ON THE CONVERGENCE OF MARKOVIAN STOCHASTIC ALGORITHMS WITH RAPIDLY DECREASING ERGODICITY RATES

LAURENT YOUNES

ABSTRACT. We analyse the convergence of stochastic algorithms with Markovian noise when the ergodicity of the Markov chain governing the noise rapidly decreases as the control parameter tends to infinity. In such a case, there may be a positive probability of divergence of the algorithm in the classic Robbins-Monro form. We provide modifications of the algorithm which ensure convergence. Moreover, we analyse the asymptotic behaviour of these algorithms and state a diffusion approximation theorem.

1. INTRODUCTION

Stochastic algorithms of Robbins-Monro type with Markovian noise form a category of processes for which almost sure convergence cannot be obtained in general. The reason is that the ergodicity of the Markov chain governing the noise may decrease when the control parameter tends to infinity, and trap the algorithm within an exploding regime. In this paper, we study rigorously a natural strategy in which more time is spent for estimating the variations of the control parameter for large values of this parameter. In particular, we give conditions under which this strategy converges.

We consider the following framework. Let Ω be a probability space, and $(\pi_\theta, \theta \in \mathbb{R}^d)$ a family of probability distributions on Ω indexed by a parameter θ. Let also, for all θ, $f(\theta, .) = f_\theta(.)$ be a function defined on Ω with values in $\mathbb{R}^d$, and :

$$h(\theta) = \int_\Omega f(\theta, x)\pi_\theta(dx)\,. \tag{1}$$

The problem is to solve the equation $h(\theta) = 0$.

For this purpose (especially when $h(\theta)$ is the gradient of a function L to maximize), one may follow the dynamical system

$$\dot{\theta} = h(\theta) \tag{2}$$

with discretized form

$$\theta_{n+1} = \theta_n + \gamma h(\theta_n)\,, \tag{3}$$

where γ is the time discretization step. Stochastic algorithms for the solution of (1) are needed when the function h cannot (for numerical reasons, or because π_θ is unknown) be efficiently computed, but when samples from π_θ can easily be obtained (by simulation or experiments). Robbins and Monro ([18]) considered the case when at step n of the algorithm (3), a sample X^{n+1} of the distribution π_{θ_n} is given, and propose an updating rule of the kind

$$\theta_{n+1} = \theta_n + \gamma_{n+1} f(\theta_n, X^{n+1})\,. \tag{4}$$

It is necessary, in this case, to use a time discretization which depends on n. Under quite general conditions on γ_n, f and π_θ, it can be shown that this procedure almost surely converges to an asymptotically stable point of (2) when such a point exists.

Date:
AMS CLASSIFICATION: PRIMARY 65U05, 62F10, SECONDARY 62M30,60K35.
KEYWORDS: STOCHASTIC APPROXIMATION - MONTE CARLO MARKOV CHAINS - MAXIMUM OF LIKELIHOOD ESTIMATION - DIFFUSION APPROXIMATION.

However, there can be situations when a sample of π_{θ_n} cannot be obtained at each step of the procedure. This happens when π_θ cannot be directly simulated, generally for the same numerical reasons for which $h(\theta)$ could not be computed. Instead (and this is the case we consider here), a feasible dynamical simulation procedure may be available for π_θ, which means that it is possible to construct a Markov chain on Ω, associated to some transition probability P_θ, which is ergodic and converges in distribution to π_θ. In such a case, one can use a procedure in which the updating of θ_n is the same as in (4), but in which the sample X^{n+1} is no more assumed to be drawn from π_{θ_n}, but from the distribution $P_{\theta_n}(X^n, .)$. Even in the case when (2) admits an asymptotically stable point, convergence cannot be obtained in general if the ergodicity of the Markov chain governed by P_θ decreases too fast when θ tends to infinity. Convergence results are provided in [1] under hypotheses which more or less assume that such a phenomenon does not occur.

Note that a necessary condition under which the sequence γ_n is a time discretization of $[0, +\infty[$ is

$$\sum_n \gamma_n = +\infty\,.$$

A typical condition needed for almost sure convergence (for example in the Robbins-Monro case) is $\sum_n \gamma_n^\alpha < \infty$ for some $\alpha > 1$. But, in the present case, one can exhibit situations (cf. paragraph 6.3) for which divergence may occur even with $\gamma_n = a/n$, if the constant a is too large. In fact, it is shown in [21] (in a special context) that a sufficient condition for almost sure convergence is $\gamma_n = a/n$ with a small enough. An important ingredient for this result was the fact that $f(\theta, .)$ was bounded, which yielded useful *a priori* bounds on the norm of θ_n. However since these bounds were *a priori*, they necessarily were very rough, and the value of the constant a which was provided in [21] was too small to be used in practice. In fact, good practical results have been obtained with larger values of a.

Note that almost sure convergence results are often given under the additional hypothesis of boundedness of the sequence (θ_n). In practice, it is always possible to use a projection procedure on some compact set, which therefore ensures boundedness. However, the choice *a priori* of the compact set on which the projection is done is, in many cases, not obvious, especially when the dimension of θ_n is large. To know that almost sure boundedness is true even without projection is valuable, for both theoretical and practical reasons.

In this paper, we study a modification of (4) in which several iterations of the simulation procedures are performed before updating θ_n, the updating being done using the average of the values of $f(\theta_n, .)$ which have been obtained. Such an approach is very often used in practice, and our purpose is to check how this can help for convergence.

More precisely, we will study the following scheme. Let θ_n and X^n be the current parameter and state. Let k_n be an integer, which may depend on θ_n. We generate a Markov chain $Y^{n+1,0}, \ldots, Y^{n+1,k_n}$ in Ω by $Y^{n+1,0} = X^n$ and $\mathbf{P}(Y^{n+1,k} \in .\,|\, Y^{n+1,k-1} = x) = P_{\theta_n}(x, .)$. Finally, we define the recursion

$$\theta_{n+1} = \theta_n + \gamma_{n+1}\left[\frac{1}{k_n}\sum_{k=1}^{k_n} f(\theta_n, Y^{n+1,k})\right] \tag{5}$$

and

$$X^{n+1} = Y^{n+1,k_n}\,. \tag{6}$$

Such a scheme has been studied in [8], in the case where the transition $(\theta_n, Y^{n+1,k-1}) \to Y^{n+1,k}$ only depends on θ_n. In such a situation, almost sure convergence is true under mild hypothesis, and the problem is whether convergence can be accelerated.

This paper is organized as follows. In the next section, we describe some examples which fall into the framework we consider. We then list a series of hypotheses relative to the algorithm, and show that under these hypotheses, almost sure convergence is true. We end the paper by a theorem stating that the algorithm also has a Gaussian behavior when it approaches convergence, which can provide hints for designing numerical strategies in practice.

2. Examples

A typical example is the problem of the maximization of an expression of the kind

$$L(\theta) = \int_\Omega F(\theta, x)\pi_\theta(x)\nu(dx)$$

for some real valued function $F(\theta, .)$; ν is a fixed probability measure on Ω and $\pi_\theta > 0$ ν-almost everywhere. If F and $\log \pi_\theta$ are differentiable in θ (and if the interchange between derivative and integral is valid), one may set $h(\theta) = \nabla_\theta L(\theta)$, and in this case $f_\theta = \nabla_\theta F + F(\theta, .)\nabla_\theta \log \pi_\theta$.

Another case is when π_θ can be written as

$$\pi_\theta(x)\nu(dx) = [\exp(-\Lambda(\theta, x))/Z_\theta]\nu(dx)\,,$$

and the aim is to maximize $l(\theta) = \log \pi_\theta(x_0)$ for a fixed $x_0 \in \Omega$. One has in this case

$$h(\theta) := \nabla_\theta l(\theta) = \int_\Omega [\nabla_\theta \Lambda(\theta, x) - \nabla_\theta \Lambda(\theta, x_0)]\pi_\theta(dx)\,.$$

This corresponds to the maximum likelihood procedure in statistical inference. In the particular case when Λ is a linear function of θ (exponential model), the function f is independent of θ.

An important example of occurence of this situation is the case when π_θ is a Gibbs distribution on a configuration space $\Omega = F^S$, where F ("state space") and S are finite sets. Considering the model

$$\pi_\theta(x) = \frac{e^{-\langle \theta\,,\, H(x)\rangle}}{Z_\theta}\,, \tag{7}$$

H being some given function defined on Ω, with values in $\mathbf{R}^d$, and $\langle\, .\,,\, .\,\rangle$ standing for the Euclidean product in $\mathbf{R}^d$. The maximization of $\log \pi_\theta(x_0)$ leads to take

$$f(\theta, x) = f(x) = H(x) - H(x_0)\,.$$

In this context, Ω is a finite set, but with enormous cardinality, so large that any sum made over all elements of Ω (such as the one which is implicit in Z_θ, or in the computation of expectations) goes far beyond the ability of computers. The fact that $\pi_\theta(x)$ is not computable implies that a direct ("static") simulation of π_θ cannot be performed. However, there exist many dynamical simulation procedures for Gibbs distributions: Gibbs sampler, Metropolis algorithm and variants, clusters algorithms... (see [19], [10] and references within). All these procedures provide transition probabilities with the property that, for all distribution ν on Ω,

$$||\nu.P_\theta^n - \pi_\theta|| \le C_\theta \lambda_\theta^n$$

for some $\lambda_\theta < 1$. In other terms, the Markov chains associated to P_θ are Doeblin-ergodic. However when θ tends to infinity, the convergence rate λ_θ tends very rapidly to 1. In fact, for large θ, λ_θ can be taken equal to $1 - \exp(-K_\theta|\theta|)$, and sharp estimates of the value of K_θ (which is bounded with respect to θ) can even be provided in some cases (see [15], for example). This is a case of rapidly decreasing ergodicity rate (at exponential speed), of the kind we are dealing with in this paper. Note that, in this situation, a (non-optimal) value of λ_θ can always be explicitely computed (cf, for example [11]).

3. Hypotheses

3.1. Basic assumptions. The almost sure convergence result partially relies on some *a priori* estimates which may be done on the sequence (θ_n). These estimates are based on assumptions on the function $f(\theta, x)$, and we will separate two cases, in which the basic assumption is that f is uniformly bounded in θ, for the first case (case (U)), and that the growth of f is at most linear in θ in the second case (case (L)). All along this paper, we shall trace in parallel two sets of hypotheses relative to each of these cases. They will be refered as (U0),..., (U6) and (L0),...,(L5). We thus first let

Hypothesis (U0) *There exists a constant D_f such that, for all θ,*

$$\sup_x |f(\theta, x)| \leq D_f \, .$$

Hypothesis (L0) *There exists a constant D_f such that, for all θ,*

$$\sup_x |f(\theta, x)| \leq D_f(1 + |\theta|) \, .$$

Although the basic assumption (U0) implies (L0), the whole set of hypotheses related to case (L) do not boil down to the hypotheses of case (U), even when (U0) is true. Both cases yield different theorems, but with similar proofs, so that it is more convenient to carry them together.

3.2. Hypotheses on the transition probabilities P_θ.

3.2.1. *Case (L).* For a function W on Ω, with values in $\mathbb{R}$, we let ΔW be the maximal oscillation of W over Ω, ie

$$\Delta W = \max_{x,y \in \Omega} (W(x) - W(y)) \, . \tag{8}$$

We set $P_\theta W(x) = \int W(y) P_\theta(x, dy)$ and $\pi_\theta W = \int W(y) \pi_\theta(dy)$.

Hypothesis (L1)

a) We assume that, for all θ, a transition probability P_θ is given such that the associated Markov chain is ergodic, with limit distribution π_θ, and exponentially fast convergence ; more precisely, we assume that , for all θ, there exist a function $\zeta_e(\theta) \geq 1$ and a positive number $\lambda_\theta < 1$ such that, for all initial distributions ν and ν' on Ω, we have

$$||\nu P_\theta^n - \nu' P_\theta^n|| \leq \zeta_e(\theta) \lambda_\theta^n ||\nu - \nu'|| \tag{9}$$

where

i) *$\nu P_\theta(A) = \int_\Omega P_\theta(x, A)\nu(dx)$ and $\nu.P_\theta^n = (\nu P_\theta^{n-1}).P_\theta$.*

ii) *the norm used to compare probability measures is the norm in variation:*

$$||\nu - \nu'|| = \sup_A (\nu(A) - \nu'(A)) \, .$$

b) We also assume that for all W, $P_\theta W$ is a differentiable function of θ. More precisely, denote by $[\nabla P_\theta]$ the operator which associates to a function W on Ω the function $\nabla_\theta(P_\theta W)$. We assume that there exists a function $\zeta_P(\theta) \geq 1$ such that, for all W (such that ΔW is finite),

$$\sup_x |\, [\nabla P_\theta] W(x) \,| \leq \zeta_P(\theta) \Delta W \, , \tag{10}$$

Similarly, we let $[\nabla \pi_\theta]$ be the functional which associates to a function W on Ω the derivative of $\pi_\theta W$. We assume

$$|\, [\nabla \pi_\theta] W \,| \leq \zeta_P(\theta) \Delta W \, , \tag{11}$$

Equation (9) is equivalent to (cf. [7]): for all W,

$$\Delta(P_\theta^n W) \leq \zeta_e(\theta) \lambda_\theta^n \Delta W \tag{12}$$

3.2.2. *Case (U).* We need to be a little more specific concerning the expression of λ_θ in case (U).
Hypothesis (U1) *We assume (L1) and that there exist constants* D_λ *and* K_λ *such that,*

$$1 - \lambda_\theta \geq K_\lambda e^{-D_\lambda . |\theta|} . \tag{13}$$

3.3. **Additional hypotheses on $f(\theta, .)$ and $\pi(\theta, .)$. Hypothesis (U2-L2)** *We assume that, for all* $y \in \Omega$ *the functions* $\pi_\theta f_\theta$ *and* $P_\theta f_\theta(y)$ *are differentiable with respect to* θ *and that there exists a function* $\zeta_f \geq 1$ *such that*

$$\nabla_\theta(\pi_\theta f_\theta) \leq \zeta_f(\theta)$$
$$\nabla_\theta(P_\theta f_\theta(y)) \leq \zeta_f(\theta)$$

3.4. **Hypotheses on the sequence γ_p and k_p.** These hypotheses are the main differences between cases (U) and (L).

