varepsilon_{n+1} \tag{3.2}$$

We make the following assumptions.

(AS.1) Assumptions about the function h. Function h: $\mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz.

(AS.2) Assumptions about an auxiliary Lyapunov function V. There exists a function V: $\mathbb{R}^d \to \mathbb{R}^+$, differentiable and such that we have the following:

(i) ∇V is Lipschitz;

(ii) There exist two constants $a > 0$ and A such that $\forall\ z \in \mathbb{R}^d$ $\langle \nabla V(z), h(z)\rangle \le -aV(z) + A$ and $\|\nabla V(z)\|^2 + \|h(z)\|^2 + \|z\|^2 \le A(1 + V(z))$;

(iii) $E(V(Z_0)) < +\infty$.

(AS.3) Assumptions about the disturbances (ε_n) and (r_n). For a set of trajectories $\Omega_0 \in \mathscr{F}_\infty$ of strictly positive probability, we have the following almost surely on Ω_0:

(i) $E(\varepsilon_{n+1}|\mathscr{F}_n) = 0$;

(ii) There exists a constant $b > 2$ such that $\sup_{n\ge 0} E(\|\varepsilon_{n+1}\|^b|\mathscr{F}_n) < \infty$;

(iii) $E(\varepsilon_{n+1}\varepsilon_{n+1}^T|\mathscr{F}_n) = c_n = C(Z_n) + \Delta_n$ with C Lipschitz from $\mathbb{R}^d$ to the set of positive symmetric matrices; there exist two constants λ_1 and λ_2 such that $0 < \lambda_1 < \lambda_{\min} C(\cdot) < \lambda_{\max} C(\cdot) < \lambda_2$, $\lambda_{\min} C(\cdot)$ [respectively, $\lambda_{\max} C(\cdot)$] denoting the smallest (resp., the largest) eigenvalue of $C(\cdot)$ and $E(\|\Delta_n\| 1_{\Omega_0}) \to 0$;

(iv) The sequence (r_n) is the sum of two sequences $(r_n^{(1)})$ and $(r_n^{(2)})$, adapted to $\mathscr{F}$, and such that $\sup_{n\ge 0} E(\|r_n^{(1)}\|^2) < \infty$, $E(\|r_n^{(1)}\| \cdot 1_{\Omega_0}) \to 0$ and $(r_n^{(2)}) \to 0$ a.s. on Ω_0.

(AS.4) Assumptions about the gains (γ_n). $\sum \gamma_n = +\infty$.

Set $s_n = \sum_{k=0}^n \gamma_k$. We define the continuous process $(Y_t)_{t\geq 0}$, interpolation of the sequence (Z_n), by

$$Y_t = Z_n + (t - s_{n-1})[h(Z_n) + r_{n+1}] + (t - s_{n-1})^{1/2}\varepsilon_{n+1} \quad \text{for } t \in [s_{n-1}, s_n] \tag{3.3}$$

and the family of processes $(Y^{(u)})_{u\geq 0}$ by $Y_t^{(u)} = Y_{u+t}$.

THEOREM 7. *Under assumptions* (AS.1)–(AS.4), *the sequence* (Z_n) *converges weakly, given* Ω_0, *to* μ, *where* μ *is the stationary distribution of the stochastic differential equation* (*SDE*)

$$dX_t = h(X_t)\,dt + C^{1/2}(X_t)\,dB_t. \tag{3.4}$$

Given Ω_0, *the family of processes* $(Y^{(u)})$ *converges weakly to the solution of* (3.4) *with initial distribution* μ.

REMARK. Assumptions (AS.2) imply that the SDE (3.4) is geometrically recurrent, and if (X_t^x) is the solution of (3.4) such that $X_0 = x$, then there exists a constant $\mathscr{C} > 0$ such that

$$\|P(X_t^x \in \cdot) - \mu\| \leq \mathscr{C}\rho^t[1 + V(x)] \quad \text{with } 0 \leq \rho < 1 \tag{3.5}$$

(see [6] or [30]); the condition $0 < \lambda_1 < \lambda_{\min} C(\cdot)$ can be replaced by this property.

3.2. *Proof of Proposition* 6 *stated in* 3.1.1.

3.2.1. *Proof of the first part of Proposition* 6. Let $B > 0$ be a generic constant. Since ∇V is globally Lipschitz, we have, by Taylor's formula,

$$\begin{aligned} V(Z_{n+1}) &\leq V(Z_n) + \langle \nabla V(Z_n), Z_{n+1} - Z_n\rangle + B\|Z_{n+1} - Z_n\|^2; \\ V(Z_{n+1}) &\leq V(Z_n) + \gamma_n\langle \nabla V(Z_n), h(Z_n) + r_{n+1}\rangle + \sigma_n\langle \nabla V(Z_n), \varepsilon_{n+1}\rangle \\ &\quad + B\big(\gamma_n^2[\|h(Z_n)\|^2 + \|r_{n+1}\|^2] + \sigma_n^2\|\varepsilon_{n+1}\|^2\big), \end{aligned}$$

with $\langle \nabla V(Z_n), h(Z_n)\rangle \leq -aV(Z_n) + A$ and $\|h(Z_n)\|^2 \leq A[1 + V(Z_n)]$.

Since $E[\langle \nabla V(Z_n), \varepsilon_{n+1}\rangle] = E[E(\langle \nabla V(Z_n), \varepsilon_{n+1}\rangle | \mathscr{F}_n)] = 0$, it follows that

$$\begin{aligned} E[V(Z_{n+1})] &\leq (1 - a\gamma_n)E[V(Z_n)] + A\gamma_n + \gamma_n E[|\langle \nabla V(Z_n), r_{n+1}\rangle|] \\ &\quad + B\gamma_n^2[A(1 + E[V(Z_n)]) + E(\|r_{n+1}\|^2)] + \sigma_n^2 E(\|\varepsilon_{n+1}\|^2). \end{aligned}$$

However, for any $\alpha > 0$,

$$\begin{aligned} |\langle \nabla V(Z_n), r_{n+1}\rangle| &\leq \frac{\alpha}{2}\|\nabla V(Z_n)\|^2 + \frac{1}{2\alpha}\|r_{n+1}\|^2 \\ &\leq \frac{A\alpha}{2}[1 + V(Z_n)] + \frac{1}{2\alpha}\|r_{n+1}\|^2, \end{aligned}$$

thus

$$E[V(Z_{n+1})] \le \left[1 - \left(a - \frac{A\alpha}{2} - O(\gamma_n)\right)\gamma_n\right] E[V(Z_n)] + O(\gamma_n)\left[1 + (1+\gamma_n)E(\|r_{n+1}\|^2) + \frac{1}{v(n)}E(\|\varepsilon_{n+1}\|^2)\right].$$

Since $\sup_n[E(\|r_{n+1}\|^2)] < +\infty$, $\sup_n[E(\|\varepsilon_{n+1}\|^2)] < +\infty$ and $(\gamma_n) \to 0$, we deduce that for all a_1, $0 < a_1 < a$, there exists n_0 such that $\forall\, n \ge n_0$,

$$E[V(Z_{n+1})] \le (1 - a_1\gamma_n)E[V(Z_n)] + O(\gamma_n).$$

Hence, by a standard lemma (see, e.g., [6]), $\sup_n E[V(Z_n)] < \infty$.

3.2.2. *Proof of the second part of Proposition* 6. First we prove that for all δ and a_2 such that $0 < \delta < a_2 < a$, there exists R, $0 \le R < +\infty$ such that, setting $s_n = \sum_{k=0}^n \gamma_k$,

$$E\big(V(Z_{n+1})1_{\{V(Z_{n+1}) \ge R\}}\big) = O\big(\sup\{\exp(-(a_2-\delta)s_n); \rho(n)\}\big)$$

with $\rho(n) = \exp(-(a_2-\delta)s_n)\sum_{j=0}^n \exp((a_2-\delta)s_j)\gamma_j/v(j)$. Since $\exp(-(a_2-\delta)s_n) = O([v(n)]^{-1})$, we prove then that $\rho(n) = O([v(n)]^{-1})$.

Set $0 < \delta < a_2 < a$, and $r > 0$ such that

$$V(z) \ge r \text{ implies } \langle \nabla V(z), h(z)\rangle \le -a_2 V(z). \tag{3.6}$$

For any R such that $r < R < \infty$, let $\phi\colon \mathbb{R} \to [0,1]$ be a twice continuously differentiable and increasing function such that $\phi(x) = 0\ \forall\, x \in]-\infty, r]$ and $\phi(x) = 1\ \forall\, x \in [R, +\infty[$.

Let us define $\Psi\colon \mathbb{R}^d \to \mathbb{R}^+$ by $\Psi(z) = \phi[V(z)]V(z)$; then

$$\nabla\Psi(z) = \phi[V(z)]\nabla V(z) + [\phi'[V(z)]\nabla V(z)]V(z).$$

Since ∇V is Lipschitz and $\nabla(\phi \circ V)(z) = [\phi'[V(z)]\nabla V(z)]$ equals zero as soon as $V(z) \notin [r, R]$, $\nabla\Psi$ is Lipschitz.

By assumptions (AT.2), $\|\nabla V(z)\|^2 \le A[1 + V(z)]$, thus there exists a constant C_1 such that $V(z) \ge r$ implies $\|\nabla V(z)\|^2 \le C_1 V(z)$. Since $\phi'[V(z)] = \phi[V(z)] = 0$ if $V(z) < r$ and $\phi'[V(z)] = 0$ if $V(z) > R$, we deduce that there exists a constant C_2 such that $\|\nabla\Psi(z)\|^2 \le C_2\Psi(z)$. We also have, $\forall\, z \in \mathbb{R}^d$,

$$\langle \nabla\Psi(z), h(z)\rangle = \phi[V(z)]\langle \nabla V(z), h(z)\rangle + [\phi'[V(z)]V(z)]\langle \nabla V(z), h(z)\rangle,$$

thus, using (3.6),

$$\begin{aligned}\langle \nabla\Psi(z), h(z)\rangle &\le -a_2 V(z)\phi[V(z)] - a_2[V(z)]^2\phi'[V(z)]\\ &\le -a_2 V(z)\phi[V(z)].\end{aligned}$$

We finally deduce that

$$\langle \nabla\Psi(z), h(z)\rangle \le -a_2\Psi(z). \tag{3.7}$$

By Taylor's formula, $\nabla\Psi$ being globally Lipschitz,

$$\Psi(Z_{n+1}) \le \Psi(Z_n) + \langle \nabla\Psi(Z_n), Z_{n+1} - Z_n\rangle + B_1\|Z_{n+1} - Z_n\|^2.$$

Using (3.7),

$$\begin{aligned}E(\Psi(Z_{n+1})) &\le (1-a_2\gamma_n)E[\Psi(Z_n)] + \gamma_n E[\langle \nabla\Psi(Z_n), r_{n+1}\rangle]\\&\quad + O\big(E\big[\gamma_n^2\|h(Z_n)\|^2 + \gamma_n^2\|r_{n+1}\|^2 + \sigma_n^2\|\varepsilon_{n+1}\|^2\big]\big).\end{aligned}$$