3.4.1. *Case (U).* **Hypothesis (U3)** *a)* $\gamma_p = a/p$ *for some* $a > 0$.
b) Let $c_l > 0$. *Let* $l_n = [\, n^{1+c_l} \,]$ *where* $[\,.\,]$ *holds for the integer part. Then, for all n,*

$$k_p \leq k_{p+1} , \text{ for } l_n < p < l_{n+1} .$$

c) k_p *is larger than the integer part of* $\kappa(\theta_p)$ *where* $\kappa(\theta) = \exp a'|\theta|$.
d) For some constants C *and* C'*, we have*

$$C'a' + C/a \geq 1 .$$

3.4.2. *Case (L). In case (L), we use loose hypotheses on* γ_p *and more stringent one on* k_p.

Hypothesis (L3)
a) (γ_n) *is decreasing,* $\sum_n \gamma_n = +\infty$ *and* $\sum_n \gamma_n^2 < +\infty$
b) There exist an increasing sequence $l_n, n \geq 0$ *of integers, and a sequence* $\overline{\gamma}_n, n \geq 0$ *such that*

$$\frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}} \leq \overline{\gamma}_p$$

unless there exists some n such that $p = l_n$. *Moreover, these sequences satisfy*

$$\sum_n \gamma_n \overline{\gamma}_n < \infty$$

and

$$\sum_n \gamma_{l_n} < \infty .$$

c) k_p *is larger than the integer part of* $\kappa(\theta_p)$ *where* $\kappa(\theta)$ *is a function which satisfies, for constants* $\mu > 0$ *and* $q_\kappa \geq 0$

$$\max_{|\theta' - \theta| \leq D_f(1+|\theta|)} \left[\zeta_\kappa(\theta')(1 - \lambda_{\theta'})^{-2}\right] \leq \mu . \kappa(\theta) , \tag{14}$$

and

$$|\nabla_\theta \log \kappa| \leq \mu , \tag{15}$$

for $\theta \in \mathbf{R}^d$, *with* $\zeta_\kappa = \zeta_e^2 \max(\zeta_P, \zeta_f)$.

Note that hypotheses b) and c) are true for the choices of γ_p and k_p which have been made in case (U), with the exception of condition (14). Note however that in both cases, we have, (θ_p) being any sequence generated by (5), for some constant $\tilde{\mu}$,

$$\max_{|\theta - \theta_p| \leq |\theta_{p+1} - \theta_p|} \left[\zeta_\kappa(\theta)(1 - \lambda_\theta)^{-2}\right] \leq \tilde{\mu} . \kappa(\theta_p) e^{(2D_\lambda - \delta)^+ |\theta_p|} , \tag{16}$$

with $\delta = a'$ in case (U), $\delta = 2D_\lambda$ in case (L), and $\zeta_\kappa(\theta) = 1$ in case (U). This comes directly from (14) in case (L), and from the fact that $|\theta_{p+1} - \theta_p|$ is uniformly bounded in case (U).

3.5. **Stability hypotheses for $\dot\theta = h(\theta)$. Hypothesis (U4)-(L4)** *We assume the existence of a function $\theta \to \mathcal{L}(\theta)$ which is twice continuously differentiable such that, for all $\theta \in \mathbb{R}^d$*

$$\max_\theta |\mathcal{L}''(\theta)| = D_L < \infty \,, \tag{17}$$

$$\langle \nabla\mathcal{L}(\theta)\,,\, h(\theta)\rangle \le 0 \tag{18}$$

and

$$1 + \mathcal{L}(\theta) \ge c_L|\theta|^2 \tag{19}$$

for some $c_L > 0$.

Hypothesis (U5)-(L5) *We assume (U4) and that there exists $\hat\theta$ such that (18) is strict for $\theta \neq \hat\theta$, and $\mathcal{L}(\theta) = 0$ if and only if $\theta = \hat\theta$.*

If we integrate equation (17), we get (maybe with another constant, but that we still denote by D_L), for all θ,

$$|\mathcal{L}'(\theta)| \le D_L(1 + |\theta|)\,. \tag{20}$$

3.6. **Polynomial growth of ζ in case U. Hypothesis (U6)** *There exists a constant r such that, for all $R > 0$*

$$\sup_{|\theta|\le R} \max(\zeta_f(\theta), \zeta_P)\zeta_e(\theta)^2 \le D_6(1 + |\theta|)^r$$

Remark: The assumptions $\zeta_e, \zeta_P, \zeta_f \ge 1$ are there merely for convenience. They clearly do not restrict the generality.

4. Convergence theorem

Theorem 1. *1) Assume conditions (U0) to (U4) and (U6). Then, there exists constants C_0 and C_0' such that, if $C > C_0$ and $C' > C_0'$ in (U3), the sequence θ_n associated to the algorithm (5), (6) is bounded with probability 1.*

2) If the asymptotic stability condition (U5) is added, θ_n converges almost surely to $\hat\theta$.

Theorem 2. *1) Assume conditions (L0) to (L4). Then, the sequence θ_n associated to the algorithm (5), (6) is bounded with probability 1.*

2) If the asymptotic stability condition (L5) is added, θ_n converges almost surely to $\hat\theta$.

5. Proof

5.1. **Preamble.** The proofs of both theorems share large common parts and therefore may be held in parallel.

In both cases, the proof of 2) is, by well-known results in stochastic approximation theory, a consequence of 1) (see [16], or [1], for example).

The proof of 1) relies on the following basic decomposition of the algorithm

$$\theta_{n+1} = \theta_n + \gamma_{n+1}h(\theta_n) + \frac{\gamma_{n+1}}{k_n}\sum_{q=1}^{k_n}[f(\theta_n, Y^{n+1,q}) - h(\theta_n)]\,.$$

The equation $\theta_{n+1} = \theta_n + \gamma_{n+1}h(\theta_n)$ is a discretization of the ordinary differential equation (mean ODE) $\dot{\theta} = h(\theta)$. The algorithm in (5) therefore is a perturbation of this discretization, the "noise" term being

$$\xi_{n+1} = \frac{1}{k_n}\sum_{q=1}^{k_n}[f(\theta_n, Y^{n+1,q}) - h(\theta_n)] .$$

The main part of the proof is to show that the contribution of the ξ_n altogether is not too large.

In the following, we will let D be an generic constant, for which we shall not trace the value, which may change from line to line although we keep the same notation. If some emphasis needs to be made on the dependence of D with respect to some quantities (say R), we will write $D(R)$.

We let $\eta = 0$ in case (U) and $\eta = 1$ in case (L) so that (U0) and (L0) can be simultaneously written

$$\sup_x |f(\theta, x)| \leq D_f(1 + |\theta|)^\eta .$$

5.2. *A priori* **bounds.** The following easy lemma is crucial. For $k > 0$, we set

$$\delta(p, p+k) = \sum_{p'=p+1}^{p+k} \gamma_{p'} .$$

Lemma 1. *If (U0) is true, we have*

$$|\theta_{p+k} - \theta_p| \leq D_f \delta(p, p+k) \tag{21}$$

and if (L0) is true

$$|\theta_{p+k} - \theta_p| \leq (1 + |\theta_p|)[\exp(D_f\delta(p, p+k)) - 1] . \tag{22}$$

Proof of lemma 1. Indeed, (21) is trivial from

$$|\theta_{n+1} - \theta_n| \leq \gamma_{n+1} D_f \tag{23}$$

and (22) may be proved by induction. □

Thus, under condition (U3), in which $\gamma_{p+1} = a/(p+1)$, we have, in case (U)

$$|\theta_n| \leq aD_f \log(n+1) . \tag{24}$$

We will also use a simple lemma on the sequence k_n:

Lemma 2. *Under condition (L3)-b, we have*

$$\sum_{p=n_0}^{n} \gamma_p \left| \frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}} \right| \leq \gamma_{n_0} + \sum_{q \geq n_0} \gamma_p \overline{\gamma}_q + 2 \sum_{l_q \geq n_0} \gamma_{l_q} \tag{25}$$

Note that, since (U3)-b implies (L3)-b with $\overline{\gamma}_p = 0$, we also get an upper bound in case (U).

Proof of lemma 2 Under (L3)-b, $\left|\frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}}\right|$ is smaller than $\frac{1}{\sqrt{k_{p-1}}} - \frac{1}{\sqrt{k_p}} + \overline{\gamma}_p$ when p is equal to no l_n. If if $p = l_n$ for some n, we may write, since $k_p \geq 1$:

$$\left| \frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}} \right| \leq 2 + \frac{1}{\sqrt{k_{p-1}}} - \frac{1}{\sqrt{k_p}} .$$

This implies that we can bound

$$\sum_{p=n_0}^{n} \gamma_p \left| \frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}} \right|$$

by

$$\sum_{p=n_0}^{n} \gamma_p \left(\frac{1}{\sqrt{k_{p-1}}} - \frac{1}{\sqrt{k_p}} \right) + \sum_{p=n_0}^{n} \gamma_p \overline{\gamma}_p + 2 \sum_{q, n_0 \leq l_q \leq n} \gamma_{l_q} .$$

The first sum is equal to

$$\sum_{p=n_0}^{n} (\gamma_{p+1} - \gamma_p)\frac{1}{\sqrt{k_p}} + \frac{\gamma_{n_0}}{\sqrt{k_{n_0-1}}} - \frac{\gamma_{n+1}}{\sqrt{k_n}} \leq \gamma_{n_0}$$

since γ_p is decreasing and $k_p \geq 1$, which yields (25). □

5.3. **Solution of the Poisson equation for P_θ.** Following [1], we introduce the solution ρ_θ of the recursive equation

$$\rho_\theta - P_\theta \rho_\theta = g_\theta ,$$

where $g_\theta(x) = f(\theta, x) - h(\theta)$. More precisely, we let

$$\rho_\theta(x) = \sum_{k=0}^{\infty} (P_\theta^k g_\theta)(x) . \tag{26}$$

We check that this expression converges. Indeed, by (12),

$$\Delta(P_\theta^k g_\theta) \leq \zeta_e(\theta)(1 + |\theta|)^\eta \lambda_\theta^k .$$

Moreover, we have $\pi_\theta(P_\theta^k g_\theta) = 0$ and this implies

$$\sup_x | P_\theta^k g_\theta(x) | = \sup_x | P_\theta^k g_\theta(x) - \pi_\theta(P_\theta^k g_\theta) | \leq \Delta(P_\theta^k g_\theta) .$$

Moreover,

Lemma 3. *We have, for some constant D_ρ,*

$$\Delta \rho_\theta \leq D_\rho \zeta_e(\theta)(1 + |\theta|)^\eta (1 - \lambda_\theta)^{-1} \tag{27}$$

with $\eta = 0$ in case (U) and $\eta = 1$ in case (L).

Moreover, letting $\hat{\rho}_\theta = P_\theta \rho_\theta$,

$$|\nabla_\theta \hat{\rho}_\theta| \leq D_\rho \zeta_\rho(\theta)(1 + |\theta|)^\eta (1 - \lambda_\theta)^{-2} . \tag{28}$$

with $\zeta_\rho = \zeta_e^2 \max(\zeta_P, \zeta_f)$.

Proof of lemma 3. Let $\epsilon = 0$ in case (U) and $\epsilon = 1$ in case (L). Applying (12), we have

$$\Delta(\rho_\theta) \leq \zeta_e(\theta) \Delta g_\theta . \sum_{k=1}^{\infty} \lambda_\theta^k \leq 2\zeta_e(\theta) D_f (1 + |\theta|)^\eta (1 - \lambda_\theta)^{-1} ,$$

which yields (27).