However,

$$\big|\langle \nabla\Psi(Z_n), r_{n+1}\rangle\big| \le \frac{\delta}{C_2}\|\nabla\Psi(Z_n)\|^2 + \frac{C_2}{4\delta}\|r_{n+1}\|^2 \le \delta\Psi(Z_n) + \frac{C_2}{4\delta}\|r_{n+1}\|^2,$$

thus

$$\begin{aligned}E[\Psi(Z_{n+1})] &\le [1-(a_2-\delta)\gamma_n]E[\Psi(Z_n)] + \left[\frac{C_2}{4\delta}\gamma_n + O(\gamma_n^2)\right]E(\|r_{n+1}\|^2)\\&\quad + O(\gamma_n^2 E[\|h(Z_n)\|^2]) + O(\sigma_n^2).\end{aligned}$$

Since $E[\|h(Z_n)\|^2] = O([1+E[V(Z_n)]]) = O(1)$ and $E(\|r_{n+1}\|^2) = O([v(n)]^{-1})$,

$$E[\Psi(Z_{n+1})] \le [1-(a_2-\delta)\gamma_n]E(\Psi(Z_n)) + O\left(\frac{\gamma_n}{v(n)}\right).$$

Therefore, using a standard lemma of stabilization (see [6], for instance), and since $\Psi(Z_{n+1}) \ge V(Z_{n+1})1_{\{V(Z_{n+1})\ge R\}}$, we finally obtain

$$E\big(V(Z_{n+1})1_{\{V(Z_{n+1})\ge R\}}\big) = O\big(\sup\{\exp(-(a_2-\delta)s_n), \rho(n)\}\big).$$

We assume first (AT.3.1). Let $s(t) = \int_0^t \gamma(s)\,ds$. We have

$$\rho(n) \sim \exp[-(a_2-\delta)s(n)]\int_0^n \frac{\exp\left((a_2-\delta)s(t)\right)\gamma(t)}{v(t)}\,dt,$$

$$\rho(n) \sim \exp[-(a_2-\delta)s(n)]\int_0^{s(n)} \frac{\exp\left((a_2-\delta)t\right)}{v[s^{-1}(t)]}\,dt.$$

Let x such that $0 < x < 1$; since $t \mapsto v[s^{-1}(t)]$ is increasing,

$$\begin{aligned}\int_0^T \frac{\exp((a_2-\delta)t)}{v[s^{-1}(t)]}\,dt &\le \frac{1}{v[s^{-1}(0)]}\int_0^{Tx}\exp((a_2-\delta)t)\,dt\\&\quad + \frac{1}{v[s^{-1}(Tx)]}\int_{Tx}^T \exp((a_2-\delta)t)\,dt;\\ \int_0^T \frac{\exp((a_2-\delta)t)}{v[s^{-1}(t)]}\,dt &\le \frac{\exp((a_2-\delta)Tx)}{(a_2-\delta)v[s^{-1}(0)]}\\&\quad + \frac{\exp((a_2-\delta)T)}{(a_2-\delta)v[s^{-1}(Tx)]}.\end{aligned}$$

It follows that

$$\left[\frac{\exp((a_2-\delta)T)}{v[s^{-1}(T)]}\right]^{-1}\int_0^T \frac{\exp((a_2-\delta)t)}{v[s^{-1}(t)]}\,dt \le \frac{\exp((a_2-\delta)T(x-1))v[s^{-1}(T)]}{(a_2-\delta)v[s^{-1}(0)]} + \frac{v[s^{-1}(T)]}{(a_2-\delta)v[s^{-1}(Tx)]}.$$

Since $t \mapsto v[s^{-1}(t)]$ varies regularly with exponent $\beta/(1-\alpha)$,

$$\limsup_{T\to\infty}\left[\frac{\exp{(a_2-\delta)T)}}{v[s^{-1}(T)]}\right]^{-1}\int_0^T \frac{\exp((a_2-\delta)t)}{v[s^{-1}(t)]}\,dt \le \frac{x^{-\beta/(1-\alpha)}}{a_2-\delta}.$$

Thus $\rho(n) = O([v(n)]^{-1})$.

We assume now (AT.3.2). Since $\beta < a\gamma_0$, there exist a_2 and δ such that $0 < \delta < a_2 < a$ and $\beta < (a_2-\delta)\gamma_0 < a\gamma_0$, and we have

$$\rho(n) = O\left[n^{-(a_2-\delta)\gamma_0}\int_0^n \frac{t^{(a_2-\delta)\gamma_0-1}}{v(t)}\,dt\right].$$

Since v varies regularly with exponent $\beta \ge 0$,

$$\lim_{T\to\infty}\left[\frac{T^{(a_2-\delta)\gamma_0}}{v(T)}\right]^{-1}\int_0^T \frac{t^{(a_2-\delta)\gamma_0-1}}{v(t)}\,dt = \frac{1}{(a_2-\delta)\gamma_0-\beta}.$$

Thus $\rho(n) = O([v(n)]^{-1})$.

3.3. *Proof of Theorem* 7 *stated in* 3.1.2.

3.3.1. *Some truncations.* We denote $\Omega_{0,N}$ the set of trajectories of Ω_0 such that

$$\sup_{n\ge 0} E\left(\left\|\varepsilon_{n+1}\right\|^b|\mathscr{F}_n\right) \le N \quad\text{and}\quad \sup_{n\ge 0}\left(\left\|r_n^{(2)}\right\|^2\right) \le N.$$

Since Ω_0 is almost surely equal to $\bigcup_N \Omega_{0,N}$, it is sufficient to prove the theorem given $\Omega_{0,N}$ for any N such that $P(\Omega_{0,N}) > 0$.

According to a method used by Lai and Wei (see, for instance, [27]), the first step of the proof of Theorem 7 consists in modifying the algorithm, without changing it on $\Omega_{0,N}$, in order to obtain *everywhere*

$$E(\varepsilon_{n+1}|\mathscr{F}_n) = 0, \qquad \sup_{n\ge 0} E(\|\varepsilon_{n+1}\|^b|\mathscr{F}_n) \le N \quad\text{and}\quad \sup_{n\ge 0}\left(\left\|r_n^{(2)}\right\|^2\right) \le N.$$

This is achieved by replacing $r_n^{(2)}$ by $\tilde{r}_n^{(2)} = r_n^{(2)} \cdot 1_{\{r_n^{(2)} \le N\}}$ and taking $\tilde{\varepsilon}_{n+1} = \varepsilon_{n+1}1_{B_n}$, with

$$B_n = \left\{E\left(\varepsilon_{n+1}|\mathscr{F}_n\right) = 0 \quad\text{and}\quad E\left(\left\|\varepsilon_{n+1}\right\|^b|\mathscr{F}_n\right) \le N\right\}.$$

In the proofs given in Sections 3.3.2, 3.3.3 and 3.3.4, we shall assume that these truncations have been made.

3.3.2. *Tightness of* $(Y^{(u)})$. $Y^{(u)}$ may be written as

$$Y^{(u)} = Y_u + H^{(u)} + R^{(u)} + \sigma^{(u)}$$

with $H_0 = R_0 = \sigma_0 = 0$, $H_t^{(u)} = H_{u+t} - H_u$, $R_t^{(u)} = R_{u+t} - R_u$, $\sigma_t^{(u)} = \sigma_{u+t} - \sigma_u$, and, for $t \in [s_{n-1}, s_n]$,

$$H_t = H_{s_{n-1}} + (t - s_{n-1})h(Z_n),$$
$$R_t = R_{s_{n-1}} + (t - s_{n-1})r_{n+1},$$
$$\sigma_t = \sigma_{s_{n-1}} + (t - s_{n-1})^{1/2}\varepsilon_{n+1}.$$

Let C_1 be a generic constant.

Step 1. The truncations made in Section 3.3.1 enable us to apply Proposition 6, and we get $\sup_{n\geq 0} E(\|Z_n\|^2) \leq \sup_{n\geq 0} E(V(Z_n)) < \infty$.

We deduce that $\sup_{t\geq 0} E(\|Y_t\|^2) \leq \sup_{t\geq 0} E(V(Y_t)) < \infty$. Therefore, given $\Omega_{0,N}$, (Y_t) is tight, and V is integrable with respect to any distribution ν, which is a closure point of (Y_t) for the weak convergence (and the moment of order 2 is finite).

Step 2. On $\Omega_{0,N}$, $r_n^{(2)} \to 0$ a.s and the modifications made in Section 3.3.1 ensure that $(r_n^{(2)})$ is bounded; thus, $E(\|r_n^{(2)}\| 1_{\Omega_{0,N}}) \to 0$.

Finally, $E(\|r_n\| 1_{\Omega_{0,N}}) \to 0$, and $E(\sup_{t\leq T} \|R_t^{(u)}\| 1_{\Omega_{0,N}}) \to 0$; $(R^{(u)})$ converges to zero.

Step 3. The condition $\sup_{n\geq 0} E(\|h(Z_n)\|^2) < \infty$ implies that

$$E(\|H_r^{(u)} - H_t^{(u)}\|^2) \leq C_1(r-t)^2. \tag{3.8}$$

Since $H_0^{(u)} = 0$, the family of processes $(H^{(u)})$ is tight. The inequality (3.8), and thus the tightness, remain true given $\Omega_{0,N}$.

Step 4 The condition $\sup_{n\geq 0} E(\|\varepsilon_n\|^b) < \infty$ ensures, by Burkholder's inequality, that

$$E(\|\sigma_r^{(u)} - \sigma_t^{(u)}\|^b) \leq C_1(r-t)^{b/2} \tag{3.9}$$

with $b/2 > 1$. Since $\sigma_0^{(u)} = 0$, the family of processes $(\sigma^{(u)})$ is tight. The inequality (3.9), and thus the tightness, remain true given $\Omega_{0,N}$.

By Steps 1 to 4, the family of processes $(Y^{(u)})$ is tight given $\Omega_{0,N}$. We shall prove that $(Y^{(u)})$ has a unique closure point for the weak convergence in Section 3.3.4, but we first need some previous results established in 3.3.3.

3.3.3. *Previous results.* We first prove that, for $0 < T < \infty$,

$$\lim_{u\to\infty} E\left(\sup_{t\leq T}\left\| H_t^{(u)} - \int_0^t h(Y_s^{(u)})\,ds\right\| \,\Big|\, \Omega_{0,N}\right) = 0. \tag{3.10}$$

In view of (3.10),

$$E(\|H_r^{(u)} - H_t^{(u)}\|^2 | \Omega_{0,N}) \leq C_1(r-t)^2.$$

Set m such that $\gamma_{m-1} \leq u \leq \gamma_m$, $m \to +\infty$. The function h being Lipschitz, for u large enough,

$$E\left(\sup_{t\leq T}\left\| H_t^{(u)} - \int_u^{u+t} h(Y_s)\,ds\right\| \,\Big|\, \Omega_{0,N}\right) \leq C_1 T \sup_{k\geq m} \sqrt{\gamma_k},$$

which implies (3.10).

We now prove

$$\lim_{u\to\infty} E\big[\big(\sigma_r^{(u)} - \sigma_t^{(u)}\big)\phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)1_{\Omega_{0,N}}\big] = 0 \tag{3.11}$$

and

$$\begin{aligned}\lim_{u\to\infty} E\bigg(\bigg[(\sigma_r^{(u)} - \sigma_t^{(u)})(\sigma_r^{(u)} - \sigma_t^{(u)})^T - \int_{t+u}^{r+u} C(Y_s)\,ds\bigg]\\ \times \phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)1_{\Omega_{0,N}}\bigg) = 0.\end{aligned} \tag{3.12}$$

We consider $0 < t_1 < t_2 < \cdots < t_k < t < r$ and ϕ a continuous and bounded real-valued function defined on $(\mathbb{R}^{2d})^k$.

Let u be large enough to have $s_{p-1} < u < s_p$ with $\gamma_{p-1} + \gamma_p \le (t - t_k)$. Let ξ_{p-1} be a bounded $\mathscr{F}_{p-1}$-measurable random variable. We have

$$E\big((\sigma_r^{(u)} - \sigma_t^{(u)})\phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)\xi_{p-1}\big) = 0.$$

We take $s_{m-1} \le t + u \le s_m$ and $s_{n-1} \le r + u \le s_n$.

Let d_j be defined by $d_m = \sqrt{\gamma_m} - \sqrt{(t+u) - s_{m-1}}$, $d_k = \sqrt{\gamma_k}$ for $m < k < n$, and $d_n = \sqrt{(r+u) - s_{n-1}}$ or $d_n = \sqrt{r - s_{n-1}} - \sqrt{t - s_{n-1}}$ if $n = m$. Then

$$E\bigg(\bigg[\big(\sigma_r^{(u)} - \sigma_t^{(u)}\big)\big(\sigma_r^{(u)} - \sigma_t^{(u)}\big)^T - \sum_{i=m}^{n} d_i^2 c_i\bigg]\phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)\xi_{p-1}\bigg) = 0.$$

Now set $\xi_p = P(\Omega_{0,N}|\mathscr{F}_p)$; the bounded martingale (ξ_p) converges towards $1_{\Omega_{0,N}}$, almost surely and in expectation. Thus, $\forall\ p$,

$$\begin{aligned}&\limsup_{u\to\infty}\big|E\big((\sigma_r^{(u)} - \sigma_t^{(u)})\phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)1_{\Omega_{0,N}}\big)\big|\\ &\qquad \le C_1\sqrt{r-t}\big\{E\big(\big|1_{\Omega_{0,N}} - \xi_{p-1}\big|^{b/(b-1)}\big)\big\}^{(b-1)/b}\end{aligned}$$

from which we deduce (3.11).

We prove in the same way that

$$\lim_{u\to\infty} E\bigg(\bigg[\big(\sigma_r^{(u)} - \sigma_t^{(u)}\big)\big(\sigma_r^{(u)} - \sigma_t^{(u)}\big)^T - \sum_{i=n}^{m} d_i^2 c_i\bigg]\phi\big((Y_{t_j}^{(u)}, \sigma_{t_j}^{(u)})_{1\le j\le k}\big)1_{\Omega_{0,N}}\bigg) = 0.$$

And, C being Lipschitz, we deduce (3.12).

3.3.4. *Convergence of* $(Y^{(u)})$. Let G be the set of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^{2d}$, equipped with the metric of uniform convergence on compact sets. The G-valued family $(Y^{(u)}, \sigma^{(u)})$ is tight, given $\Omega_{0,N}$. Let $\mathscr{G}$ be the Borel σ-algebra of G. Let $(u(n))$ be an increasing sequence such that, given $\Omega_{0,N}$,

$$\big(Y^{(u(n))}, \sigma^{(u(n))}\big) \Rightarrow Q,\ \ Q \text{ probability on } (G, \mathscr{G}).$$

We denote by $\omega = ([X_t(\omega), S_t(\omega)]_{t\ge 0})$ a point ω of G and by $\mathscr{G}_t$ the σ-field spanned by $(X_r, S_r)_{r\le t}$.

Let $0 \le t_1 < t_2 < \cdots < t_k < t < r$. Given $\Omega_{0,N}$, $([Y_{t_j}^{(u(n))}, \sigma_{t_j}^{(u(n))}]_{1\le j\le k}, \sigma_t^{(u(n))}, \sigma_r^{(u(n))})$ converges weakly to the distribution of $([X_{t_j}, S_{t_j}]_{1\le j\le k}, S_t, S_r)$ on $(G, \mathscr{G}, Q)$. The relations (3.9), (3.11) and (3.12) imply

$$E_Q\big((S_r - S_t)\phi\big((X_{t_j}, S_{t_j})_{1\le j\le k}\big)\big) = 0;$$

$$E_Q\bigg(\bigg[(S_r - S_t)(S_r - S_t)^T - \int_t^r C(X_s)ds\bigg]\phi\big((X_{t_j}, S_{t_j})_{1\le j\le k}\big)\bigg) = 0.$$

On $(G, \mathscr{G}, Q)$, $(S_t)_{t\ge 0}$ is a square-integrable continuous martingale; it is adapted to $(\mathscr{G}_t)_{t\ge 0}$, and its increasing process is $(\int_0^t C(X_s)\,ds)_{t\ge 0}$.

We deduce that $dS_t = C^{1/2}(X_t)\,dB_t$, where $(B_t)_{t\ge 0}$ is a Brownian motion. From (3.10),

$$(H_t^{(u(n))})_{t\ge 0} \Rightarrow \bigg(\int_0^t h(X_s)\,ds\bigg)_{t\ge 0}.$$

Thus,

$$X_t = X_0 + \int_0^t h(X_s)\,ds + \int_0^t C^{1/2}(X_s)\,dB_s$$

where $(B_t)_{t\ge 0}$ is a Brownian motion adapted to $(\mathscr{G}_t)_{t\ge 0}$.

Let a probability on $\mathbb{R}^d$, ν, be a closure point for the weak convergence of the distributions, given $\Omega_{0,N}$, of $(Y_t)_{t\ge 0}$. According to Step 1 of Section 3.3.2, there exists a sequence $u(n)$, increasing to infinity such that $Y_{u(n)} \Rightarrow \nu$ (given $\Omega_{0,N}$) and $\nu(V) < +\infty$.

Let μ be the unique stationary distribution of (3.4). In order to conclude the proof of Theorem 7, it is enough to prove that any closure point for the weak convergence conditional on $\Omega_{0,N}$ of $(Y^{(u(n))})$ is a solution of the SDE (3.4) with initial distribution μ.

Let ϕ: $\mathbb{R}^d \to \mathbb{R}$ be continuous and bounded. For any $t > 0$, $(Y_{u(n)-t})_{n\ge 0}$ is tight, given $\Omega_{0,N}$, and there exists a subsequence $(w(n))$ of $(u(n))$ such that $Y_{w(n)-t} \Rightarrow \nu_1$ (given $\Omega_{0,N}$), where $\nu_1(V) < \infty$. Given $\Omega_{0,N}$, $(Y^{(w(n)-t)})$ converges weakly to $(X_t)_{t\ge 0}$, the solution of (3.4) with initial distribution ν_1. Then, it results from (3.5) that

$$\bigg|\frac{E(\phi(Y_{w(n)})1_{\Omega_{0,N}})}{P(\Omega_{0,N})} - \mu(\phi)\bigg| \le \big|E_{\nu_1}\big(\phi(X_t)\big) - \mu(\phi)\big| + \bigg|\frac{E_{\nu_1}(\phi(Y_t^{(w(n)-t)})1_{\Omega_{0,N}})}{P(\Omega_{0,N})} - E(\phi(X_t))\bigg|$$

with $|E_{\nu_1}(\phi(X_t)) - \mu(\phi)| = O(\rho^t)$ where $0 \le \rho < 1$ since $\nu_1(V) < \infty$. Thus $|\nu(\phi) - \mu(\phi)| = O(\rho^t)$ for all $t > 0$ and $\nu(\phi) = \mu(\phi)$. Hence $\nu = \mu$.

3.4. *Proof of Theorem* 1. First we show in Section 3.4.1 that the weak convergence rate of (Z_n) can be obtained by proving the same rate for a locally similar sequence $(\widetilde{Z}_n)$. Then we prove the tightness of $\sqrt{v(n)}(\widetilde{Z}_n - z^*)$ in 3.4.2, which will finally enable us to apply, in 3.4.3, Theorem 7 to the strongly disturbed algorithm defined by $\widetilde{Y}_n = \sqrt{v(n)}(\widetilde{Z}_n - z^*)$.

3.4.1. *Definition of* $(\widetilde{Z}_n)$. We can choose M as small as we want in (A1.2). Once M is fixed, set $\Gamma_N = \Gamma(z^*) \cap \{\sup_{n\geq N} \|Z_n - z^*\| \leq M\}$.

To prove the weak convergence of $\sqrt{v(n)}(Z_n - z^*)$ given $\Gamma(z^*)$, it is sufficient to prove it given Γ_N for all N.

For a given N, let $(\widetilde{Z}_n)_{n\geq N}$ be defined by $\widetilde{Z}_N = Z_N 1_{\{\|Z_N - z^*\| \leq M\}}$, and

$$\widetilde{Z}_{n+1} = \widetilde{Z}_n + \gamma_n F(\widetilde{Z}_n) + [\sigma_n \varepsilon_{n+1} + \gamma_n r_{n+1}] 1_{\{\|Z_n - z^*\| \leq M\}} \text{ if } n \geq N,$$

where $F(z) = h(z) 1_{\{\|z - z^*\| \leq M\}} - K(z - z^*) 1_{\{\|z - z^*\| > M\}}$, with a constant $K > 0$ to be specified later on.

The sequences $(Z_n)_{n\geq N}$ and $(\widetilde{Z}_n)_{n\geq N}$ agree on Γ_N. Thus, we only have to prove the weak convergence of $\sqrt{v(n)}(\widetilde{Z}_n - z^*)$ given Γ_N.