We now prove (28). For this, we need another lemma:

Lemma 4. *Let (w_θ) be a family of functions on Ω such that $\nabla_\theta w_\theta$ exists (as a bounded function over Ω) for all θ. Let $\hat{w}_\theta = w_\theta - \pi_\theta w_\theta$. Let finally $\hat{w}_\theta^n = P_\theta^n \hat{w}_\theta$. We have, for $n \geq 0$,*

$$\begin{aligned} \nabla_\theta \hat{w}_\theta^n(x) &= [\nabla \pi_\theta](P_\theta^n w_\theta(x) - P_\theta^n w_\theta(.)) \\ &+ P_\theta^n \nabla_\theta w_\theta(x) - \pi_\theta \nabla_\theta w_\theta \\ &+ \sum_{i=0}^{n-1} \pi_\theta \left(P_\theta^i [\nabla P_\theta] P^{n-i-1} w_\theta(x) - P_\theta^i [\nabla P_\theta] P^{n-i-1} w_\theta(.) \right) \end{aligned} \tag{29}$$

Proof of lemma 4: Equation (29) may be rewritten as

$$\begin{aligned} \nabla_\theta \hat{w}_\theta^n(x) &= -[\nabla \pi_\theta] P_\theta^n w_\theta \\ &+ P_\theta^n \nabla_\theta w_\theta(x) - \pi_\theta \nabla_\theta w_\theta \\ &+ \sum_{i=0}^{n-1} P_\theta^i [\nabla P_\theta] P^{n-i-1} w_\theta(x) - \pi_\theta [\nabla P_\theta] P^{n-i-1} w_\theta(.) , \end{aligned} \tag{30}$$

which can be proved by induction using

$$P_\theta^{n+1}(w_\theta - \pi_\theta w_\theta) = P_\theta^n(P_\theta w_\theta - \pi_\theta P_\theta w_\theta) .$$

□

End of the proof of lemma 3: Since $u_\theta(x) := P_\theta g_\theta(x) = P_\theta f_\theta(x) - \pi_\theta P_\theta f_\theta$ lemma 4 may be applied to $w_\theta = P_\theta f_\theta$. This yields, assuming (L1) and (U2)-(L2)

$$\begin{aligned}|\nabla_\theta P_\theta^n u_\theta(x)| &\leq 2D_f\zeta_P(\theta)\zeta_e(\theta)(1+|\theta|)^\eta\lambda_\theta^{n+1}\\ &\quad +2\zeta_e(\theta)\zeta_f(\theta)\lambda_\theta^{n+1}\\ &\quad +2nD_f\zeta_e(\theta)^2\zeta_P(\theta)(1+|\theta|)^\eta\lambda_\theta^n\\ &\leq 6nD_f\zeta_e(\theta)^2\max(\zeta_P(\theta),\zeta_f(\theta))(1+|\theta|)^\eta\lambda_\theta^n\end{aligned}$$

Since

$$\hat\rho_\theta(x) = P_\theta\rho_\theta(x) = \sum_{n\geq 0} P_\theta^n u_\theta(x)\,,$$

we get the fact that $\hat\rho$ is differentiable and the upper bound (28). □

5.4. **Convergence of the noise term.** Set, for $R>0$, $0\leq n_0\leq n$,

$$A_n^{n_0}(R) = \sum_{p=n_0}^{n}\gamma_{p+1}d_p^R\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\xi_{p+1}\rangle\,,$$

where d_p^R is the indicator of the event $\{|\theta_q|\leq R,\ q\leq p\}$. We have

$$\begin{aligned}A_n^{n_0} &= \sum_{p=n_0}^{n}\gamma_{p+1}d_p^R\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\frac{1}{k_p}\sum_{q=1}^{k_p}g_{\theta_p}(Y^{p+1,q})\right\rangle\\ &= \sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\sum_{q=1}^{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\rho_{\theta_p}(Y^{p+1,q})-P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q})\right\rangle\\ &= \sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\sum_{q=1}^{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\rho_{\theta_p}(Y^{p+1,q})-P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q-1})\right\rangle\\ &\quad+\sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\sum_{q=1}^{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q-1})-P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q})\right\rangle\end{aligned}$$

Thus, remembering that, by construction, $Y^{p+1,0}=Y^{p,k_{p-1}}=X^p$, we have

$$\begin{aligned}A_n^{n_0} &= \sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\sum_{q=1}^{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\rho_{\theta_p}(Y^{p+1,q})-P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q-1})\right\rangle\\ &\quad+\sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,P_{\theta_p}\rho_{\theta_p}(X^p)-P_{\theta_p}\rho_{\theta_p}(X^{p+1})\right\rangle\end{aligned}$$

Let

$$B_n^{n_0} = \sum_{p=n_0}^{n}\frac{\gamma_{p+1}}{k_p}d_p^R\sum_{q'=1}^{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,\rho_{\theta_p}(Y^{p+1,q'})-P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q'-1})\right\rangle.$$

Let $\mathcal{F}_p$ be the σ-algebra generated by the random variables $Y^{p',q'}$ for $p'\leq p$, $q'\leq k_{p'}$. Clearly θ_p, k_p and d_p^R are $\mathcal{F}_p$-measurable, which implies that the sequence $B_p^{n_0}$ is a martingale adapted to $\mathcal{F}_p$.

We have, denoting by $\mathbf{E}$ the expectation with respect to the distribution of the sequence $Y^{p,q}$,

$$\begin{aligned}
\mathbf{E}[(B_n^{n_0})^2] &= \sum_{p=n_0}^{n} \gamma_{p+1}^2 \mathbf{E}\left[\frac{d_p^R}{k_p^2}\sum_{q'=1}^{k_p}\left\langle \nabla_\theta \mathcal{L}(\theta_p)\,,\,\rho_{\theta_p}(Y^{p+1,q'}) - P_{\theta_p}\rho_{\theta_p}(Y^{p+1,q'-1})\right\rangle^2\right] \\
&\leq \sum_{p=n_0}^{n} \gamma_{p+1}^2 \mathbf{E}\left[\frac{d_p^R}{k_p}|\nabla_\theta \mathcal{L}(\theta_p)|^2 \Delta(\rho_{\theta_p})^2\right] \\
&\leq \sum_{p=n_0}^{n} \gamma_{p+1}^2 D_L^2 \mathbf{E}\left[\frac{d_p^R}{k_p}(1+|\theta_p|)^{2+2\eta}\zeta_e(\theta_p)^2(1-\lambda_{\theta_p})^{-2}\right]
\end{aligned}$$

The last inequality comes from lemma 3 and equation (20).

Assume first that we are in case (U). We have, by (13), that

$$(1-\lambda_{\theta_p})^{-2} \leq K_\lambda^{-2} e^{2D_\lambda|\theta_p|}\,.$$

Moreover,

$$1/k_p \leq e^{-a'|\theta_p|}\,,$$

and assumptions (U6) implies $|\zeta_e(\theta)| \leq D_6(1+|\theta|)^r$. Since $|\theta_p| \leq R$ when $d_p^R \neq 0$, this gives

$$\begin{aligned}
\mathbf{E}[(B_n^{n_0})^2] &\leq D_L^2 D_6^2 K_\lambda^{-2}\sum_{p=n_0}^{n}\gamma_{p+1}^2 \mathbf{E}[d_p^R(1+|\theta_p|)^{2r+2}e^{(2D_\lambda - a')|\theta_p|} \\
&\leq D_L^2 D_6^2 K_\lambda^{-2}(1+R)^{2r+2}e^{(2D_\lambda-a')^+R}\sum_{p=n_0}^{n}\gamma_{p+1}^2
\end{aligned}$$

where α^+ is the positive part of the real number α.

Now, let's take case (L). We have

$$\mu k_p \geq \mu\kappa(\theta_p) \geq (1-\lambda_{\theta_p})^{-2}\zeta_\rho\,,$$

and $\zeta_\rho = \zeta_\kappa \geq \zeta^2$, so that

$$\mathbf{E}[(B_n^{n_0})^2] \leq (1+R)^4 D_L^2\mu\sum_{p=n_0}^{n}\gamma_{p+1}^2$$

We now study the remaining part of $A_n^{n_0}$, which is

$$\begin{aligned}
C_n^{n_0} &= \sum_{p=n_0}^{n}\frac{\gamma_{p+1}d_p^R}{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,P_{\theta_p}\rho_{\theta_p}(X^p) - P_{\theta_p}\rho_{\theta_p}(X^{p+1})\right\rangle \\
&= \sum_{p=n_0+1}^{n}\left\{\frac{\gamma_{p+1}d_p^R}{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p)\,,\,P_{\theta_p}\rho_{\theta_p}(X^p)\right\rangle - \frac{\gamma_p d_{p-1}^R}{k_{p-1}}\left\langle\nabla_\theta\mathcal{L}(\theta_{p-1})\,,\,P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\right\rangle\right\} \\
&\quad -\frac{\gamma_{n+1}d_n^R}{k_n}\left\langle\nabla_\theta\mathcal{L}(\theta_n)\,,\,P_{\theta_n}\rho_{\theta_n}(X^{n+1})\right\rangle + \frac{\gamma_{n_0+1}d_{n_0}^R}{k_{n_0}}\left\langle\nabla_\theta\mathcal{L}(\theta_{n_0})\,,\,P_{\theta_{n_0}}\rho_{\theta_{n_0}}(X^{n_0})\right\rangle
\end{aligned}$$

The last two terms may be bounded from above by $2K_\lambda\gamma_{n_0+1}(1+R)^{(1+r)}e^{(D_\lambda-a')^+R}$ (case (U)), and by $2D_\rho\mu\gamma_{n_0+1}(1+R)^2$ (case (L)).

To estimate the sum, we use the decomposition

$$\begin{aligned}
\epsilon_p^{(1)} &= \tfrac{\gamma_{p+1}d_p^R}{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_p) - \nabla_\theta\mathcal{L}(\theta_{p-1})\,,\,P_{\theta_p}\rho_{\theta_p}(X^p)\right\rangle \\
\epsilon_p^{(2)} &= \tfrac{\gamma_{p+1}d_p^R}{k_p}\left\langle\nabla_\theta\mathcal{L}(\theta_{p-1})\,,\,P_{\theta_p}\rho_{\theta_p}(X^p) - P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\right\rangle \\
\epsilon_p^{(3)} &= \tfrac{1}{k_p}(\gamma_{p+1}-\gamma_p)d_p^R\left\langle\nabla_\theta\mathcal{L}(\theta_{p-1})\,,\,P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\right\rangle \\
\epsilon_p^{(4)} &= \gamma_p\left(\tfrac{1}{k_p} - \tfrac{1}{k_{p-1}}\right)d_p^R\left\langle\nabla_\theta\mathcal{L}(\theta_{p-1})\,,\,P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\right\rangle
\end{aligned}$$

$\epsilon_p^{(5)} = \frac{\gamma_p}{k_{p-1}}(d_p^R - d_{p-1}^R)\langle \nabla_\theta \mathcal{L}(\theta_{p-1})\, ,\, P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\rangle$
to prove the lemma

Lemma 5. *In case (U), we have, for a constant D_C, depending on the previously introduced constants, but not on R:*

$$|C_n^{n_0}| \leq D_C(1+R)^{r+1}e^{(2D_\lambda - a')^+ R}(\gamma_{n_0+1} + \sum_{p=n_0+1}^{n} \gamma_p^2 + \sum_{q, n_0+1 \leq l_q \leq n} \gamma_{l_q})\,. \tag{31}$$

In case (L)

$$|C_n^{n_0}| \leq D_C(1+R)^2(\gamma_{n_0+1} + \sum_{p=n_0+1}^{n} \gamma_p^2 + \sum_{q \geq n_0} \gamma_p \overline{\gamma}_q + \sum_{q, n_0+1 \leq l_q \leq n} \gamma_{l_q})\,. \tag{32}$$

Proof of lemma 5.

This lemma is a consequence of upper bounds on the $\epsilon^{(j)}$, which are obtained with the help of the following hints. For $j = 1$, we use the inequality (23) and the fact that the second derivative of U is bounded. This yields the estimate

$$\sum_{p=n_0+1}^{n} |\epsilon_p^{(1)}| \leq D(1+R)^{1+r}e^{(D_\lambda - a')^+ R} \sum_{p=n_0+1}^{n} \gamma_{p+1}^2 \sum_{p=n_0+1}^{n} \gamma_{p+1}^2$$

in case (U) and

$$\sum_{p=n_0+1}^{n} |\epsilon_p^{(1)}| \leq D(1+R)^2 \sum_{p=n_0+1}^{n} \gamma_{p+1}^2$$

in case (L).

For $j = 2$, we use

$$|P_{\theta_p}\rho_{\theta_p}(X^p) - P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)| \leq 2|\theta_p - \theta_{p-1}|\zeta_\rho(\theta_p^*)(1+|\theta_p^*|)(1-\lambda_{\theta_p^*})^{-2}\,,$$

where θ_p^* lies on the segment $[\theta_{p-1}, \theta_p]$. After division by k_n, this is bounded by, in case (U),

$$D.\gamma_p^2(1+R)^{1+r}e^{(2D_\lambda - a')^+ R}\,,$$

and in case (L) (using (16))

$$D.\gamma_p^2(1+R)^2\,.$$

For $j = 3$, we bound

$$(1/k_p)d_p^R\langle \nabla_\theta \mathcal{L}(\theta_{p-1})\, ,\, P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\rangle$$

by $D.(1+R)^{1+r}e^{(D_\lambda - a')^+ R}$ in case (U) and by $D.(1+R)^2$ in case (L), and use the fact that γ_p is decreasing which yields $\sum_{p \geq n_0+1} |\gamma_p - \gamma_{p-1}| = \gamma_{n_0}$.

For $j = 4$, we write

$$\frac{1}{k_p} - \frac{1}{k_{p-1}} = (\frac{1}{\sqrt{k_{p-1}}} + \frac{1}{\sqrt{k_p}})(\frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}})\,.$$

The part

$$(\frac{1}{\sqrt{k_{p-1}}} + \frac{1}{\sqrt{k_p}})d_p^R\langle \nabla_\theta \mathcal{L}(\theta_{p-1})\, ,\, P_{\theta_{p-1}}\rho_{\theta_{p-1}}(X^p)\rangle$$

may be bounded by $D.(1+R)^{r+1}e^{(D_\lambda - a'/2)^+ R}$ in case (U) and by $D.(1+R)^2$ in case (L). Applying lemma 2, we get

$$\sum_{p=n_0+1}^{n} \epsilon_p^{(4)} \leq D.(1+R)^{r+1}e^{(D_\lambda - a'/2)^+ R}\left(\gamma_{n_0} + \sum_{q \geq n_0} \gamma_p \overline{\gamma}_q + \sum_{l_q \geq n_0} \gamma_{l_q}\right)$$

in case (U) and

$$\sum_{p=n_0+1}^{n} \epsilon_p^{(4)} \leq D.(1+R)^2 \left(\gamma_{n_0} + \sum_{q\geq n_0} \gamma_p \overline{\gamma}_q + \sum_{l_q \geq n_0} \gamma_{l_q}\right)$$

Finally, for $j=5$, we have

$$\begin{aligned}\sum_{p=n_0+1}^{n} |\epsilon_p^(5)| &\leq \max_{p=n_0}^{n} d_{p-1}^R \left| \frac{\gamma_p}{k_{p-1}} \langle \nabla_\theta \mathcal{L}(\theta_{p-1}) \,,\, P_{\theta_{p-1}} \rho_{\theta_{p-1}}(X^p) \rangle \right| \\ &\leq \gamma_{n_0} (1+R)^{1+\eta} \max_{p=n_0}^{n} d_{p-1}^R \frac{\zeta_e(\theta_{p-1})}{k_{p-1}}\end{aligned}$$

and this is smaller than $D_6(1+R)^{1+r}\gamma_{n_0+1}$ in case (U) and than $(1+R)^2\gamma_{n_0+1}$ in case (L).

□

Returning to $A_n^{n_0}$, we have shown (using Doob's inequality for the martingale part)

Lemma 6. *For some constant $D_A > 0$,*

$$\mathbf{E}\left[\sup_{n\geq n_0} \left| \sum_{p=n_0}^{n} \gamma_{p+1} d_p^R (\nabla_\theta \mathcal{L}(\theta_p) \,|\, \xi_{p+1}) \right|^2 \right] \leq D_A (1+R)^{2+2r} e^{2(2D_\lambda - \delta)^+ R}$$

$$\left(\gamma_{n_0} + \sum_{n\geq n_0} \gamma_n^2 + \sum_{q\geq n_0} \gamma_p \overline{\gamma}_q + \sum_{q, n_0+1 \leq l_q \leq n} \gamma_{l_q}\right)^2,$$

with $\delta = a'$ in case (U) and $\delta = 2D_\lambda$ in case (L), $r=1$ in case (L).

5.5. **End of the proof.** We now show that (θ_n) is almost surely bounded. Following [1], we analyse the exploding trajectories of this process. We have

(33) $$\mathcal{L}(\theta_{n+1}) = \mathcal{L}(\theta_n) + \gamma_{n+1}\langle \nabla_\theta \mathcal{L}(\theta_n) \,,\, \xi_{n+1}\rangle + 2\gamma_{n+1}\langle \nabla_\theta \mathcal{L}(\theta_n) \,,\, h(\theta_n)\rangle + D_L(|\theta_{n+1} - \theta_n|^2)$$

From (U4), we have

(34) $$\mathcal{L}(\theta_{n+1}) \leq \mathcal{L}(\theta_n) + \gamma_{n+1}\langle \nabla_\theta \mathcal{L}(\theta_n) \,,\, \xi_{n+1}\rangle + D_L|\theta_{n+1} - \theta_n|^2$$

Let R_n be an increasing sequence which grows to infinity. We let σ_n be the first integer p for which $\mathcal{L}(\theta_p) > R_n$. Let τ_n be the last p smaller than σ_{n+1} for which $\mathcal{L}(\theta_{p-1}) \leq R_n$ and $\mathcal{L}(\theta_p) > R_n$. Note that $\sigma_n \leq \tau_n$ and if $p < \sigma_n$, $c_U|\theta_p|^2 < 1 + R_p$. We let $\tilde{R}_p = \sqrt{1+R_p}/\sqrt{c_U}$.

Now, assume that there exists a deterministic sequence (s_n) such that $s_n \leq \sigma_n$ almost surely. Using (34), we have, if $\sigma_{n+1} < \infty$

$$(\mathcal{L}(\theta_{\sigma_{n+1}}) - \mathcal{L}(\theta_{\tau_n - 1}))^2 \leq 2[\sum_{p=\tau_n-1}^{\sigma_{n+1}-1} \gamma_{p+1}\langle \nabla_\theta \mathcal{L}(\theta_p) \,,\, \xi_{p+1}\rangle]^2 + 2D_L^2[\sum_{p=\tau_n-1}^{\sigma_{n+1}-1} |\theta_{p+1} - \theta_p|^2]^2$$

The last sum is smaller than $D(1+\tilde{R}_{n+1})^4[\sum_{p\geq s_n} \gamma_p^2]^2$. Writing $\epsilon_p = \gamma_{p+1}(\nabla_\theta \mathcal{L}(\theta_p) \,|\, \xi_{p+1})$, we have

$$\begin{aligned}[\sum_{p=\tau_n-1}^{\sigma_{n+1}-1} \epsilon_p]^2 &= [\sum_{p=s_n-1}^{\sigma_{n+1}-1} \epsilon_p - \sum_{p=s_n-1}^{\tau_n-1} \epsilon_p]^2 \\ &\leq 2 \sup_{s_n-1\leq k\leq \sigma_{n+1}-1} [\sum_{p=s_n-1}^{k} \epsilon_p]^2 \\ &= 2 \sup_{s_n-1\leq k\leq \sigma_{n+1}-1} [\sum_{p=s_n-1}^{k} d_p^{\tilde{R}_{n+1}} \epsilon_p]^2 \\ &\leq 2 \sup_{s_n-1\leq k} [A_k^{s_n-1}(\tilde{R}_{n+1})]^2\end{aligned}$$

We have obtained

(35)

$$\mathbf{E}[(\mathcal{L}(\theta_{\sigma_{n+1}}) - \mathcal{L}(\theta_{\tau_n - 1}))^2 \mathbf{1}_{\sigma_{n+1}<\infty}] \leq \overline{D}\mathbf{E}\left[(\sup_{p\geq s_n - 1} A_p^{s_n-1}(1+\tilde{R}_{n+1}))^2 + \tilde{R}_{n+1}^4 (\sum_{p\geq s_n} \gamma_p^2)^2 \right].$$

Since we also have $\mathcal{L}(\theta_{\sigma_{n+1}}) - \mathcal{L}(\theta_{\tau_n - 1}) \geq (R_{n+1} - R_n)$, this yields

$$P(\sigma_{n+1} < \infty) \leq \overline{D}\mathbf{E}\left[(\sup_{p\geq s_n - 1} A_p^{s_n-1}(\tilde{R}_{n+1}))^2 + (1+\tilde{R}_{n+1})^4 (\sum_{p\geq s_n} \gamma_p^2)^2 \right] / (R_{n+1} - R_n)^2 . \tag{36}$$

We now separate cases (U) and (L). In case (U), we have, $|\theta_p| \leq |\theta_0| + aD_f \log(p+1)$. We thus have $\mathcal{L}(\theta_p) \leq D_L(1+|\theta_0|+aD_f\log(p+1))^2$, and if we choose $s_n = \exp(\alpha\sqrt{R_n})$ with $\alpha < 1/aD_f\sqrt{D_L}$ (say $\alpha = 1/2aD_f\sqrt{D_L}$) we have $s_n \leq \sigma_n$ at least for large n.

Moreover, $\sum_{p\geq s_n} \gamma_p^2$ is smaller than a^2/s_n, and $\sum_{p^{1+c_l}\geq s_n} \gamma_{p^{1+c_l}}$ is smaller than $a/[s_n^{\frac{c_l}{1+c_l}}]$. Using lemma 6, (36) yields

$$\mathbf{P}(\sigma_{n+1} < \infty) < \overline{D}' \frac{(1+R_{n+1})^{q+1}}{(R_{n+1} - R_n - 1)^2} e^{(2D_\lambda - a')^+ \sqrt{\frac{R_{n+1}}{c_U}} - \frac{c_l\sqrt{R_n}}{2(1+c_l)aD_f\sqrt{D_L}}} .$$

If

$$(a'/c_U) + c_l/(2aD_f(1+c_l)\sqrt{D_L}) > 2D_\lambda , \tag{37}$$

we get, letting $R_n = n^2$ the fact that $\mathbf{P}(\sigma_{n+1} < \infty) \to 0$. This gives the result in this case, the constants C_0 and C_0' being provided by (37).

In case (L), we have

$$|\theta_p| < (1+|\theta_0|)e^{2D_f\delta(0,p)}$$

and we can set $s_n = \min\{p; (1+|\theta_0|)e^{2D_f\delta(0,p)} > \sqrt{R_n/D_L}\}$ which tends to infinity.

In this case, we have

$$\mathbf{P}(\sigma_{n+1} < \infty) < \overline{D}' \frac{(1+R_{n+1})^2}{(R_{n+1} - R_n)^2} \omega_n ,$$

with

$$\omega_n = (\gamma_{s_n} + \sum_{n\geq s_n} \gamma_n^2 + \sum_{q\geq s_n} \gamma_p \overline{\gamma}_q \sum_{q, s_n+1 \leq l_q} \gamma_{l_q}) \to 0 .$$

and $R_n = 2^n$ gives the conclusion. □

6. Remarks

6.1. Dependance on the state variable. Since our emphasis is on the relation between the growth of θ and the ergodicity of the Markov chains associated to P_θ, we have avoided additional technicalities, by always assuming uniform upper bounds with respect to the state variable x. This essentially restrict the range of applications of our results to finite or compact state spaces. However, we could have replaced this simplifying hypothesis by weaker assumptions enabling to control the moments of the variables $Y^{p,q}$, as in [1] (part 2, chapter 3). Most of the methods used in this reference to control the variable x can be transcribed to our case.

6.2. Remarks on the choice of the sequence k_p. We have two hypotheses on the sequence k_p. One says that k_p should not be smaller than a function $\kappa(\theta)$, which is large when the Markov chains associated to P_θ looses its ergodicity. Note that it is always possible to take $\kappa(\theta) =$ constant on a bounded set, which means that the use of large values of k_n may be used as a reaction to some exploding trajectory of the algorithm. Such an event only happens with a small probability, so that it is likely that large values of k_n may not be used too often in applications. This is a significant improvement to the procedure described in ([21]) (which corresponds to case (U) with

$a' = 0$) for which the steps γ_n are assumed to be small enough from the beginning, in order to prevent rare exploding trajectories.

The other condition assumes that k_p should not decrease too much too often. This may be seen as a technical condition, and we do not know whether this is necessary or not for convergence. It was required when dealing with the sequence $\epsilon_q^{(4)}$ in the proof on theorems 1, 2, and was not used elsewhere. This implies, in particular, that, if this condition is dropped, the convergence of the noise term reduces to the convergence of $\sum \epsilon_q^{(4)}$. Since $\sum \gamma_p$ diverges, it seems hard to get some control over the noise term without some regularity assumption on the behaviour of k_n. Without changing the previous proof, any condition ensuring that

$$\sum_{p=n_0}^{n} \gamma_p \left| \frac{1}{\sqrt{k_p}} - \frac{1}{\sqrt{k_{p-1}}} \right|$$

is small will work.

6.3. **Consequences of the existence of metastable states.** In this paragraph, we show that problems can really occur when no assumption is made on γ_p and k_p, in particular when there exist metastable states for the family of transition probabilities (P_θ). [1]

To simplify, we let (in this paragraph) $d = 1$, ie. θ is a real number, and let Ω be finite (we will then consider P_θ as a stochastic matrix : $P_\theta = (p_\theta(x, y), x, y \in \Omega))$. A metastable state is a state $x_m \in \Omega$ such that

$$\begin{cases} \lim_{\theta \to +\infty} p_\theta(x_m, x_m) &= 1 \\ \lim_{\theta \to +\infty} \pi_\theta(x_m) &= 0 \end{cases} .$$

We will moreover assume that the convergence of $p_\theta(x_m, x_m)$ to 1 is very fast :

$$1 - P_\theta(x_m, x_m) \simeq e^{-\rho\theta} \text{ for some } \rho > 0 .$$

This situation is typical in the context of Gibbs distributions.

Still for simplification, assume that $f(\theta, x) = f(x)$ is independent of θ and that $f(x_m) > 0$. Consider the situation when the states $Y^{p,q}$ all remain equal to x_m, ie. the process is trapped in the metastable state. In that case, the procedure diverges, and we now check whether this may occur with positive probability.

Let A_{n+1} be the event : " at step n, the Markov chain $Y^{n,1}, \ldots, Y^{n,k_n}$ with transition P_{θ_n} satisfies $Y^{n,l} = x_m$ for all l." Let B_n be the intersection of all A_p for $p \leq n$. If B_n is true, we have $\theta_{p+1} = \theta_p + \gamma_{p+1} f(x_m)$ for $p \leq n$, which implies that $\theta_{n+1} = \theta_0 + f(x_m) \sum_{p=1}^{n+1} \gamma_p$.

Then

$$\begin{aligned} P(B_{n+1})/P(B_n) &= P(B_{n+1} \,|\, B_n) \\ &= P(A_{n+1} \,|\, B_n) \\ &= P(A_{n+1} \,|\, X^{n+1} = x_m, \theta_{n+1} = \theta_0 + f(x_m) \sum_{p=1}^{n+1} \gamma_p) \\ &= P_{\theta_{n+1}}(x_m, x_m)^{k_{n+1}} \end{aligned}$$

We therefore have

$$\log P(B_n) = \log P(B_0) + \sum_{q=1}^{n} k_q \log P_{\theta_q}(x_m, x_m) .$$

Thus, for the algorithm to converges with probability 1, it is necessary that

$$\sum_{n \geq 1} \kappa(\theta_n) \log P_{\theta_n}(x_m, x_m) = -\infty . \tag{38}$$

[1] The idea of using metastable states to prove that the algorithms may diverge is due to H. Künsch (personal communication).