3.4.2. *Tightness of* $([v(n)]^{1/2}(\widetilde{Z}_n - z^*))$. Let us show that, if M has been chosen small enough, then, for a suitable K,

$$\sup_{n\geq N} E(v(n)\|\widetilde{Z}_n - z^*\|^2) < \infty,$$

$$\begin{aligned}\widetilde{Z}_{n+1} - z^* = {} & \big[I + \gamma_n \rho(\widetilde{Z}_n - z^*) H\big](\widetilde{Z}_n - z^*) \\ & + \gamma_n \big[h(\widetilde{Z}_n) - \rho(\widetilde{Z}_n - z^*) H(\widetilde{Z}_n - z^*)\big] 1_{\{\|\widetilde{Z}_n - z^*\| \leq M\}} \\ & + \gamma_n \big[-\rho(\widetilde{Z}_n - z^*) H - KI\big](\widetilde{Z}_n - z^*) 1_{\{\|\widetilde{Z}_n - z^*\| > M\}} \\ & + \big[\sigma_n \varepsilon_{n+1} + \gamma_n r_{n+1}\big] 1_{\{\|Z_n - z^*\| \leq M\}}.\end{aligned}$$

Set $t > 0$, and let Q_t be an invertible matrix such that $Q_t H Q_t^{-1} = H_t$ is a matrix whose diagonal terms are the eigenvalues of H, the terms below being 0 or t, and the other ones equal to zero. For $S_n = Q_t(\widetilde{Z}_n - z^*)$, we have

$$\begin{aligned}S_{n+1} = {} & [I + \gamma_n \rho(\widetilde{Z}_n - z^*) H_t] S_n + \gamma_n \big[-\rho(\widetilde{Z}_n - z^*) H_t - KI\big] S_n 1_{\{\|\widetilde{Z}_n - z^*\| > M\}} \\ & + \gamma_n Q_t \big[h(\widetilde{Z}_n) - \rho(\widetilde{Z}_n - z^*) H(\widetilde{Z}_n - z^*)\big] 1_{\{\|\widetilde{Z}_n - z^*\| \leq M\}} \\ & + \big[\sigma_n Q_t \varepsilon_{n+1} + \gamma_n Q_t r_{n+1}\big] 1_{\{\|Z_n - z^*\| \leq M\}}.\end{aligned}$$

For A and δ such that $0 < A < 2L$ and $0 < \delta < 2L - A$, there exists t, $0 < t \leq 1$, such that for n large enough, $n \geq n_0$, $\forall\, x \in \mathbb{R}^d$,

$$\big\|I + \gamma_n \rho(x) H_t\big\|^2 \leq \big[1 - (A + \delta)\gamma_n\big].$$

According to (A1.1), there exists M satisfying (A1.2) such that

$$\big\|z - z^*\big\| \leq M \text{ implies } \big\|h(z) - \rho(z - z^*) H\big(z - z^*\big)\big\| \leq C_1 \big\|z - z^*\big\|^2$$

and such that

$$\big\|z - z^*\big\| \leq M \text{ implies } \big\|h(z) - \rho(z - z^*) H(z - z^*)\big\| \leq c\|z - z^*\|,$$

with $c\|Q_t\|\|Q_t\|^{-1} < \delta/4$ and $C_1 > 0$. Finally, we choose $K > 0$ such that

$$y^T[-\rho(x)H_t - KI]y \le 0 \qquad \forall\ y \in \mathbb{R}^d,\ \forall\ x \in \mathbb{R}^d.$$

Then for any $n \ge n_0$, n_0 large enough,

$$\begin{aligned} E(\|S_{n+1}\|^2) \le \left(1 - \left[A + \frac{\delta}{2} - \delta' - O(\gamma_n)\right]\gamma_n\right) E(\|S_n\|^2) \\ + O\big(\sigma_n^2 E\big(\|\varepsilon_{n+1}\|^2 1_{\{\|Z_n - z^*\| \le M\}}\big)\big) \\ + O\big(\gamma_n E\big(\|r_{n+1}\|^2 1_{\{\|Z_n - z^*\| \le M\}}\big)\big), \end{aligned}$$

with an arbitrary $\delta' > 0$. For n large enough, we have

$$E(\|S_{n+1}\|^2) \le (1 - A\gamma_n)E(\|S_n\|^2) + O\left(\frac{\gamma_n}{v(n)}\right),$$

from which we deduce that $\sup_n E[v(n)\|S_{n+1}\|^2] < \infty$.

Finally $\sup_n E[v(n)\|\widetilde{Z}_{n+1} - z^*\|^2] < \infty$.

3.4.3. *Application of Theorem* 7. Let $\widetilde{Y}_{n+1} = \sqrt{v(n)}[\widetilde{Z}_{n+1} - z^*]$; we have

$$\sup_n E\big[\|\widetilde{Y}_n\|^2\big] < \infty. \tag{3.13}$$

The expression $\widetilde{Y}_{n+1}$ may be written in the following way:

$$\widetilde{Y}_{n+1} = \widetilde{Y}_n + \gamma_n\big[\alpha I + \rho\big(\widetilde{Z}_n - z^*\big)H\big]\widetilde{Y}_n + \gamma_n \widetilde{r}_{n+1} + \sqrt{\gamma_n}\widetilde{\varepsilon}_{n+1}, \tag{3.14}$$

with $\widetilde{\varepsilon}_{n+1} = \varepsilon_{n+1} 1_{\{\|Z_n - z^*\| \le M\}}$ and with

$$\begin{aligned} \widetilde{r}_{n+1} &= \widetilde{u}_n \widetilde{Y}_n + \sqrt{v(n)}\big[r_{n+1}^{(1)} + r_{n+1}^{(2)} + r_{n+1}^{(3)}\big], \\ &\qquad (\widetilde{u}_n) \text{ deterministic sequence converging to } 0, \\ r_{n+1}^{(1)} &= \big[h(\widetilde{Z}_n) - \rho\big(\widetilde{Z}_n - z^*\big)H\big(\widetilde{Z}_n - z^*\big)\big] 1_{\{\|\widetilde{Z}_n - z^*\| \le M\}}, \\ r_{n+1}^{(2)} &= \big[-\rho\big(\widetilde{Z}_n - z^*\big)H - KI\big]\big(\widetilde{Z}_n - z^*\big) 1_{\{\|\widetilde{Z}_n - z^*\| > M\}}, \\ r_{n+1}^{(3)} &= r_{n+1} 1_{\{\|\widetilde{Z}_n - z^*\| \le M\}}. \end{aligned}$$

We verify now that the strongly disturbed algorithm (3.14) fulfills the assumptions of Theorem 7.

Assumptions (AS.1) *are fulfilled*. According to the properties of the function ρ, $\rho(\widetilde{Z}_n - z^*) = \rho(\widetilde{Y}_n)$, and the function $f\colon z \mapsto [\alpha I + \rho(z)H]z$ is Lipschitz.

Auxiliary Lyapunov function. The Lyapunov function $V\colon \mathbb{R}^d \to \mathbb{R}$ defined by $V(z) = \|z\|^2$ fulfills the assumption (AS.2) with $a = \alpha - L$.

Properties of $(\widetilde{r}_n)$. According to (3.13) and since $(\widetilde{u}_n) \to 0$,

$$\sup_n E\big[\big(\widetilde{u}_n\|\widetilde{Y}_n\|\big)^2\big] < \infty, \qquad E\big[\widetilde{u}_n\|\widetilde{Y}_n\|1_{\Gamma_N}\big] \to 0.$$

According to the choice of M in Section 3.4.2, we have $\|r_{n+1}^{(1)}\| \le C_1\|\widetilde{Z}_n - z^*\|^2 1_{\{\|\widetilde{Z}_n - z^*\| \le M\}}$, and, applying (3.13), we deduce that

$$\sup_n E\Big[\Big(\sqrt{v(n)}\|r_{n+1}^{(1)}\|\Big)^2\Big] \le \sup_n E\big[v(n)\|\widetilde{Z}_n - z^*\|^2\big] < \infty,$$

$$E\Big[\Big(\sqrt{v(n)}\|r_{n+1}^{(1)}\|\Big)1_{\Gamma_N}\Big] \le C_1 E\Big[\sqrt{v(n)}\|\widetilde{Z}_n - z^*\|^2\Big] \le \frac{C_1}{\sqrt{v(n)}} \to 0.$$

We have $\|r_{n+1}^{(2)}\| \le C_2\|\widetilde{Z}_n - z^*\| 1_{\{\|\widetilde{Z}_n - z^*\| > M\}}$, thus

$$\sup_n E\Big[\Big(\sqrt{v(n)}\|r_{n+1}^{(2)}\|\Big)^2\Big] < \infty.$$

Since (Z_n) and $(\widetilde{Z}_n)$ agree on Γ_N, we have $1_{\{\|\widetilde{Z}_n - z^*\| > M\}} \cdot 1_{\Gamma_N} = 1_{\{\|Z_n - z^*\| > M\} \cap \Gamma_N} = 0$, and

$$E\Big[\sqrt{v(n)}\|r_{n+1}^{(2)}\| 1_{\Gamma_N}\Big] = 0.$$

Since $(r_{n+1}^{(3)})$ fulfills the conditions required by the assumptions (A1.2), we finally deduce that

$$\sup_n E\big[\|\widetilde{r}_{n+1}\|^2\big] < \infty \quad \text{and} \quad E\big[\|\widetilde{r}_{n+1}\| 1_{\Gamma_N}\big] \to 0.$$

Properties of $(\widetilde{\varepsilon}_n)$. $(\widetilde{\varepsilon}_n)$ fulfills the assumptions (AS.3) with $C(z) = \Gamma$ for all $z \in \mathbb{R}^d$.

According to Theorem 7, given Γ_N, $(\widetilde{Y}_n) \Rightarrow \mu$, where μ is the stationary distribution of (2.3), and the first part of Theorem 1 is proved.

Let ϕ be a real continuous and bounded function, and ξ a $\mathscr{F}_\infty$-mesurable random variable. Replacing $\Gamma(z^*)$ by $\Gamma(z^*) \cap \{\phi(\xi) < t\}$ as soon as this event is nonnegligible, we obtain the second assertion. □

3.5. *Proof of Theorem* 2. We simplify the proof by taking V instead of $(V - \inf V)$, that is, by assuming $\inf V = 0$.

Following Gelfand and Mitter's idea [11], we define in Section 3.5.1 a continuous time interpolation $(U_t)_{t \ge 0}$ of (Z_n). This interpolated process looks like an annealing diffusion

$$dY_t = -a(t)\nabla V(Y_t)\,dt + dB_t,$$

the function a, slowly increasing to infinity, being precised below. In Section 3.5.2 we precise how long the process (U_{t+u}) follows (Y_{t+u}) given $U_u = Y_u = x$. Then, in 3.5.3, we obtain Theorem 2, transfering properties of the annealing diffusion process to the algorithm.

We set $t_n = \sum_{k=0}^n \sigma_k^2$, and $K(t) = \int_0^t \sigma^2(s)\,ds$. Under (A2.3.1) and (A2.3.2), σ^2 varies regularly with exponent $(-\alpha)$, $\alpha < 1$. Thus K varies regularly with exponent $1 - \alpha > 0$ and $K(t) \to \infty$ as $t \to \infty$. This property holds under (A2.3.3) too.

3.5.1. *Definition of the interpolation* (U_t) *and tightness of* $(V(U_t))$.