Since $1 - P_\theta(x_m, x_m) \simeq e^{-\rho\theta}$, this yields (letting $\tau_n = \sum_{p\leq n} \gamma_p$)

$$\sum_{n\geq 1} \kappa(\theta_0 + f(x_m)\tau_n) e^{-\rho H_{x_m} \tau_n} = \infty . \tag{39}$$

Thus, κ and γ_n may not be chosen arbitrarily, since this condition has to be true for all metastable point. For example, let $\gamma_n = \frac{a}{n}$, and approximate τ_n by $a \log n + \text{const}$. Then (39) gives

$$\sum_{n\geq 1} \frac{\kappa(\theta_0 + f(x_m)\tau_n)}{n^{a\rho f(x_m)}} = \infty .$$

If a is small enough, convergence is obtained without any condition on κ (this is the situation considered in [21]). To be able to use larger values of a, one must be ready to use functions $\kappa(\theta)$ which tend to infinity with θ.

7. Gaussian approximation of the algorithm

7.1. **Preliminaries.** The algorithm

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \left(\frac{1}{k_n} \sum_{q=1}^{k_n} f(\theta, Y^{q,n+1}) \right) . \tag{40}$$

being considered as a perturbation of the discretisation of the mean ODE

$$\dot{\theta} = h(\theta) = \pi_\theta f_\theta , \tag{41}$$

we measure the difference between the trajectories of (40) and those of the ODE, when the gains γ_n become small. For this, we consider a family of such algorithms, indexed by $N > 0$, each of them being associated to a sequence of gains $(\gamma_n^N, n \geq 1)$, and an initial parameter θ_0^N. This also requires a family of sequence (k_n^N), and states $Y^{N,p,q}$ (since the notation is becoming heavy, we shall let the superscript N appear only when it is necessary for good understanding). To define a time continuous version of these sequences, we let $t_n^N = \gamma_1^N + \ldots + \gamma_n^N$, and denote by $\theta^N(t)$, $\gamma^N(t)$ $\ldots$ the piecewise constant functions, equal to θ_n^N, γ_n^N $\ldots$ on the interval $[t_n^N, t_{n+1}^N[$. Note that, by inequality (24), we have, for $s < t$, in case (L),

$$|\theta^N(t) - \theta^N(s)| \leq D_f(1 + |\theta^N(s)|)(e^{|t-s|} - 1) . \tag{42}$$

We denote by $\overline{\theta}(t, \overline{\theta}_0)$ the solution of the mean ODE $\dot{\theta} = h(\theta)$, with initial condition $\overline{\theta}(0, \overline{\theta}_0) = \overline{\theta}_0$. Fixing a $\overline{\theta}_0$, we set

$$U_n^N = (\theta_n^N - \overline{\theta}(t_n^N, \overline{\theta}_0))/\sqrt{\gamma_n^N} .$$

abd let $U^N(t)$ be the piecewise constant function equal to U_n^N on $[t_n^N, t_{n+1}^N[$.

Our purpose is to check that the process U^N converges in distribution to a diffusion when N tends to infinity, under some additional conditions which are described now.

7.2. **Assumptions.** We assume (L0), (L1), (L2), (L4) and (L5), and we replace (L3) by a more specific condition which is

Hypothesis (L3)'

We take $\gamma_{n+1}^N = a/(n + N + 1)^b$, *with* $1/2 < b \leq 1$.

Let $\kappa(\theta)$ *be a function which satisfies, for constants* $\mu > 0$ *and* $q_\kappa \geq 0$.

$$\max_{|\theta'-\theta| \leq D_f(1+|\theta|)} \left[\zeta_\kappa(\theta')(1 - \lambda_{\theta'})^{-2} \right] \leq \mu.\kappa(\theta) , \tag{43}$$

and

$$|\nabla_\theta \log \kappa| \leq \mu , \tag{44}$$

for $\theta \in \mathbf{R}^d$.

Let $l_n = n^{1/b_1}$ *with* $b_1 < b/2$ *and* $\overline{\gamma}_p = p^{-b_2}$ *with* $b_2 > 1 - b/2$. *We let* $l_n^N = l_{n+N}^0$ *and the sequences* k_n^N *be given by*

- *If there exists n such that $p = l_n$, $k_p^N = [\kappa(\theta_p^N)]$*
- *If there exists no n such that $p = l_n$,*

$$k_p^N = \max \left[\frac{1}{(1/\sqrt{k_{p-1}^N} + \overline{\gamma}_p^N)^2}, [\kappa(\theta_p^N)] \right]$$

• *Moreover, we assume that for some fixed $R_0 > 0$, there exists a constant k_0, such that $\kappa(\theta) \equiv k_0$ for $|\theta| \leq R_0$, and that*

(45) $$\lim_{N\to\infty} \mathbf{P}[k_0^N \neq k_0] = 0$$

One can check that the constraints on γ_n^N and k_n^N for fixed n are stronger in (L3)' than in (L3), so that, for each N, the sequence θ_n^N converges almost surely to $\hat{\theta}$. We complete (L5) by (R_0 being the constant introduced in (L3)').

Hypothesis (L6)

• $|\hat{\theta}| < R_0$ *(where $\hat{\theta}$ is the asymptotically stable point of the ODE).*

For a comparison to hold, the starting points of (40) and (41) must be close to each other. Denote by χ_0^N the probability distribution of $U^N(0)$.

Hypothesis (L7)

We assume that χ_0^N converge to a distribution χ_0 with finite second moments when N tends to infinity.

This implies, letting $Q(R) = \sup_N P[|U^N(0)| > R]$, that $Q(R) \to 0$ if R tends to infinity. As a consequence, we have, if $R > |\overline{\theta}_0|$,

$$\begin{aligned} P(|\theta_0^N| > R) &\leq P\left(|U^N(0)| > \frac{R - |\overline{\theta}_0|}{\sqrt{\gamma_n^N}}\right) \\ &\leq P\left(|U^N(0)| > \frac{R - |\overline{\theta}_0|}{\sqrt{\gamma_0^N}}\right) \\ &\leq \sup_{N'} P\left(|U^{N'}(0)| > \frac{R - |\overline{\theta}_0|}{\sqrt{\gamma_0^N}}\right) \\ &\leq Q\left(\frac{R - |\overline{\theta}_0|}{\sqrt{\gamma_0^N}}\right) \end{aligned}$$

We thus have

Lemma 7. *For all $R > |\overline{\theta}_0|$, we have*

(46) $$\lim_{N\to\infty} \mathbf{P}[|\theta_0^N| > R] = 0$$

Finally, since $\hat{\theta}$ is asymptotically stable, one can find $R_1 < R_0$ and $R_2 > 0$ such that $R_2 < R_0$ and

(47) $$|\overline{\theta}_0| \leq R_1 \Rightarrow |\overline{\theta}(t, \overline{\theta}_0)| \leq R_2,\ t \geq 0\,.$$

We fix such values and make the last assumption, which ensures that (46) is true with $R = R_1$:

Hypothesis (L8)

$$|\overline{\theta}_0| < R_1$$

We then have the theorem :

Theorem 3. *Let the assumptions above be true, and define, for all* θ, *the matrix*

$$S(\theta) = \pi_\theta(g_\theta\, {}^t g_\theta) + 2\sum_{k\geq 1} \pi_\theta(g_\theta\, {}^t(P_\theta^k g_\theta)) . \tag{48}$$

Then for all $T > 0$, *the sequence of processes,* (U^N), *converges to a diffusion on* $[0, T]$, *with generator* L_t *and initial distribution* χ_0, *with*

$$L_t(\psi)(U) = \left\langle \psi'(U)\,,\, \left(\frac{\delta_1(b)}{2a} I + h'(\overline{\theta}(t, \overline{\theta}_0))\right).U \right\rangle + \frac{1}{2}\left\langle \psi''(U)\,,\, \frac{S(\overline{\theta}(t, \overline{\theta}_0))}{k_0} \right\rangle .$$

with $\delta_1(b) = 1$ *if* $b = 1$ *and 0 otherwise, the last term indicating the sum of termwise products of matrices* $\psi''(U)$ *and* S/k_0.

The proof of theorem 3 is adapted from the one in [1] (part 2, chapter 4). The same techniques, with the required modifications due to the new form of the algorithm, apply in our context. It is provided in section 8.

7.3. **Interpretation and consequences.** Let us describe some particular cases of application of this theorem. First, assume that $\theta_0^N = \overline{\theta}_0^N \equiv \theta_0$ for some θ_0 with $|\theta_0| < R_1$: this corresponds to comparing (40) to its associated ODE when they both start from the same point, and we see that the rescaled difference behaves like a diffusion when the steps are small.

Now, suppose that $\overline{\theta}_0 = \hat{\theta}$, the asymptotically stable point of the ODE. In this case, $\overline{\theta}(t, \overline{\theta}_0) \equiv \hat{\theta}$ and the generator of the limit diffusion is

$$L_t(\psi)(U) = \left\langle \psi'(U)\,,\, h'(\hat{\theta}).U(t) \right\rangle + \frac{1}{2}\left\langle \psi''(U)\,,\, \frac{S(\hat{\theta})}{k_0} \right\rangle .$$

This may provide some hint for the choice of k_0. Assume that $\chi_0^N \equiv \chi_0$, and assume that it is Gaussian and centered, with variance Σ_0 . In this case, the limit process, $U(t)$, remains Gaussian and is given by

$$U(t) = e^{th'(\hat{\theta})}U(0) + \frac{1}{\sqrt{k_0}} \int_0^t e^{(t-s)h'(\hat{\theta})} S^{1/2} dW_s$$

where W_s is a standard Brownian motion. The variance of $U(t)$ then is

$$\Sigma_t = e^{t\, h'(\hat{\theta})} \Sigma_0 e^{t\, {}^t h'(\hat{\theta})} + \frac{1}{k_0} \int_0^t e^{s\, h'(\hat{\theta})} S e^{s\, {}^t h'(\hat{\theta})} ds$$

This expression must be expressed in function of the true computation cost at time t. We define the computation cost as the number of variables $Y^{p,q}$ which have been simulated at time t. If t corresponds to n steps of the algorithm, the cost is $C = k_0 n$. Moreover, we have

$$t = a \sum_{p=1}^{n} (p + N)^{-b}$$

which is close to $a[(n + N)^{1-b} - N^{1-b}]$ when $b < 1$ and to $a \log(1 + n/N)$ when $b = 1$. This may be rewritten

$$t = aN \left[\left(1 + \frac{C}{Nk_0}\right)^{1-b} - 1 \right], \text{ if } b < 1, \text{ or,}$$

$$t = \log\left(1 + \frac{C}{Nk_0}\right) \text{ if } b = 1.$$

For a fixed cost C, Σ_t is a function which depends on k_0. It is in theory possible to minimise its trace (ie. the mean quadratic error of the algorithm), and one can check that this provides, in general, a non-trivial value of k_0. Unfortunately, in practice, the matrices S, h' (and Σ_0) cannot be computed, so that an exact optimal value of k_0 cannot be found beforehand.

Finally, let us consider the case when $\theta_0^N = \theta_N^0 = \theta_N$ for a fixed algorithm, and $\overline{\theta}_0^N = \hat{\theta}$. This corresponds to the comparison of the tail of the sequence $\theta_p, p \geq 0$ to its limit value $\hat{\theta}$. To apply the theorem, we must show that condition (L7) is true, and thus study the distribution of $(\theta_N - \hat{\theta})/\gamma_N$.

Let us make the following additional assumption:

Hypothesis (L9): *There exists $r_0 > 0$ such that, for all θ such that $|\theta - \hat{\theta}| \leq r_0$, we have*

$$\left\langle \theta - \hat{\theta}\,,\, h(\theta) \right\rangle < -\beta|\theta - \hat{\theta}|^2$$

Moreover, if $\gamma = a/n$ (ie. $b = 1$), we have $2a\beta > 1$.

For $R > 0$ and $N > 0$ consider the stopping time $\tau = \tau_R^N$ given by the first $n > N$ for which $k_n \neq k_0$ or $|\theta_n - \hat{\theta}| > R$. Since the sequence θ_n converges almost surely to $\hat{\theta}$, we have, for all fixed $R > 0$:

$$\lim_{N\to\infty} P\left[\tau_R^N < \infty\right] = 0 \tag{49}$$

We have the lemma

Lemma 8. *For all $R < r_0$ and all large enough N, there exists a constant $D(R, N)$such that, for all $p \geq N$,*

$$\mathbf{E}(|\theta_p - \hat{\theta}|^2 \mathbf{1}_{\tau_R^N > p}) \leq D(R, N)\gamma_p \tag{50}$$

Proof of Lemma 8 Although this proof is very close to the proof of ([1], part 2, theorem 1-24), we provide it here for completeness.

We fix N and $R < r_0$ and let $\tau = \tau_R^N$. We have

$$\begin{aligned} |\theta_{n+1} - \hat{\theta}|^2 &= |\theta_n - \hat{\theta}|^2 + 2\gamma_{n+1}\left\langle h(\theta_n)\,,\, \theta_n - \hat{\theta}\right\rangle \\ &\quad + 2\gamma_{n+1}\left\langle \xi_{n+1}\,,\, \theta_n - \hat{\theta}\right\rangle + \gamma_{n+1}^2 |h(\theta_n) + \xi_{n+1}|^2 \end{aligned}$$

so that, since $\tau > n+1 \Rightarrow \tau > n$, and, if $\theta_n \leq R$, there exists $D(R)$ such that $|h(\theta_n) + \xi_{n+1}| \leq D(R)$,

$$\begin{aligned} |\theta_{n+1} - \hat{\theta}|^2 \mathbf{1}_{\tau > n+1} &\leq (1 - 2\gamma_{n+1}\beta_R)|\theta_n - \hat{\theta}|^2 \mathbf{1}_{\tau > n} \\ &\quad + 2\gamma_{n+1}\left\langle \xi_{n+1}\,,\, \theta_n - \hat{\theta}\right\rangle \mathbf{1}_{\tau > n} + D(R)\gamma_{n+1}^2 \end{aligned}$$

Since

$$\begin{aligned} \xi_{n+1} &= \frac{1}{k_n}\sum_{q=1}^{k_n}[f(\theta_n, Y^{n+1,q}) - h(\theta_n)] \\ &= \frac{1}{k_n}\sum_{q=1}^{k_n}[\rho_{\theta_n}(Y^{n+1,q}) - P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q})] \\ &= \frac{1}{k_n}\sum_{q=1}^{k_n}[\rho_{\theta_n}(Y^{n+1,q}) - P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q-1})] \\ &\quad + \frac{1}{k_n}\sum_{q=1}^{k_n}[P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q-1}) - P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q})]\,, \end{aligned}$$

we have ($k_n = k_0$ if $\tau > n$)

$$\begin{aligned} \mathbf{E}\left[\left\langle \xi_{n+1}\,,\, \theta_n - \hat{\theta}\right\rangle \mathbf{1}_{\tau > n} | \mathcal{F}_n\right] &= \frac{1}{k_0}\sum_{q=1}^{k_0}\left\langle P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q-1}) - P_{\theta_n}\rho_{\theta_n}(Y^{n+1,q})\,,\, \theta_n - \hat{\theta}\right\rangle \mathbf{1}_{\tau > n} \\ &\quad + \frac{1}{k_0}\left\langle P_{\theta_n}\rho_{\theta_n}(X^n) - P_{\theta_n}\rho_{\theta_n}(X^{n+1})\,,\, \theta_n - \hat{\theta}\right\rangle \mathbf{1}_{\tau > n} \end{aligned}$$

Letting $Z_{n+1} = (1/k_0)\left\langle P_{\theta_n}\rho_{\theta_n}(X^{n+1})\,,\,\theta_n - \hat\theta\right\rangle$, using straightforward estimates deriving from our hypotheses, one can prove that

$$\mathbf{E}\left[\left\langle \xi_{n+1}\,,\,\theta_n - \hat\theta\right\rangle \mathbf{1}_{\tau>n}|\mathcal{F}_n\right] = \mathbf{1}_{\tau>n}(Z_n - Z_{n+1}) + r_n\mathbf{1}_{\tau>n}$$

with

$$|r_n| \le D(R)\gamma_{n+1}$$

on the set $\tau > n$. Thus, we get, with a new constant $D(R)$,

$$\begin{aligned}\mathbf{E}\left[|\theta_{n+1} - \hat\theta|^2\mathbf{1}_{\tau>n+1}\right] &\le (1 - 2\gamma_{n+1}\beta_R)\mathbf{E}\left[|\theta_n - \hat\theta|^2\mathbf{1}_{\tau>n}\right]\\ &\quad +2\gamma_{n+1}\mathbf{E}\left[\mathbf{1}_{\tau>n}(Z_n - Z_{n+1})\right] + D(R)\gamma_{n+1}^2\end{aligned}$$

Let $\alpha_n = \mathbf{E}\left[|\theta_n - \hat\theta|^2\mathbf{1}_{\tau>n}\right]$. An iteration of the previous estimate yields the following formula. For $p \le n$, let

$$V_p^n = \prod_{k=p}^{n}(1 - 2\beta_R\gamma_{k+1})\,,$$

(if $p > n$ we let $V_p^n = 1$). We have, for all $n \ge N$:

$$\begin{aligned}\alpha_n &\le V_N^{n-1}\alpha_N + D(R)\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p^2\\ &\quad +2\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p\mathbf{E}\left[\mathbf{1}_{\tau>p-1}(Z_{p-1} - Z_p)\right]\end{aligned}$$

But we can write

$$\begin{aligned}\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p\mathbf{1}_{\tau>p-1}(Z_{p-1} - Z_p) &= V_{N+1}^{n-1}\gamma_{N+1}\mathbf{1}_{\tau>N}Z_N - \gamma_n\mathbf{1}_{\tau>n-1}Z_n\\ &\quad + \sum_{q=N+1}^{n-1}\left[V_{q+1}^{n-1}\gamma_{q+1}\mathbf{1}_{\tau>q} - V_q^{n-1}\gamma_q\mathbf{1}_{\tau>q-1}\right]Z_q\\ &= V_{N+1}^{n-1}\gamma_{N+1}\mathbf{1}_{\tau>N}Z_N - \gamma_n\mathbf{1}_{\tau>n-1}Z_n\\ &\quad + \sum_{q=N+1}^{n-1}\left[V_{q+1}^{n-1}\gamma_{q+1} - V_q^{n-1}\gamma_q\right]\mathbf{1}_{\tau>q}Z_q\\ &\quad - \sum_{q=N+1}^{n-1} V_q^{n-1}\gamma_q Z_q\mathbf{1}_{q>\tau>q-1}\end{aligned}$$

We have, for large enough N and $n > N$:

$$V_N^{n-1}\gamma_N \le V_{N+1}^{n-1}\gamma_{N+1} \le \cdots \le V_{n-1}^{n-1}\gamma_{n-1} \le \gamma_n\,.$$

(we skip the easy checking of this result; the assumption $2a\beta_R < 1$ is needed when $b = 1$). Moreover, when $\tau > q-1$, lemma 3 implies that $|Z_q|$ is bounded by a constant $D(R)$. This yields

$$\begin{aligned}\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p\mathbf{1}_{\tau>p-1}(Z_{p-1} - Z_p) &\le 2D(R)\gamma_n + D(R)\left[\gamma_n - V_{N+1}^{n-1}\gamma_{N+1}\right]\\ &\quad +\gamma_n D(R)\sum_{q=N+1}^{n-1}\mathbf{1}_{q>\tau>q-1}\\ &\le 4D(R)\gamma_n\end{aligned}$$

so that we have proved that, for some constant $D(R)$,

$$\alpha_n \le D(R)\gamma_n + V_N^{n-1}\alpha_N + D(R)\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p^2$$

But, letting

$$\alpha'_n = V_N^{n-1}\alpha_N + D(R)\sum_{p=N+1}^{n} V_p^{n-1}\gamma_p^2$$

for a large enough constant $D'(R)$, we have $\alpha'_n \le D(R)\alpha_n$. Indeed, if $\alpha_n \le D'\gamma_n$,

$$\begin{aligned} \alpha'_{n+1} &= (1-2\beta_R\gamma_{n+1})\alpha'_n + D\gamma_{n+1}^2 \\ &\le \gamma_{n+1}(D' + (D-2D'\beta_R)\gamma_{n+1}) \\ &\ge \gamma_{n+1}(D' + (D-D'/a)\gamma_{n+1}) \end{aligned}$$

so that, if $D' > aD$, and is chosen so that $\alpha_N \le D'\gamma_N$, we get $\alpha'_n \le D'\gamma_n$ for all n, which yields the statement of the lemma.

□

Lemma 8 and (49) trivially imply that, for all $\epsilon > 0$, there exists $A > 0$ such that, for all N, $P(U^N(0) > A) \le \epsilon$, so that the family of the distributions of $U^N(0)$ is tight. Assumption (L9) implies that all the eigenvalues of $h'(\hat\theta) + \delta_b(b)/2a$ are negative so that the diffusion with generator L_t has a stationary Gaussian distribution. Theorem 3 and a compactness argument (cf. [1], lemma II-4-14) imply that $U^N(0)$ must converge in distribution towards this stationary Gaussian distribution, which is centered and with variance $S(\hat\theta)$.

In practice, this central limit theorem may be used to obtain an indicator of the fact that θ_N^0 has reached a stationnary regime, in which the effect of the O.D.E has become comparable with the diffusion noise. In such a case, we know that the convergence speed will be in $\sqrt{\gamma_n} = \sqrt{a}n^{-b/2}$. A good idea at this point would be to increase b, wait again for stabilization, getting in that way a faster convergence in the asymptotics. This procedure may obviously be iterated.

8. Proof of theorem 2.

The proof is partially adapted from [1] (part 2, chapter 4).

Like in section 5, D will represent a generic constant, $D(R)$, or $D(R,T)$ indicating that it only depends on R or on R and T. Moreover, the notation ω^N will indicate some sequence depending on N, such that $\lim_{N\to\infty} = 0$. There again, we shall not trace the explicit value of ω^N, which may change from line to line.

8.1. Preliminary results.

8.1.1. *Bounds on the sequence k_p^N.*

Lemma 9. *If $\gamma_p^N = a(N+p)^{-b}$, with $1/2 < b \le 1$, $l_n^N = (n+N)^{1/b_1}$ with $0 < b_1 < b/2$, $\overline{\gamma}_p^N = (N+p)^{-b_2}$ with $b_2 \ge 1 - b/2$ and*

$$\frac{1}{\sqrt{k_p^N}} - \frac{1}{\sqrt{k_{p-1}^N}} \le \overline{\gamma}_p^N$$

if there exists no n such that $p = l_n$, we have, for some constant D depending on b, b_1, b_2,

$$\sum_{p=n_0}^{n} \gamma_p^N \left| \frac{1}{k_p^N} - \frac{1}{k_{p-1}^N} \right| \le D.\max(n_0, N)^{-b/2}$$

Proof of lemma 9. We are in the situation of lemma 2, and, since we have

$$\left|\frac{1}{k_p^N} - \frac{1}{k_{p-1}^N}\right| \leq 2\left|\frac{1}{\sqrt{k_p^N}} - \frac{1}{\sqrt{k_{p-1}^N}}\right|$$

we can use inequality (25). This yields

$$\sum_{p=n_0}^{n} \gamma_p^N \left|\frac{1}{k_p^N} - \frac{1}{k_{p-1}^N}\right| \leq 2(n_0+N)^{-b} + 2\sum_{q\geq n_0}(q+N)^{-b-b_2} + 4\sum_{(q+N)^{1/b_1}\geq n_0}[(q+N)^{1/b_1}+N]^{-b}$$

The first sum is bounded by $D(n_0+N)^{1-b-b_2}$. The second one may be compared to the integral

$$\begin{aligned}\int_{n_0^{b_1}-N}^{\infty}\left[(t+N)^{\frac{1}{b_1}}+N\right]^{-b}dt &= \int_{n_0^{b_1}}^{\infty}[t^{\frac{1}{b_1}}+N]^{-b}dt \\ &= N^{b_1-b}\int_{(\frac{n_0}{N})^{b_1}}^{\infty}[u^{\frac{1}{b_1}}+1]^{-b}du\end{aligned}$$

with the change of variables $t = N^{b_1}u$. This may be bounded either by

$$N^{b_1-b}\int_0^{\infty}[u^{\frac{1}{b_1}}+1]^{-b}du$$

or by

$$N^{b_1-b}\int_{(\frac{n_0}{N})^{b_1}}^{\infty}u^{\frac{-b}{b_1}}du = \frac{b}{b-b_1}n_0^{b_1-b}$$

which yields the bound $D\max(n_0,N)^{b_1-b}$ for some constant D depending on b and b_1 and ends the proof of the lemma. □

8.1.2. *Bound on* $\sum\xi_p^N$. Let

$$\xi_{p+1}^N = \frac{1}{k_n^N}\sum_{q=1}^{k_p^N}[f(\theta_p^N, Y^{N,p+1,q}) - h(\theta_p^N)]\,.$$

The next lemma can be obtained similarly to equation (6).

Lemma 10. *We have*

$$\mathbf{E}\left[\sup_{n\geq n_0}\left|\sum_{p=n_0}^{n}\gamma_{p+1}^N d_p^{N,R}\xi_{p+1}^N\right|^2\right] \leq D(\gamma_{n_0}^N + \sum_{n\geq n_0}(\gamma_n^N)^2 + \sum_{q\geq n_0}\gamma_p^N\overline{\gamma}_q^B + \sum_{q,n_0+1\leq l_q\leq n}\gamma_{l_q}^N)^2\,,$$

where $d_p^{N,R} = \mathbf{1}[\theta_q \leq R, q\leq p]$.

In the present case, the upper bound is smaller that $D\max(n_0,N)^{-b/2}$ since we are in the conditions of lemma 9.

8.1.3. *Approximation by the ODE.* The following result is true for all fixed N. We have defined $\tau_R = \tau_R^N$ to be the stopping time

$$\inf\{n\geq 0; |\theta_n^N - \hat{\theta}| > R\}\,.$$

Lemma 11. *For all* $R>0$ *and* $T>0$, *we have, with a suitable constant* $D_o(T,R)$

$$\mathbf{E}\left[\sup_{n\leq\tau_R\wedge m^N(T)}|\theta_n^N - \overline{\theta}(t_n^N)|^2\right] \leq D_o(T,R)\left[\Gamma^N + E\left(\mathbf{1}_{[\tau_R>0]}|\theta_0^N - \overline{\theta}_0|^2\right)\right) \tag{51}$$

with

$$\Gamma^N = \gamma_1^N + \sum_{q\geq 1}(\gamma_n^N)^2 + \sum_{q\geq 1}\gamma_q^N\overline{\gamma}_q^N\sum_{q,1\leq l_q^N\leq n}\gamma_{l_q^N}^N)^2)$$

Note that under condition (L3)', $\Gamma^N \leq N^{-b/2}$.