Step 1 [Definition of (U_t)]. We may build up the sequence $(\sigma_n \xi_{n+1})$ from a Brownian motion $B = (B_t)_{t \geq 0}$ independent of Z_0 by setting $\sigma_n \xi_{n+1} = B_{t_n} - B_{t_{n-1}}$. Then, we define the process $(U_t)_{t \geq 0}$, interpolation of the sequence (Z_n), by

$$U_t = Z_n - (t - t_{n-1}) v(n) \nabla V(Z_n) + B_t - B_{t_{n-1}} \quad \text{if } t \in [t_{n-1}, t_n].$$

Let $a: \mathbb{R}^+ \to \mathbb{R}^+$ be a continuously differentiable, concave and slowly increasing function such that $a(t_{n-1}) = v(n)$ for all integer n. The term U_t may be written as

$$U_t = Z_n - (t - t_{n-1}) a(t_{n-1}) \nabla V(Z_n) + B_t - B_{t_{n-1}} \text{ if } t \in [t_{n-1}, t_n].$$

Step 2 [Tightness of $(V(U_t))$]. Let C_1 be a generic constant. Proposition 6 may be applied by choosing V as the auxiliary Lyapunov function and by setting $h = -\nabla V$: $\sup_n E(V(Z_n)) < \infty$ and there exists R, $0 \leq R < \infty$, such that

$$\sup_n E\big(v(n) V(Z_n) 1_{\{V(Z_n) \geq R\}}\big) < \infty. \tag{3.15}$$

Let $t \in [t_{n-1}, t_n]$, and Ψ be defined in the same way as in the proof of Proposition 6. Since Ψ is globally Lipschitz,

$$\Psi(U_t) \leq \Psi(Z_n) + \langle \nabla \Psi(Z_n), U_t - Z_n \rangle + C_1 \|U_t - Z_n\|^2.$$

Since there exists a constant $A_1 > 0$ such that $\langle \nabla \Psi(z), -\nabla V(z) \rangle \leq -A_1 \Psi(z)$, we have

$$\begin{aligned} \Psi(U_t) &\leq \Psi(Z_n)\big[1 - A_1 (t - t_{n-1}) v(n)\big] + \langle \nabla \Psi(Z_n), B_t - B_{t_{n-1}} \rangle \\ &\quad + C_1 \big[(t - t_{n-1})^2 [v(n)]^2 (1 + V(Z_n)) + \|B_t - B_{t_{n-1}}\|^2\big]; \\ E(\Psi(U_t)) &\leq C_1 \big[E(\Psi(Z_n)) + (t - t_{n-1})^2 [v(n)]^2 (1 + E(V(Z_n))) + (t - t_{n-1})\big]. \end{aligned}$$

Since $\sup_n E(V(Z_n)) < \infty$ and $\sup_n E(v(n) \Psi(Z_n)) < \infty$, we deduce that

$$E(\Psi(U_t)) = O\bigg(\frac{1}{v(n)}\bigg).$$

Finally

$$E(V(U_t) 1_{\{V(U_t) \geq R\}}) = O\bigg(\frac{1}{a(t)}\bigg). \tag{3.16}$$

3.5.2. *Previous result on the annealing diffusion process.* The stochastic differential equation $dY_t = a(t) \nabla V(Y_t)\, dt + dB_t$ has been studied in [5], [13], [16], [32] and, recently, in [29] under our assumptions on V and under the additional assumption: for a constant $c^* < 1/\Lambda$ and for t large enough, $a(t) \leq c^* \ln t$.

We first check that there exists $c^* < 1/\Lambda$ such that $a(t) \leq c^* \ln t$, before enunciating a result proved in [29].

Under (A2.3.1) or (A2.3.2): since σ^2 varies regularly with exponent $(-\alpha)$, $\frac{1}{2} \leq \alpha < 1$, we have $K(t) \sim (t\sigma^2(t))/(1-\alpha)$. Thus K varies regularly with exponent $1-\alpha$; that is, $K(t) = t^{1-\alpha}L(t)$, with L varying slowly.

Since $v(t) \leq c[1-\alpha]\ln(t)$, we have $v(t) \leq c \cdot \ln K(t) - c \cdot \ln L(t)$. But $c < 1/\Lambda$ and $\lim_{t\to\infty}(\ln L(t)/\ln K(t)) = 0$, thus there exists $\bar{c} < 1/\Lambda$ such that, for t large enough, $v(t) \leq \bar{c}\ln K(t)$.

Under (A2.3.3):

$$\sigma^2(t) = \frac{\sigma_0^2}{t\ln(\ln t)}, \frac{\sigma_0^2 \ln t}{\ln(\ln t)} \leq K(t) \leq \sigma_0^2 \ln t, \text{ and } v(t) = c\ln(\ln t).$$

Thus $v(t) \leq c\ln K(t)[1+o(1)]$, and there exists $\bar{c} < 1/\Lambda$ such that, for t large enough, $v(t) \leq \bar{c}\ln K(t)$.

In both cases, $a(K^{-1}(t_{n-1})) \leq \bar{c}\ln n$, for n large enough. However $t_{n-1} \sim K(n)$; thus there exists $\tilde{c} < 1/\Lambda$ such that $a(n) \leq \tilde{c}\ln n$ for n large enough. Set $t \in [n, n+1]$, n large enough. We have

$$a(t) = a(n) + a'(t^*)(t-n) \quad \text{with } t^* \in [n, t].$$

$a'(t) = o(a(t)/t)$, thus $a'(t^*) \leq (\tilde{c}\ln(n+1)/n)$, which enables us to conclude.

A result of Márquez [29]. Let ϕ be a continuous function with compact support, and let K_1 be a suitable compact set that contains the support of ϕ. Precising upper bounds in the works of Chiang, Hwang and Sheu [5], [16] or of Royer [32], Márquez [29] has proved that there exist a positive continuous and increasing function τ with $\tau(t) = O(t^{2/3})$ and a constant $\rho_1 > 0$ such that, uniformly for $x \in K_1$,

$$P\Big(\sup_{t\leq s\leq t+\tau(t)} \|Y_s\| \geq C | Y_t = x\Big) \leq \exp[-\rho_1 a(t)]; \tag{3.17}$$

$$\big|E(\phi(Y_{t+\tau(t)})|Y_t = x) - G_{1/2a(t+\tau(t))}(\phi)\big| \leq \exp[-\rho_1 a(t)]. \tag{3.18}$$

3.5.3. *Following the annealing diffusion process.*

Step 1 (Following the annealing diffusion process under the provisional assumption "∇V is bounded"). Here we assume that ∇V is bounded.

For all $u > 0$, let $B^{(u)} = (B_t^{(u)})_{t\geq 0}$ with $B_t^{(u)} = B_{u+t} - B_u$.

Let $Y^{(u)}$ and $U^{(u)}$ be defined by

$$\begin{aligned} U_0^{(u)} &= Y_0^{(u)} = x, \\ dY_t^{(u)} &= -[a(u+t)]\nabla V(Y_t^{(u)})\,dt + dB_t^{(u)}, \\ dU_t^{(u)} &= [-a(u+t)\nabla V(U_t^{(u)}) + h_{t+u}]\,dt + dB_t^{(u)}, \end{aligned}$$

where $h_t = -a(t_{m-1})\nabla V(U_{t_{m-1}}) + a(t)\nabla V(U_t)$ for $t \in [t_{m-1}, t_m]$.

Set $h_t = H_t((U_s^{(u)}), s \leq t)$.

Let $\mathscr{C}$ be the set of the $\mathbb{R}^d$-valued continuous functions defined on $\mathbb{R}^+$ and assume that $\mathscr{C}$ is equipped with its Borel σ-field. We denote a function of

$\mathscr{C}$ as $\omega = (X_t(\omega))_{t\geq 0}$ and $(\mathscr{G}_t) = (\sigma\{X_s,\ s \leq t\})$. By Girsanov's theorem, $(Y^{(u)} - x)$ and $(U^{(u)} - x)$ have canonical versions with distributions P_Y and P_U conditionnally on $\mathscr{G}_t$. We are going to prove that for all $\Gamma \in \mathscr{G}_t$,

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| = O\left(\frac{1}{a(u)}\right).$$

By Girsanov's theorem, for each t, P_Y and P_U have a density, denoted by L_t^Y and L_t^U respectively, with respect to the Wiener measure W. Moreover, for all $\Gamma \in \mathscr{G}_t$,

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| \leq E_W\left[L_t^Y\left|1 - \frac{L_t^U}{L_t^Y}\right|\right] \leq \left[E_Y\left(\left[1 - \frac{L_t^U}{L_t^Y}\right]^2\right)\right]^{1/2},$$

where, for the probability P_Y,

$$\begin{aligned}\ln\frac{L_t^U}{L_t^Y} &= \int_0^t \langle H_s[(X_r + x)_{r\leq s}], dX_s\rangle \\ &\quad - \frac{1}{2}\int_0^t \Big\{\big\|H_s[(X_r + x)_{r\leq s}]\big\|^2 \\ &\qquad + 2\Big\langle H_s[(X_r + x)_{r\leq s}], -a(s+\tau)\nabla V(X_s + x)\Big\rangle\Big\}\, ds.\end{aligned}$$

Moreover, under P_Y, $\ln(L_t^U/L_t^Y)$ has the same distribution as $H_t - \frac{1}{2}J_t$ has, with

$$H_t = \int_u^{u+t} \langle H_s[(Y_r)_{r\leq s}], dB_s\rangle; \qquad J_t = \int_u^{u+t} \big\|H_s[(Y_r)_{r\leq s}]\big\|^2\, ds.$$

Since the expectations of $\exp(H_t - \frac{1}{2}J_t)$ and $\exp(4H_t - 8J_t)$ are equal to 1, we have

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| \leq E_Y\left[\left(1 - \exp\left(H_t - \tfrac{1}{2}J_t\right)\right)^2\right] = E_Y[\exp(2H_t - J_t)] - 1,$$

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| \leq E_Y\left[\exp\left(2H_t - 4J_t\right)\exp(3J_t)\right] - 1 \leq [E_Y(\exp(6J_t))]^{1/2} - 1.$$

Thus

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| \leq \sup_{t\geq 0}\left|[E_Y(\exp(6J_t))]^{1/2} - 1\right|.$$

Let us study $E_Y(\exp(6J_t))$. For $u \in [t_{n-1}, t_n]$, we have

$$J_t = O(|t_n - t_{n-1}|) + \int_{t_n}^{u+t} \|k_s\|^2\, ds \tag{3.19}$$

with $k_s = \nabla V(Y_s)[a(s) - a(t_{m-1})] - v(m)[\nabla V(Y_{t_{m-1}}) - \nabla V(Y_s)]$, for $s \in [t_{m-1}, t_m]$, $m > n$.

Since ∇V is assumed to be Lipschitz and bounded, there exists a constant C_1 such that

$$\|k_s\|^2 \leq C_1\{[v(m)]^4(t_m - t_{m-1})^2 + [v(m)]^2\|B_s - B_{t_{m-1}}\|^2\}.$$

The study of $E_Y(\exp(6J_t))$ requires the study of $E[\exp(C_1[v(m)]^2 \int_{t_{m-1}}^{t_m} \|B_s - B_{t_{m-1}}\|^2\, ds)]$. We have

$$E\left[\exp\left(C_1[v(m)]^2 \int_{t_{m-1}}^{t_m} \|B_s - B_{t_{m-1}}\|^2\, ds\right)\right] = E\left[\exp\left(C_1[v(m)]^2 (t_m - t_{m-1})^2 \int_0^1 \|B_s\|^2\, ds\right)\right].$$

By Jensen's inequality,

$$E\left[\exp\left(C_1[v(m)]^2 (t_m - t_{m-1})^2 \int_0^1 \|B_s\|^2\, ds\right)\right] \le \int_0^1 E\big[\exp\big(C_1[v(m)]^2 (t_m - t_{m-1})^2 \|B_s\|^2\big)\big]\, ds.$$

Hence, Φ denoting the Laplace transform of the chi-square distribution with d degrees of freedom,

$$E\left[\exp\left(C_1[v(m)]^2 \int_{t_{m-1}}^{t_m} \|B_s - B_{t_{m-1}}\|^2\, ds\right)\right] \le \Phi[C_1[v(m)]^2 (t_m - t_{m-1})^2].$$

It follows that, for m large enough,

$$E\left[\exp\left(C_1[v(m)]^2 \int_{t_{m-1}}^{t_m} \|B_s - B_{t_{m-1}}\|^2\, ds\right)\right] \le \left[1 - 2C_1[v(m)]^2 \big(t_m - t_{m-1}\big)^2\right]^{-d/2}.$$