Proof of lemma 11 This is to relate to ([1]: theorem II-1-9). We have

$$\overline{\theta}(t_n) - \overline{\theta}(t_{n-1}) = \int_{t_{n-1}}^{t_n} h(\overline{\theta}(u))du$$

so that

$$\theta_n - \overline{\theta}(t_n) = \theta_{n-1} - \overline{\theta}(t_{n-1}) + \int_{t_{n-1}}^{t_n} (h(\theta_{n-1}) - h(\overline{\theta}(u)))du + \gamma_n \xi_n \,.$$

Denote by d_n^R the indicator function of the set $n < \tau_R$. We have

$$d_n^R(\theta_n - \overline{\theta}(t_n)) = d_n^R(\theta_{n-1} - \overline{\theta}(t_{n-1})) + d_n^R \int_{t_{n-1}}^{t_n} (h(\theta_{n-1}) - h(\overline{\theta}(u)))du + d_n^R \gamma_n \xi_n \,.$$

so that

$$\begin{aligned} d_n^R(\theta_n - \overline{\theta}(t_n)) \quad &= \quad d_{p+1}^R(\theta_p - \overline{\theta}(t_p)) + \sum_{q=p}^{n-1} d_{q+1}^R \int_{t_q}^{t_{q+1}} (h(\theta_q) - h(\overline{\theta}(u)))du \\ & \quad + \sum_{q=p}^{n-1} d_{q+1}^R \gamma_{q+1} \xi_{q+1} \end{aligned}$$

When $p \leq \tau_R$, θ_p is bounded, and since $\hat{\theta}$ is asymptotically stable, $\theta(.\overline{\theta}_0)$ is also bounded. This implies that there exists a constant $D(R)$ such that

$$d_{q+1}^R |h(\theta_q) - h(\overline{\theta}(u), \overline{\theta}_0)| \leq D(R) d_{q+1}^R |\theta_q - \overline{\theta}(u, \overline{\theta}_0)| \,.$$

Since $d_{q+1}^R \leq d_q^R$, we get

$$\begin{aligned} d_n^R |\theta_n - \overline{\theta}(t_n)| \quad &\leq \quad d_p^R |\theta_p - \overline{\theta}(t_p)| + D \sum_{q=p}^{n-1} d_q^R \int_{t_q}^{t_{q+1}} |\theta_q - \overline{\theta}(u)| du \\ & \quad + \quad |\sum_{q=p}^{n-1} \gamma_{q+1} d_{q+1}^R \xi_{q+1}| \end{aligned}$$

Now, if $u \in [t_q, t_{q+1}[$, we have

$$|\,|\theta_q - \overline{\theta}(u)| - |\theta_q - \overline{\theta}(t_q)|\,| \leq |\overline{\theta}(u) - \overline{\theta}(t_q)| \leq D\gamma_{q+1}$$

so that

$$\begin{aligned} d_n^R |\theta_n - \overline{\theta}(t_n)| \quad &\leq \quad d_p^R |\theta_p - \overline{\theta}(t_p)| + D \sum_{q=p}^{n-1} d_q^R \gamma_{q+1} |\theta_q - \overline{\theta}(t_q)| \\ & \quad + \quad D^2 \sum_{q=p}^{n-1} \gamma_{q+1}^2 + |\sum_{q=p}^{n-1} \gamma_{q+1} d_R^{q+1} \xi_{q+1}| \end{aligned}$$

and, letting $V_n = d_n^R|\theta_n - \overline{\theta}(t_n)|$,

$$\begin{aligned} V_n^2 &\leq 3V_p^2 + 3D\left[\sum_{q=p}^{n-1} \gamma_{q+1} V_q\right]^2 \\ &\quad +3\left[D^2 \sum_{q=p}^{n-1} \gamma_{q+1}^2 + |\sum_{q=p}^{n-1} \gamma_{q+1} d_R^{q+1} \xi_{q+1}|\right]^2 \\ &\leq 3V_p^2 + 3D\left(\sum_{q=p}^{n-1} \gamma_{q+1}\right) \sum_{q=p}^{n-1} \gamma_{q+1} V_q^2 \\ &\quad +3\left[D^2 \sum_{q=p}^{n-1} \gamma_{q+1}^2 + |\sum_{q=p}^{n-1} \gamma_{q+1} d_R^{q+1} \xi_{q+1}|\right]^2 \end{aligned}$$

from which we deduce,

$$\begin{aligned} E[\sup_{h\leq n} V_h^2] &\leq 3E(V_0^2) + 3DT \sum_{q=0}^{n-1} \gamma_{q+1} E(\sup_{h\leq q} V_q^2) \\ &\quad +3E\left[D^2 \sum_{q=0}^{n-1} \gamma_{q+1}^2 + |\sum_{q=0}^{n-1} \gamma_{q+1} d_R^{q+1} \xi_{q+1}|\right]^2 \end{aligned}$$

So, letting $\alpha_q = E(\sup_{h\leq q} V_q^2)$, we get

$$\alpha_n \leq 3E(V_0^2) + 3DT \sum_{q=0}^{n-1} \gamma_{q+1}\alpha_q + D(\gamma_1 + \sum_{q\geq 1} \gamma_n^2 + \sum_{q\geq 1} \gamma_p^N \overline{\gamma}_q^B \sum_{q,1\leq l_q\leq n} \gamma_{l_q})^2$$

and a simple induction argument shows that this implies that, for some constant $D(T,R)$,

$$\alpha_n \leq D(E(V_0^2) + (\gamma_1 + \sum_{q\geq 1} \gamma_n^2 + \sum_{q,1\leq l_q\leq n} \gamma_{l_q})^2)$$

which finishes the proof of lemma 11 □

8.2. **Basic lemma.** We now prove what corresponds to ([1], prop. II-4-4).

Let $\tilde{\gamma}_n^N = \sqrt{\gamma_n^N}$. We fix a $T > 0$, and for all N, we let $m_T^N = \sup\{n, t_n^N \leq T\}$ and $\overline{T}^N = t_{m^N}$. We also denote by d_R^N the indicator function of the set $|\theta_0^N| \leq R$

Consider a sequence of random linear operators $(\Psi_n^N,\ n \geq 0)$, such that Ψ_n^N is $\mathcal{F}_n^N$-measurable, where $\mathcal{F}_n^N$ is the σ-algebra generated by the random variables $Y^{N,p',q'}$ for $p' \leq n$, $q' \leq k_{p'}$. For all n, N, we assume that Ψ_n^N is a linear operator from $\mathbb{R}^c$ to $\mathbb{R}^{c'}$ for some fixed c, c'. We furthermore assume that there exists a constant $D_\psi(T,R)$ such that, for all n,

$$d_R^N|\Psi_n^N| \leq D_\psi$$

and

$$d_R^N|\Psi_n^N - \Psi_{n-1}^N| \leq D_\psi \tilde{\gamma}_n^N .$$

Consider a family of functions u_θ, defined on Ω, with values in $\mathbb{R}^c$, such that u_θ is continuous in θ and $P_\theta u_\theta$ is continuously differentiable in θ. We shall bound the L^ν norm of sums of the kind:

$$\sum_{n\leq p\leq m} d_R^N \epsilon_p^N ,$$

with

$$\epsilon_p^N = \frac{\tilde{\gamma}_{p+1}^N}{k_p^N} \sum_{q=1}^{k_p^N} \Psi_p^N (u_{\theta_p^N}(Y^{N,p+1,q}) - P_{\theta_p^N} u_{\theta_p^N}(Y^{N,p+1,q})) .$$

We fix N in the computation that follows, and drop it from the notation. Like in the proof of theorems 1, 2, we write $\epsilon_p = \epsilon_p^{(1)} + \epsilon_p^{(2)}$, with

$$\epsilon_p^{(1)} = \frac{\tilde{\gamma}_{p+1}}{k_p^N} \sum_{q=1}^{k_p^N} \Psi_p(u_{\theta_p}(Y^{p+1,q}) - P_{\theta_p} u_{\theta_p}(Y^{p+1,q-1})) .$$

and

$$\epsilon_p^{(2)} = \frac{\tilde{\gamma}_{p+1}}{k_p^N} \Psi_p(P_{\theta_p} u_{\theta_p}(X^p) - P_{\theta_p} u_{\theta_p}(X^{p+1})) .$$

Inequality (42) implies that the function $\theta^N(t), t \leq T$ is bounded by $(1+\theta_0^N)e^{D_f T}$. Thus, there exists a constant $D_u(T,R)$ such that, for all N, for all $p \leq m^N(T)$, and, for all $x \in \Omega$,

$$d_R|\Delta u_{\theta_p}| \leq D_u(T,R) \tag{52}$$

$$d_R|P_{\theta_p} u_{\theta_p}(x) - P_{\theta_q} u_{\theta_q}(x)| \leq D_u(T,R)|\theta_p - \theta_q| \tag{53}$$

Since in turn, if $q \leq p$,

$$|\theta_p - \theta_q| \leq D_f(1+|\theta_q|)(e^{t_p - t_q} - 1)$$

we also have (with a possible change of the value of D_u), for $q, p \leq m^N(T)$,

$$d_R|P_{\theta_p} u_{\theta_p}(x) - P_{\theta_q} u_{\theta_q}(x)| \leq D_u(T,R)|t_p - t_q| \tag{54}$$

Let $\nu \geq 2$ be given. Since $d_R \epsilon_p^{(1)}$ is a martingale increment, we have (Burkholder 's inequality), with an universal constant D_ν,

$$\mathbf{E}\left(\sup_{n \leq p \leq m} \left| \sum_{p'=1}^{p} d_R \epsilon_{p'}^{(1)} \right|^\nu \right) \leq D_\nu \mathbf{E}[\sum_{p=n}^{m} (d_R \epsilon_p^{(1)})^2]^{\nu/2} .$$

Using (52), we bound $\mathbf{E}(d_R(\epsilon_p^{(1)})^2)$ by $\tilde{\gamma}_p^2 D_\psi^2 D_u^2 \mathbf{E}(1/k_p)$, so that

$$\mathbf{E}\left(\sup_{n \leq p \leq m} \left| \sum_{p'=1}^{p} \epsilon_{p'}^{(1)} \right|^\nu \right) \leq D_\nu D_\psi^\nu D_u^\nu |t_m - t_n|^{\nu/2} .$$

We now write (letting, for short $\hat{u}_p^q = P_{\theta_p} u_{\theta_p}(X^q)$,

$$\begin{aligned} \sum_{p=n}^{m} \epsilon_p^{(2)} &= \sum_{p=n+1}^{m-1} (\tilde{\gamma}_{p+1} - \tilde{\gamma}_p) \frac{1}{k_p} \Psi_p . \hat{u}_p^p \\ &+ \sum_{p=n+1}^{m-1} \tilde{\gamma}_p (\frac{1}{k_p} - \frac{1}{k_{p-1}}) \Psi_p . \hat{u}_p^p \\ &+ \sum_{p=n+1}^{m-1} \frac{\tilde{\gamma}_p}{k_{p-1}} \Psi_p . (\hat{u}_p^p - \hat{u}_{p-1}^p) \\ &+ \sum_{p=n+1}^{m-1} \frac{\tilde{\gamma}_p}{k_{p-1}} (\Psi_p - \Psi_{p-1}) . \hat{u}_{p-1}^p \\ &+ \frac{\tilde{\gamma}_{n+1}}{k_n} \Psi_n . \hat{u}_n^n - \frac{\tilde{\gamma}_{m+1}}{k_m} \Psi_m . \hat{u}_m^{m+1} \end{aligned}$$

Let us rapidly study the sums $A_i = d_R \sum_{p=n+1}^{m-1} |\tilde{\epsilon}_p^{(i)}|$, where $\tilde{\epsilon}_p^{(i)}$ is one of the terms involved in the 4 sums in the right hand side of the previous formula. For $i=1$ we can use the lemma (the proof is left to the reader)

Lemma 12. *The ratio $(\tilde{\gamma}_{p+1} - \tilde{\gamma}_p)/\tilde{\gamma}_p^3$ converges to 0 if $b < 1$ and to $1/2a$ if $b = 1$.*

In all cases, the previous ratio is bounded, so that, for some constant $D(T,R)$, A_1 is smaller than $D(T,R)\sum_{p=n+1}^{m-1}\gamma_p^{3/2}$ which is in turn smaller that $D(T,R)\tilde{\gamma}_n|t_n-t_m|$. Using (53), the same bound may be obtained for $i=3$.

For $i=2$, lemma 2, implies that

$$\sum_{p=n+1}^{m-1}\tilde{\gamma}_p(\frac{1}{k_p}-\frac{1}{k_{p-1}})\Psi_p.\hat{u}_p^p \leq \tilde{\gamma}_n+\sum_{q=n}^{m}\tilde{\gamma}_q\overline{\gamma}_q+\sum_{n\leq l_q\leq m}\tilde{\gamma}_{l_q}$$

with $\tilde{\gamma}_q=\sqrt{a}(q+N)^{-b/2}$, $\overline{\gamma}_q=(q+N)^{-b_2}$, we have $\tilde{\gamma}_q\overline{\gamma}_q=a^{-1/2}\gamma_q(q+N)^{b/2-b_2}$ so that

$$\sum_{q=n}^{m}\tilde{\gamma}_q\overline{\gamma}_q\leq a^{-1/2}\gamma_n^{\frac{1}{2}-\frac{b_2}{b}}(t_m-t_n)\,.$$

Since $l_q=(q+N)^{1/b_1}$, a similar argument as in the proof of lemma 9 yields the fact that

$$\begin{aligned}\sum_{n\leq l_q\leq m}\tilde{\gamma}_{l_q} &\leq \sqrt{a}N^{b_1-b/2}\int_{(\frac{n}{N})^{b_1}}^{(\frac{n}{N})^{b_1}}[u^{\frac{1}{b_1}}+1]^{-b/2}du\\ &\leq \sqrt{a}N^{b_1-b/2}\int_0^{\infty}[u^{\frac{1}{b_1}}+1]^{-b/2}du\end{aligned}$$

and this vanishes when N tends to infinity, since $b_1<b/2$.

The sum for $j=4$ is the only one of order $|t_n-t_m|$ (it is bounded by $D_uD_\psi|t_n-t_m|$). Finally, the last term is smaller than $D_uD_\psi\tilde{\gamma}_n$.

We summarize what we have shown in the next lemma, in which $B_{m,n}^N$ corresponds to the term A_4 above.