We deduce that

$$E_Y\left[\exp\left(6 \int_{t_{m-1}}^{t_m} \|k_s\|^2\, ds\right)\right] \le \big(1 - 2C_1[v(m)]^2 (t_m - t_{m-1})^2\big)^{-d/2} \times \exp\big[C_1[v(m)]^4 (t_m - t_{m-1})^3\big].$$

For any u large enough such that $t_{n-1} \le u < t_n$, we have, in view of (3.19),

$$\sup_{t \ge 0} E_Y\left(\exp 6 \int_{t_n}^{t+u} \|k_s\|^2\, ds\right) \le M(n)$$

with

$$M(n) = \prod_{j=n}^{\infty} \big(1 - 2C_1[v(j)]^2 (t_j - t_{j-1})^2\big)^{-d/2} \exp\left(C_1 \sum_{j=n}^{\infty} [v(j)]^4 (t_j - t_{j-1})^3\right).$$

When n tends to infinity,

$$M_n \sim \exp\left(C_1 \sum_{j=n}^{\infty} [v(j)]^2 (t_j - t_{j-1})^2\right) = \exp\left(C_1 \sum_{j=n}^{\infty} \gamma_j^2\right).$$

It follows that, for $u \in [t_{n-1}, t_n]$,

$$\sup_{t \ge 0}[E_Y(\exp(6J_t))] = O\left[\exp\left(\sum_{j=n}^{+\infty} \gamma_j^2\right)\right].$$

Thus, for all u such that $t_{n-1} \le u \le t_n$,

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| = O\left(\sum_{j=n}^{+\infty} \gamma_j^2\right).$$

Under (A2.3.1) and (A2.3.3), γ varies regularly with exponent $(-\alpha)$, $\alpha > 1/2$. Thus $t \mapsto \int_0^t \gamma^2(s)\,ds$ varies regularly with exponent $1-2\alpha < 0$. Since v varies slowly, we have $\int_t^{+\infty} \gamma^2(s)\,ds = O([v(t)]^{-1})$. This property is still true under (A2.3.2). Finally, $|P_U(\Gamma) - P_Y(\Gamma)| = O([v(n)]^{-1})$, and

$$\left|P_U(\Gamma) - P_Y(\Gamma)\right| = O\left(\frac{1}{a(u)}\right), \tag{3.20}$$

and this upper bound is independent of x.

Step 2 (Following the annealing diffusion process). Coming back to assumptions (A2.1) to (A2.3), let ϕ be a continuously differentiable function with compact support; applying (3.16), there exists a compact set K_1, which contains the support of ϕ, and such that

$$P(U_t \notin K_1) = O\left(\frac{1}{a(t)}\right). \tag{3.21}$$

We may take in (3.17) C such that $K_1 \subset \{x; \|x\| \le C\}$.

Set $C_2 = C + 2\sup_{x\in K_1} \|x\|$, and let $\widetilde{V}$ be a twice continuously differentiable function, which equals V on the Euclidian sphere with radius C_2, and such that $\nabla\widetilde{V}$ is bounded. Let $\widetilde{Y}^{(t)}$ and $\widetilde{U}^{(t)}$ be defined by

$$\begin{aligned}
\widetilde{Y}_0^{(t)} &= \widetilde{U}_0^{(t)} = x, \qquad x \in K_1,\\
d\widetilde{Y}_u^{(t)} &= -[a(t+u)]\nabla\widetilde{V}(\widetilde{Y}_u^{(t)})\,dt + dB_u^{(t)},\\
d\widetilde{U}_u^{(t)} &= [-a(t+u)\nabla\widetilde{V}(\widetilde{U}_u^{(t)}) + h_{t+u}]\,dt + dB_u^{(t)}.
\end{aligned}$$

Applying (3.17) and (3.18) to $Y^{(t)}$, we have, uniformly for $x \in K_1$,

$$P\left(\sup_{0\le s\le\tau(t)} \|Y_s^{(t)}\| \ge C | Y_t = x\right) \le \exp[-\rho_1 a(t)],$$

$$\left|E\left(\phi(Y_{\tau(t)}^{(t)}) | Y_t = x\right) - G_{1/2a(t+\tau(t))}(\phi)\right| \le \exp[-\rho_1 a(t)],$$

from which we deduce that, uniformly for $x \in K_1$,

$$P\left(\sup_{0\le s\le\tau(t)} \|\widetilde{Y}_s^{(t)}\| \ge C_2 | \widetilde{Y}_t = x\right) \le \exp[-\rho_1 a(t)],$$

$$\left|E\left(\phi(\widetilde{Y}_{\tau(t)}^{(t)}) | \widetilde{Y}_t = x\right) - G_{1/2a(t+\tau(t))}(\phi)\right| \le \exp[-\rho_1 a(t)].$$

Then, applying (3.20), we have, uniformly for $x \in K_1$,

$$P\Big(\sup_{0\le s\le\tau(t)} \|\widetilde{U}_s^{(t)}\| \ge C_2 | \widetilde{U}_t = x\Big) = O\Big(\frac{1}{a(t)}\Big),$$

$$\big|E(\phi(\widetilde{U}_{t+\tau(t)}) | \widetilde{U}_0^{(\tau)} = x) - G_{1/2a(t+\tau(t))}(\phi)\big| = O\Big(\frac{1}{a(t)}\Big).$$

Thus, almost surely,

$$P\Big(\sup_{0\le s\le\tau(t)} \|U_s^{(t)} - \widetilde{U}_s^{(t)}\| > 0 | U_t = \widetilde{U}_t \in K_1\Big) = O\Big(\frac{1}{a(t)}\Big),$$

$$|E(\phi(U_{t+\tau(t)}) | U_t \in K_1) - G_{1/2a(t+\tau(t))}(\phi)| = O\Big(\frac{1}{a(t)}\Big).$$

Applying (3.21), we deduce that

$$|E(\phi(U_{t+\tau(t)})) - G_{1/2a(t+\tau(t))}(\phi)| = O\Big(\frac{1}{a(t)}\Big). \tag{3.22}$$

Since $\tau(t) = O(t^{2/3})$, we finally deduce that

$$\big|E(\phi(U_t)) - G_{1/2a(t)}(\phi)\big| = O\Big(\frac{1}{a(t)}\Big), \tag{3.23}$$

and this result is uniform for the functions ϕ whose support is included in a same compact and such that $\sup_{z\in\mathbb{R}^d} |\phi(z)| \le 1$.

3.5.4. *Proof of the first assertion of Theorem* 2. For any $r > 0$, let us take $R > r$ such that (3.16) is fulfilled, and an arbitrary $r_1 \in]0, r[$. We consider a positive function λ defined on $\mathbb{R}^+$, and for any $t > 0$ a continuous function ϕ_t such that

$$1_{\{r\lambda(t)\le V<R\}} \le \phi_t \le 1_{\{r_1\lambda(t)\le V<2R\}}.$$

Applying (3.23) with $\phi = \phi_t$ yields

$$P(\{V(U_t) \ge r\lambda(t)\}) \le G_{1/2a(t)}(V \ge r_1\lambda(t)) + O\Big(\frac{1}{a(t)}\Big),$$

$$P(\{V(U_t) \ge r_1\lambda(t)\}) \ge G_{1/2a(t)}(V \ge r\lambda(t)) + O\Big(\frac{1}{a(t)}\Big).$$

Thus, r_1 being arbitrary $(0 < r_1 < r)$,

$$P(\{V(U_t) \ge r\lambda(t)\}) = G_{1/2a(t)}(V \ge r\lambda(t)) + O\Big(\frac{1}{a(t)}\Big). \tag{3.24}$$

If we take $\lambda(t) = 1\ \forall\, t \in \mathbb{R}^+$ [(3.24) will also be helpful for proving Theorem 3], we have

$$P(\{V(U_t) \ge r\}) = G_{1/2a(t)}(V \ge r) + O\Big(\frac{1}{a(t)}\Big).$$

Now assumption (A2.1) implies that, for all $r > 0$, $\lim_{T\to 0} T\ln(G_T(V \geq r)) = -r$. Thus

$$P(\{V(U_t) \geq r\}) = O\left(\frac{1}{a(t)}\right) \quad \text{and} \quad P(\{V(Z_n) \geq r\}) = O\left(\frac{1}{v(n)}\right).$$

On the other hand, we have

$$E\big(V(Z_n)1_{\{V(Z_n)\geq r\}}\big) \leq R \cdot P(\{V(Z_n) \geq r\}) + E\big(V(Z_n)1_{\{V(Z_n)\geq R\}}\big).$$

Finally, by (3.15), the first part of Theorem 2 is proved.

3.5.5. *Proof of the second assertion of Theorem* 2. Under the additional assumption, there exists $r > 0$ such that $V(z) < r$ implies $\|\nabla V(z)\|^2 \geq \rho_V V(z)$, and we can write (2.4) as

$$Z_{n+1} = Z_n + \gamma_n h(Z_n) + \gamma_n r_{n+1} + \sigma_n \xi_{n+1},$$

with

$$h(z) = -\nabla V(z)1_{\{V(z)<r\}},$$

$$r_{n+1} = -\nabla V(Z_n)1_{\{V(Z_n)\geq r\}}.$$

Following the proof of the second part of Proposition 6 (Section 3.2.2), we show that for any A, $0 < A < \rho$, there exists n_0, such that for all $n \geq n_0$,

$$E(V(Z_{n+1})) \leq (1 - A\gamma_n)E(V(Z_n)) + O\left(\frac{\gamma_n}{v(n)}\right),$$

and finally

$$\sup_n E(v(n)V(Z_n)) < \infty,$$

which completes the proof of Theorem 2. □

3.6. *Proof of Theorem* 3.

3.6.1. *Weak convergence.* Let H_a be a random variable with distribution $G_{1/2a}$. Since g varies regularly with exponent $(-\eta)$, the Laplace transform of $4aV(H_a)$, $t \mapsto g(2a(2t+1))/g(2a)$, converges as $a \to \infty$ to the Laplace transform of $\gamma(\eta, 1/2)$, $t \mapsto (2t+1)^{-\eta}$. Thus, $4aV(H_a) \Rightarrow \gamma(\eta, 1/2)$ as $a \to +\infty$ and, for any $r > 0$,

$$G_{1/2a}[4aV \geq r] = P[4aV(H_a) \geq r] \to \gamma(\eta, \tfrac{1}{2})([r, +\infty[) \qquad \text{when } a \to \infty.$$

Applying (3.24) with $\lambda(t) = 1/4a(t)$, we have

$$P(4a(t)V(U_t) \geq r) = G_{1/2a(t)}(4a(t)V \geq r) + O\left(\frac{1}{a(t)}\right),$$

$$P[4a(t)V(U_t) \geq r] \to \gamma(\eta, \tfrac{1}{2})([r, +\infty[).$$

Thus,

$$4a(t)V(U_t) \Rightarrow \gamma\big(\eta, \tfrac{1}{2}\big) \quad \text{and} \quad 4v(n)V(Z_n) \Rightarrow \gamma\big(\eta, \tfrac{1}{2}\big).$$

3.6.2. *Small deviations.* Applying (3.24) with $\lambda(t) = f(a(t))[\ln a(t)]/a(t)$, we have

$$P\left[V(U_t) \geq \frac{rf(a(t))\ln a(t)}{a(t)}\right] = G_{1/2a(t)}\left[V \geq \frac{rf(a(t))\ln a(t)}{a(t)}\right] + O\left(\frac{1}{a(t)}\right).$$

Thus,

$$\ln\left(P\left[V(U_t) \geq \frac{rf(a(t))\ln a(t)}{a(t)}\right]\right)$$
$$= -\ln g(2a(t)) + \ln \int_{V \geq r\lambda(t)} \exp(-2a(t)V(x))\,dx + O\left(\frac{1}{a(t)}\right),$$

$$\ln\left(P\left[V(U_t) \geq \frac{rf(a(t))\ln a(t)}{a(t)}\right]\right) \sim \eta \ln a(t) - 2rf(a(t))\ln a(t) + O\left(\frac{1}{a(t)}\right),$$

from which we deduce that

$$\frac{1}{f(a(t))\ln a(t)} \ln\left(P\left[V(U_t) \geq \frac{rf(a(t))\ln a(t)}{a(t)}\right]\right) \to -2r,$$
$$\frac{1}{f(v(n))\ln v(n)} \ln\left(P\left[V(Z_n) \geq \frac{rf(v(n))\ln v(n)}{v(n)}\right]\right) \to -2r.$$

3.7. *Proof of Theorem* 4. We take up the notations introduced in Section 3.5.3. To establish Theorem 4, we have to prove the convergence of

$$E\left[\phi\left(U_t, \sqrt{a(t)}\nabla V(U_t)\right)\right] \quad \text{as } t \to +\infty,$$

which amounts to proving the convergence of

$$E\left[\phi\left(U_{t+\alpha(t)}, \sqrt{a(t+\alpha(t))}\nabla V(U_{t+\alpha(t)})\right)\right] \quad \text{as } t \to +\infty,$$

or to proving this convergence given $U_t \in K_1$. Thus, we can assume that the second- and third-order derivatives of V are bounded, in the same way as in the proof of Theorem 2.

Let $W_{n+1} = \sqrt{v(n)}\nabla V(Z_{n+1})$. We have

$$W_{n+1} = W_n - \gamma_n D^2 V(Z_n) W_n + \gamma_n r_{n+1} + \sqrt{\gamma_n} D^2 V(Z_n)\xi_{n+1}$$

with $E(\|r_{n+1}\|^2) \to 0$.

We apply the method of the stochastic differential equation introduced in Section 3.1.2 to the couple (Z_n, W_n). We come back here to the interpolation used in 3.1.2 and we define (X_t, Y_t) if $t \in [s_{n-1}, s_n]$ by

$$X_t = Z_n - (t - s_{n-1})\nabla V(Z_n) + \frac{\sqrt{t - s_{n-1}}}{\sqrt{v(n)}}\xi_{n+1} \tag{3.25}$$

and

$$\begin{aligned} Y_t = W_n &- (t - s_{n-1})D^2V(Z_n)W_n + (t - s_{n-1})r_{n+1} \\ &+ \sqrt{t - s_{n-1}}D^2V(Z_n)\xi_{n+1} \end{aligned} \tag{3.26}$$

with $s_n = \sum_{k=0}^{n} \gamma_k$.

We define the family of processes $(X^{(u)}, Y^{(u)})_{u\geq 0}$ by

$$(X_t^{(u)}, Y_t^{(u)}) = (X_{u+t}, Y_{u+t}). \tag{3.27}$$

Theorem 2 implies that $\sup_n E(\|W_n\|^2) < \infty$. Consequently, (X_t, Y_t) is tight and $\sup_n E(\|Y_t\|^2) < \infty$. Then $(X^{(u)}, Y^{(u)})_{u\geq 0}$ is tight, and any closure point for the weak convergence is a solution of the stochastic differential equation

$$dP_t = -\nabla V(P_t)\,dt, \tag{3.28}$$

$$dR_t = -D^2V(P_t)R_t\,dt + D^2V(P_t)\,dB_t, \tag{3.29}$$

with (P_0, R_0) independent of (B_t).

However, $(X_t) \Rightarrow G_0$, thus $(X^{(u)})$ converges weakly to (P_t) defined by $P_t = P_0$, P_0 with distribution G_0. Given $P_0 = z^* \in \operatorname{Argmin} V$, the solution of (3.29) is a geometrically recurrent linear diffusion with stationary distribution $\nu_{z^*} = \mathcal{N}(0, \frac{1}{2}D^2V(z^*))$.

For all continuous and bounded function ϕ defined on $\mathbb{R}^{2d}$, for all $t > 0$,

$$|E[\phi(P_t, R_t)|P_0 = z^*,\ R_0 = r] - \int \phi(z^*, y)\,d\nu_{z^*}(y)| = O([\rho(z^*)]^t(1 + \|r\|^2))$$

with $\rho(z^*) < 1$.

Let μ, probability on $\mathbb{R}^{2d}$, be a closure point for the weak convergence of (X_t, Y_t). The first marginal distribution of μ is G_0, and the second has a bounded moment of order 2. Let us consider a sequence $(u(n))$, increasing to infinity, such that $(X_{u(n)}, Y_{u(n)}) \Rightarrow \mu$.

Let ϕ be a continuous and bounded function, and set $t > 0$. By the tightness of $(X^{(u)}, Y^{(u)})$, there exists a subsequence of $(u(n))$, denoted by $(w(n))$, such that $(X^{(w(n)-t)}, Y^{(w(n)-t)})$ converges weakly to the diffusion solution of (3.28) and (3.29). Moreover,

$$\left|E[\phi(P_t, R_t)] - \sum_{z^*\in\operatorname{Argmin} V} G_0(z^*)\int \phi(z^*, y)\,d\nu_{z^*}(y)\right| = O\Big(\sup_{z^*\in\operatorname{Argmin} V}\{\rho(z^*)\}^t\Big).$$

Thus $\mu = \sum_{z^*\in\operatorname{Argmin} V} G_0(z^*)\delta_{z^*}\otimes\nu_{z^*}$, which completes the proof of Theorem 4.

3.8. *Proof of Theorem* 5.

3.8.1. *Preliminary.* In order to prove Theorem 5, we first establish convergence rates for the simulated annealing algorithm

$$\widehat{Z}_{n+1} = \widehat{Z}_n - \gamma_n[\nabla V(\widehat{Z}_n) + R_{n+1}] + \sigma_n \xi_{n+1},$$

with strong assumptions on the disturbance (R_n). Then, we show in Section 3.7.2 how Theorem 5 may be proved by using this preliminary result.

On a probability space $(\Omega, \mathscr{A}, P)$ equipped with a filtration $\mathscr{F} = (\mathscr{F}_n)_{n\geq 0}$, let $(\widehat{Z}_n)$ be defined by

$$\widehat{Z}_{n+1} = \widehat{Z}_n - \gamma_n[\nabla V(\widehat{Z}_n) + R_{n+1}] + \sigma_n \xi_{n+1}. \tag{3.30}$$

We make the following assumptions.

(AA.1) (R_n) and (ξ_n) are adapted to $\mathscr{F}$ and R_{n+1} and ξ_{n+1} are independent conditionally on $\mathscr{F}_n$.

(AA.2) ξ_{n+1} is independent of $\mathscr{F}_n$ and has the distribution $\mathscr{N}(0, I)$.

(AA.3) $E(\|R_{n+1}\|^2|\mathscr{F}_n) \leq n^{-\delta}L$, where L is an almost surely finite random variable and $\delta > 1 - \alpha$.

LEMMA 8. *Under assumptions* (A2.1) *to* (A2.3), *and* (AA.1) *to* (AA.3), *Theorems* 2 *and* 3 *can be applied to* $(\widehat{Z}_n)$. *Under the additional assumption* (A2.4), *Theorem* 4 *can be applied to* $(\widehat{Z}_n)$.

PROOF.

Step 1. For any $A > 0$, set $\overline{R}_n = R_n 1_{\{\|R_n\| \leq \sqrt{An^{-\delta}}\}}$, and

$$\overline{Z}_{n+1} = \overline{Z}_n - \gamma_n[\nabla V(\overline{Z}_n) + \overline{R}_{n+1}] + \sigma_n \xi_{n+1}.$$

Then $(\overline{Z}_n)$ and $(\widehat{Z}_n)$ agree on $\{L \leq A\}$. L being almost surely finite, $P(L > A) \to 0$ as $A \to +\infty$; therefore, it is enough to prove the lemma assuming L deterministic.

Step 2. We define $(Z_n)_{n\geq N}$ by $Z_N = \widehat{Z}_N$ and

$$Z_{n+1} = Z_n - \gamma_n \nabla V(Z_n) + \sigma_n \xi_{n+1} \quad \text{for } n \geq N.$$

For $i < j$, let us denote by $\Pi^{i,j}$ and $\widehat{\Pi}^{i,j}$ the distributions of Z_j and $\widehat{Z}_j$ given $\mathscr{F}_i$, respectively.

The density of $\Pi^{n,n+1}$ with respect to Lebesgue measure is

$$z_{n+1} \mapsto \exp\left[-\frac{\|z_{n+1} - Z_n + \gamma_n \nabla V(Z_n)\|^2}{2\sigma_n^2}\right],$$

and the density of $\widehat{\Pi}^{n,n+1}$ is

$$z_{n+1} \mapsto E\left(\exp\left[-\frac{\|z_{n+1} - \widehat{Z}_n + \gamma_n \nabla V(\widehat{Z}_n) + \gamma_n R_{n+1}\|^2}{2\sigma_n^2}\right]\middle|\mathscr{F}_n\right).$$

Let K be the Kullback information. Since ξ_{n+1} and R_{n+1} are independent given $\mathscr{F}_n$, we have, given $Z_n = \widehat{Z}_n$,

$$K\big[\Pi^{n,n+1}, \widehat{\Pi}^{n,n+1}\big] = E\bigg[\frac{1}{2\sigma_n^2}\big(2\langle\gamma_n R_{n+1}, \sigma_n\xi_{n+1}\rangle + \gamma_n^2\|R_{n+1}\|^2\big)\Big|\mathscr{F}_n\bigg],$$

$$K\big[\Pi^{n,n+1}, \widehat{\Pi}^{n,n+1}\big] = \frac{\gamma_n^2}{2\sigma_n^2}E(\|R_{n+1}\|^2|\mathscr{F}_n) \le \frac{\gamma_n n^{-\delta}v(n)}{2}L.$$

Given $Z_N = \widehat{Z}_N$,

$$K\big[\Pi^{N,N+n}, \widehat{\Pi}^{N,N+n}\big] \le L\bigg(\sum_{j=N}^{N+n}\frac{\gamma_j j^{-\delta}v(j)}{2}\bigg).$$

Therefore, $\|\cdot\|_{\mathrm{var}}$ being the total variation of a measure, given $Z_N = \widehat{Z}_N$,

$$\big\|\Pi^{N,\,N+n} - \widehat{\Pi}^{N,\,N+n}\big\|_{\mathrm{var}} = O\bigg(\bigg(\sum_{j=N}^{N+n}\frac{\gamma_j j^{-\delta}v(j)}{2}\bigg)^{1/2}\bigg).$$

Since $t \mapsto t^{-\delta}\gamma(t)v(t)$ varies regularly with exponent $-(\alpha+\delta)$, $\delta > 1-\alpha$,

$$\bigg(\sum_{j=N}^{+\infty} j^{-\delta}\gamma_j v(j)\bigg)^{1/2} = O\bigg(\frac{1}{v(N)}\bigg).$$

Let ϕ be a continuous function with compact support. In view of the inequality (3.22) of the proof of Theorem 2, uniformly for $Z_n = \widehat{Z}_n = z$ in a compact set K,

$$\big|E\big(\phi(U_{t_n+\tau(t_n)})\big) - G_{1/2a(t_n+\tau(t_n))}(\phi)\big| = O\bigg(\frac{1}{v(n)}\bigg).$$

Let β be an increasing function such that $t_n + \tau(t_n) \in [t_{n+\beta(n)-1}, t_{n+\beta(n)}[$. We have then

$$\begin{aligned}&\big|E\big(\phi(Z_{n+\beta(n)})\big) - G_{1/2a(t_n+\tau(t_n))}(\phi)\big|\\ &\qquad\le O\bigg(\frac{1}{v(n)}\bigg) + \big|E\big(\phi(Z_{n+\beta(n)})\big) - E\big(\phi(U_{t_n+\tau(t_n)})\big)\big|;\end{aligned}$$

$$\big|E\big(\phi(Z_{n+\beta(n)})\big) - G_{1/2a(t_n+\tau(t_n))}(\phi)\big| = O\bigg(\frac{1}{v(n)}\bigg).$$

We deduce that, given $\widehat{Z}_N = Z_N = z$, uniformly for $z \in K$,

$$\big|E\big(\phi(\widehat{Z}_{N+\beta(N)})\big) - G_{1/2a(t_N+\tau(t_N))}(\phi)\big| = O\bigg(\frac{1}{v(N)}\bigg).$$

Step 3. On the other hand, $E(\|R_{n+1}\|^2) \le Ln^{-\delta}$, thus $\sup_n E(v(n)\|R_{n+1}\|^2) < \infty$. Proposition 6 may be applied. Therefore $\sup_n E(V(\widehat{Z}_n)) < \infty$, and there exists $R > 0$ such that $E(V(\widehat{Z}_n)1_{\{V(\widehat{Z}_n)\ge R\}}) = O(\frac{1}{v(n)})$.

Then, we may conclude by following the end of the proofs of Theorems 2, 3 and 4. □

3.8.2. *Proof of Theorem* 5. We use the averaging method of the disturbance (ε_n) introduced by Walk [28] or by Hwang and Sheu [17].

For $N \geq 1$, let us define $(\overline{\varepsilon}_n)_{n\geq N}$ by $\overline{\varepsilon}_N = 0$ and, for $n \geq N$,

$$\overline{\varepsilon}_{n+1} = \zeta_n \sum_{j=N}^{n} \frac{\gamma_j}{\zeta_j} \varepsilon_{j+1},$$

where $\zeta_n = \prod_{j=N}^{n}(1-\gamma_j)$. Let $(\overline{Z}_n)_{n\geq N}$ be defined by $\overline{Z}_N = Z_N$ and, for $n \geq N$,

$$\overline{Z}_{n+1} = \overline{Z}_n - \gamma_n\big[\nabla V(Z_n) + \overline{\varepsilon}_n + r_{n+1}\big] + \sigma_n \xi_{n+1}.$$

With ∇V being globally Lipschitz, $\nabla V(Z_n) = \nabla V(\overline{Z}_n) + O(\|\overline{Z}_n - Z_n\|)$, and

$$\overline{Z}_{n+1} = \overline{Z}_n - \gamma_n[\nabla V(\overline{Z}_n) + R_{n+1}] + \sigma_n \xi_{n+1},$$

with $R_{n+1} = \overline{\varepsilon}_n + r_{n+1} + O(\|\overline{Z}_n - Z_n\|)$.

Thus, it is sufficient to prove that Lemma 5 can be applied to $(\overline{Z}_n)$ and that $\overline{Z}_n - Z_n = o([v(n)]^{-1})$ almost surely in order to establish Theorem 4.

However, $\overline{Z}_{n+1} - Z_{n+1} = \overline{Z}_n - Z_n + \gamma_n \overline{\varepsilon}_n - \gamma_n \varepsilon_{n+1}$, thus $\overline{Z}_{n+1} - Z_{n+1} = -\overline{\varepsilon}_{n+1}$. It is finally enough to prove that there exists $\delta > 1 - \alpha$ such that

$$\overline{\varepsilon}_n^2 = O(n^{-\delta}) \quad \text{a.s.}$$

From Chow's theorem (see, e.g., [14]), for all $\beta > 0$,

$$\|\overline{\varepsilon}_n\|^2 = O\left[\zeta_n^2\left[\sum_{j=N}^{n} \frac{\gamma_j^2}{\zeta_j^2}\right]\left(\ln\left[\sum_{j=N}^{n} \frac{\gamma_j^2}{\zeta_j^2}\right]\right)^{1+\beta}\right] \quad \text{a.s.}$$

Under (A3.3.1): Let $G(t) = \int_N^t \gamma(s)\,ds$. We have $\zeta_n \sim \exp(-G(n))$ and

$$\sum_{j=N}^{n} \frac{\gamma_j^2}{\zeta_j^2} \sim \int_N^n \gamma^2(s) e^{2G(s)}\,ds \sim \int_{G(N)}^{G(n)} \gamma(G^{-1}(s)) e^{2s}\,ds \sim \gamma_n e^{2G(n)}.$$

Thus,

$$\zeta_n^2\left[\sum_{j=N}^{n} \frac{\gamma_j^2}{\zeta_j^2}\right]\left(\ln\left[\sum_{j=N}^{n} \frac{\gamma_j^2}{\zeta_j^2}\right]\right)^{1+\beta} \sim \gamma_n\big[\ln\big(\gamma_n e^{2G(n)}\big)\big]^{1+\beta} \sim \gamma_n[G(n)]^{1+\beta},$$

and $t \mapsto \gamma(t)[G(t)]^{1+\beta}$ varies regularly with exponent $\delta = -\alpha + (1-\alpha)(1+\beta)$, for an arbitrary $\beta > 0$.

Since $\alpha > \frac{2}{3}$, there exists $\beta > 0$ such that $\delta > 1 - \alpha$ and $\|\overline{\varepsilon}_n\|^2 = O(n^{-\delta})$ almost surely.

Under (A3.3.2): since $\gamma_n = \gamma_0/n$, we have $\sum_{k=N}^{n} \gamma_k = O(\ln n)$, $\zeta_n = O(n^{-1})$ and $\sum_{k=N}^{n}(\gamma_k^2/\zeta_k^2) = O(n)$.

Consequently, $\forall\ \beta > 0$,

$$\|\overline{\varepsilon}_n\|^2 = O\left[\frac{1}{n}[\ln n]^{1+\beta}\right] \quad \text{a.s.}$$

Thus $\|\overline{\varepsilon}_n\|^2 = O(n^{-\delta})$ a.s. for all δ such that $0 < \delta < 1$. □

Acknowledgments. I gratefully acknowledge the constant help of Marie Duflo. I also thank the referees and the editors for their thoughtful comments.

REFERENCES

[1] BENAIM, M. (1996). A dynamical system approach to stochastic approximation. *SIAM J. Control Optim.* **34** 437–472.

[2] BENVENISTE, A., METIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximation.* Springer, New York.

[3] BOUTON, C. (1988). Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne. *Ann. Inst. H. Poincaré* **24** 131–155.

[4] CHEN, H. F., GUO, L. and GAO, A. J. (1988). Convergence and robustness of the Robbins–Monro algorithm truncated at randomly varying bounds. *Stochastic Process. Appl.* **27** 217–231.

[5] 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.

[6] DUFLO, M. (1996). *Algorithmes Stochastiques. Collection Mathématiques et Applications*. Springer, New York.

[7] DUPUIS, P. and KUSHNER, H. J. (1989). Stochastic approximation and large deviations: upper bounds and w.p. 1 convergence. *SIAM J. Control Optim.* **27** 1108–1135.

[8] FEIGIN, P. D. (1985). Stable convergence of semimartingales. *Stochastic Process. Appl.* **19** 125–134.

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

[10] FORT, J. C. and PAGES, G. (1996). Convergence of stochastic algorithms: from the Kushner–Clark theorem to the Lyapunov functional method. *Adv. in Appl. Probab.* **28** 1072–1094.

[11] GELFAND, S. B. and MITTER, S. K. (1991). Recursive stochastic algorithms for global optimization in $\mathbb{R}^d$. *SIAM J. Control Optim.* **29** 999–1018.

[12] GELFAND, S. B. and MITTER, S. K. (1993). Metropolis-type annealing algorithms for global optimization in $\mathbb{R}^d$. *SIAM J. Control Optim.* **31** 111–131.

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

[14] HALL, P. and HEYDE, C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

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

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

[17] HWANG, C. R and SHEU, S. J. (1990). On the behavior of a stochastic algorithm with annealing. Technical report, Academia Sinica, Taiwan.

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

[19] KANIOVSKI, YU. (1988). Limit theorems for processes of stochastic approximation when the regression function has several roots. *Kibernetika* **2** 136–138. (In Russian, 120–122; English trans.)

[20] KANIOVSKI, YU. and PFLUG, G. (1994). Non-standard limit theorems for urn models and stochastic approximation procedures. Preprint.

[21] KERSTING, G. D. (1978). A weak convergence theorem with application to the Robbins–Monro process. *Ann. Probab.* **6** 1015–1025.

[22] KIEFER, J. and WOLFOWITZ, J. (1952). Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist.* **23** 462–466.

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

[24] KUSHNER, H. J. and CLARK, D. S. (1978). Stochastic approximation for constrained and unconstrained systems. *Appl. Math. Sci.* **26**. Springer, New York.

[25] KUSHNER, H. J. and HUANG, H. (1979). Rates of convergence for stochastic approximation type algorithms. *SIAM J. Control Optim.* **17** 607–617.

[26] KUSHNER, H. J. and HUANG, H. (1981). Asymptotic properties of stochastic approximations with constant coefficients. *SIAM J. Control Optim.* **19** 87–105.

[27] LAI, T. Z. and WEI, C. Z. (1983). A note on martingale difference sequences satisfying the local Marcinkiewicz–Zigmund condition. *Bull. Institute Math. Acad. Sinica* **11** 1–13.

[28] LJUNG, L., PFLUG, G. and WALK, H. (1992). *Stochastic Approximation and Optimization of Random Systems.* Birkhäuser, Boston.

[29] MARQUEZ, D. (1997). Convergence rates for annealing diffusion processes. *Ann. Appl. Probab.* **7** 1118–1139.

[30] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III: Foster–Lyapunov criteria for continuous time processes with examples. *Adv. in Appl. Probab.* **25** 518–548.

[31] NEVEL'SON, M. B. and HA'SMINSKII, R. Z. (1972). Stochastic approximation and recursive estimation. *Transl. Math. Monogr.* **47**. (In Russian, Nauka.)

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

[33] SENETA, E. (1976). *Regularly Varying functions. Lecture Notes in Math.* **508**. Springer, Berlin.

[34] TOUATI, A. (1993). Two theorems on convergence in distribution for stochastic integrals and statistical applications. *Theory Probab. Appl.* **38** 95–117.

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: pelletier@math.uvsq.fr