Lemma 13. *When $n\leq m\leq m_T^N$, the sum*

$$V_{n,m}^N=\sum_{n\leq p\leq m}\frac{\tilde{\gamma}_{p+1}^N}{k_p^N}\sum_{q=1}^{k_p^N}\Psi_p^N.(u_{\theta_p^N}(Y^{N,p+1,q})-P_{\theta_p^N}u_{\theta_p^N}(Y^{N,p+1,q}))$$

may be decomposed into three parts : $M_{n,m}^N+B_{n,m}^N+R_{n,m}^N$, with

$$M_{n,m}^N=\sum_{n\leq p\leq m}\frac{\tilde{\gamma}_{p+1}^N}{k_p^N}\sum_{q=1}^{k_p^N}\Psi_p^N(u_{\theta_p^N}(Y^{N,p+1,q})-P_{\theta_p^N}u_{\theta_p^N}(Y^{N,p+1,q-1}))\,,$$

$$B_{m,n}^N=\sum_{n\leq p\leq m}\frac{\tilde{\gamma}_{p+1}^N}{k_p}(\Psi_{p+1}^N-\Psi_p^N).P_{\theta_p}u_{\theta_p}(X^p)\,.$$

M^N is a martingale and

$$\mathbf{E}(d_R^N\sup_{n\leq p\leq m}|M_{n,p}^N|^\nu)\leq D_\nu D(T,R)|t_m-t_n|^{\nu/2}\,,$$

$$\sup_{n\leq p\leq m}d_R^N|B_{n,p}^N|\leq D(T,R)|t_m-t_n|\,,$$

and

$$sup_{n\leq p\leq m}d_R^N|R_{n,p}^N|\leq D(T,R)\omega^N(1+|t_m-t_n|)$$

with $\omega^N\to 0$ if $N\to\infty$.

8.3. **Tightness.** This lemma will be used to show that the sequence of processes U^N is tight (in the set of all processes which are right continuous, and have a left limit at all points). To prove this, it suffices to show that ([3])

i) For all fixed $t \in [0, T]$, the family of random variables $U^N(t)$ is tight.
ii) The oscillations of the functions U^N are uniformly bounded : for $0 < s < t \leq T$, denote

$$\omega^N(s,t) = \sup_{s \leq u < v \leq t} |U^N(u) - U^N(v)|.$$

The condition is that, for all $T > 0$, $\eta > 0$ and $\epsilon > 0$, there exists $\delta > 0$ and N_0 such that, for all $N \geq N_0$, for all $t < T$,

$$\mathbf{P}(\omega^N(t, t+\delta) > \eta) \leq \delta\epsilon \tag{55}$$

We therefore evaluate the variations of the function $U^N(t)$. Fix $t < t'$, and let n, m be such that $t_n^N \leq t \leq t_{n+1}^N$ and $t_m^N \leq t' \leq t_{m+1}^N$. Then,

$$U^N(t') - U^N(t) = \sum_{p=m}^{n-1} (U^N(t_{p+1}^N) - U^N(t_p^N)) .$$

Replacing U^N by its expression in function of θ_n and $\overline{\theta}$ yields.

$$\begin{aligned} U^N(t_{p+1}^N) - U(t_p^N) &= \frac{\theta_{p+1}^N - \overline{\theta}(t_{p+1}^N, \overline{\theta}_0)}{\sqrt{\gamma_{p+1}^N}} - \frac{\theta_p^N - \overline{\theta}(t_p^N, \overline{\theta}_0)}{\sqrt{\gamma_p^N}} \\ &= \frac{\sqrt{\gamma_{p+1}^N} - \sqrt{\gamma_p^N}}{\sqrt{\gamma_p^N}} U^N(t_p) \\ &\quad + \gamma_{p+1}^N \frac{h(\theta_p^N) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))}{\sqrt{\gamma_{p+1}^N}} \\ &\quad - \frac{1}{\sqrt{\gamma_{p+1}^N}} \int_{t_p^N}^{t_{p+1}^N} [h(\overline{\theta}(s, \overline{\theta}_0)) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))] ds \\ &\quad + \sqrt{\gamma_{p+1}^N} \xi_{p+1}^N , \end{aligned} \tag{56}$$

where $\xi_{p+1}^N = \frac{1}{k_p^N} \sum_{q=1}^{k_p^N} (\rho_{\theta_p^N}(Y^{N,p+1,q} - P_{\theta_p^N}\rho_{\theta_p^N}(Y^{N,p+1,q}))$.

From lemma 12, we know that, if $b < 1$ (resp. $b = 1$), the ratio

$$\frac{\sqrt{\gamma_{p+1}^N} - \sqrt{\gamma_p^N}}{(\gamma_p^N)^{3/2}}$$

tends to 0 (resp. $1/2a$) if N tends to infinity. We denote by $\delta_1(b)$ the Dirac fonction at $b = 1$.

Let d_R^N be the indicator of the set $|\theta^N(0)| \leq R$). By assumption (L8), $\overline{\theta}(t_p^N, \overline{\theta}_0)$ lies in a fixed compact set. Moreover, if $|\theta_0^N| \leq R$, because of equation (42), θ_n^N lies in a compact set which only depends on R and T provided $t_n^N \leq T$. Thus, there exists a constant $D(T, R)$ such that, for all p such atht $t_p^N \leq T$,

$$d_R^N \left| \frac{h(\theta_p^N) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))}{\sqrt{\gamma_{p+1}^N}} \right| \leq D(T, R) d_R^N |U^N(t_n^N)|$$

and

$$\frac{1}{\sqrt{\gamma_{p+1}^N}} \int_{t_p^N}^{t_{p+1}^N} [h(\overline{\theta}(s, \overline{\theta}_0)) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))] ds \leq D(\gamma_{p+1}^N)^{3/2}$$

Adding the obtained estimates, and returning to continuous-time notation yield

$$\sup_{t\le s\le t'} |U^N(s)-U^N(t)| \le D.\int_t^{t'} |U^N(s)|ds + D.\sup_{n\le p\le m} |\sum_{p'=n}^{p} \xi_{p'+1}^N| + \omega^N|t_m-t_n|. \tag{57}$$

when ω^N tends to 0 if N tends to $+\infty$. Indeed, since $|t_m-t_n| = \sum_n^m \gamma_p$, d^N can be bounded by

$$\left| \frac{\sqrt{\gamma_{n+1}^N}-\sqrt{\gamma_n^N}}{(\gamma_n^N)^{3/2}} - \frac{\delta_1(b)}{2a}\right| + \gamma_n^N$$

up to some multiplicative constant.

This implies in particular that

$$\begin{aligned}\mathbf{E}[\sup_{s\le t'} d_R^N|U^N(s)|^2] \quad \le \quad & D.\mathbf{E}(d_R^N|U^N(0)|^2) + D.\int_0^{t'} \mathbf{E}(d_R^N|U^N(s)|^2)ds \\ & +D.\mathbf{E}[\sup_{n\le p\le m} d_R^N|\sum_{p'=n}^{p}\xi_{p'+1}^N|^2] \\ & +D.\omega^N|t_m-t_n|^2\end{aligned} \tag{58}$$

Applying lemma 13 (with $\Psi_n^N=$ Identity) to the sum, we get

$$\begin{aligned}\mathbf{E}[\sup_{s\le t'} d_R^N|U^N(s)|^2] \quad \le \quad & D.\mathbf{E}(d_R^N|U^N(0)|^2) + D.\int_0^{t'} \mathbf{E}(d_R^N|U^N(s)|^2)ds \\ & +D.(|t-t'|+|t-t'|^2)+\omega^N\end{aligned}$$

with a possibly new sequence ω^N.

Since we have assumed that $\mathbf{E}(d_R^N|U^N(0)|^2)$ is bounded, an application of Gronwall's lemma implies that $M_T = \sup_N \mathbf{E}\left(\sup_{t\le T} d_R^N|U^N(t)|^2\right)$ is finite.

This implies that

$$\mathbf{P}\left[d_R^N\int_t^{t'}|U^N(s)|ds>\eta\right] \le \frac{|t-t'|}{\eta^2}\int_t^{t'}\mathbf{E}(d_R^N|U^N(s)|^2) \le \frac{|t-t'|^2}{\eta^2}M_T \tag{59}$$

from which we get

$$\mathbf{P}\left[\int_t^{t'}|U^N(s)|ds>\eta\right] \le \frac{|t-t'|^2}{\eta^2}M_T + \mathbf{P}(|\theta_0^N-\hat\theta|>R) \tag{60}$$

Another application of lemma 13 with $\nu=4$ gives

$$\mathbf{E}\left[\sup_{n\le p\le m} d_R^N|\sum_{p'=n}^{p}\xi_{p'+1}^N|^4\right] \le D(T,R)[|t-t'|^2+|t-t'|^4]+\omega^N$$

thus

$$\mathbf{P}[\sup_{n\le p\le m}|\sum_{p'=n}^{p}\xi_{p'+1}^N|>\eta] \le \frac{D(T)}{\eta^4}[|t-t'|^2+|t-t'|^4]+\omega^N+\mathbf{P}(|\theta_0^N-\hat\theta|>R) \tag{61}$$

It is now easy to prove (55), using (58), (60), (61). Indeed, let $\eta>0, \epsilon>0$ be given. Assume that N is large enough so that the sequence ω^N in (57) is smaller than $\eta/3T$. Then (57) implies that

$$\begin{aligned}\mathbf{P}\left[\sup_{t\le s\le t+\delta}|U^N(s)-U^N(t)|>\eta\right] \quad \le \quad & \mathbf{P}\left[\int_t^{t+\delta}|U^N(s)|ds>\frac{\eta}{3D}\right] \\ & +\mathbf{P}[\sup_{n\le p\le m}|\sum_{p'=n}^{p}\xi_{p'+1}^N|>\frac{\eta}{3D}]\end{aligned}$$

where n and m are such that $t_n^N \leq t < t_{n+1}^N$ and $t_m^N \leq t+\delta < t_{m+1}^N$. Equations (60) and (61) then imply that, for some constant D, and a sequence ω^N which tends to 0,

$$\mathbf{P}\left[\sup_{t\leq s\leq t+\delta} |U^N(s) - U^N(t)| > \eta\right] \leq D\delta^2 + \omega^N + \mathbf{P}(|\theta_0^N - \hat{\theta}| > R)$$

Take $R > R_1 + |\hat{\theta}|$ so that, by lemma 7, $\mathbf{P}(|\theta_0^N - \hat{\theta}| > R)$ tends to 0 when N tends to infinity. It suffices to take $\delta = \epsilon/3D$, N_0 such that $\omega^N + \mathbf{P}(|\theta_0^N - \hat{\theta}| > R) \leq \delta\epsilon/3$ for $N \geq N_0$ to obtain (55). Moreover, we have proved that

$$M_T(R) = \sup_N \mathbf{E}(\sup_{t<T} d_R^N |U^N(t)|^2) < \infty \tag{62}$$

which implies that

$$P(|U^N(t)| > \alpha) \leq \frac{M_T(R)}{\alpha^2} + P(|\theta_0^N - \hat{\theta}| > R)$$

which can be used to prove the tightness of $(U^N(t))$ for fixed t. We thus have proved

Proposition 1. *The sequence of processes $(U^N())$ is tight.*

This implies that

Lemma 14. *Let k_n^N and k_0 be as in condition (L3)'. We have*

$$\lim_{N\to\infty} \mathbf{P}\left[\exists n \geq 0, \quad k_n^N \neq k_0\right] = 0\,.$$

Proof of lemma 14: Assume that $k_0^N = k_0$. We will then have $k_n^N \equiv k_0$ as soon as $\theta_n^N < R_0$ for all n. By lemma 7, we have, with a probablity which tends to 1 when N tends to infinity, that $|\theta_0^N| < R_1$. We will have $\theta_n^N < R_0$ for all N as soon as $|\theta_n^N - \overline{\theta}(t_n^N, \overline{\theta}_0)| < R_0 - R_2$ that is

$$U^N(t_n^N) < \frac{R_0 - R_2}{\sqrt{\gamma_n^N}} \leq \frac{R_0 - R_2}{\sqrt{\gamma_1^N}}$$

we thus have

$$\mathbf{P}\left[\exists n \geq 0, \quad k_n^N \neq k_0\right] \leq \sqrt{\gamma_1^N}\frac{M_T(R_1 + |\hat{\theta}|)}{R_0 - R_2} + \mathbf{P}(|\theta_0^N| > R_1) + \mathbf{P}(k_0^N \neq k_0)$$

which tends to 0 when N tends to infinity. □

8.4. **Identification of the limit.** The end of the proof consists in showing that any weak limit of a converging subsequence of $U^N()$ is equal to a diffusion, independent of the subsequence. Denote by L_t the generator of the limit diffusion, which has to be identified . We show that any weak limit of the U^N is the solution of the martingale problem associated to L_t, ie. that for any ψ, which is C^∞ with compact support, the process

$$\left(\psi(U(t)) - \psi(U(0)) - \int_0^t L_s\psi(U(s))ds \ ; \ t \geq 0\right)$$

is a martingale. Following ([1]), one can check that it suffices to show that for all N, the following decomposition is valid, for any ψ

$$\psi(U^N(t)) = \psi(U^N(0)) + \int_0^t L_s\psi(U^N(s))ds + M^N(t) + r^N(t)\,, \tag{63}$$

where M^N is a martingale and, for some $R > 0$, for all t, $\mathbf{E}(d_R^N|r^N(t)|) \to 0$ when $N \to \infty$.

We thus study the variation $\psi(U^N(t)) - \psi(U^N(0))$ for $t \leq T$. Fix n such that $t_n^N \leq t < t_{n+1}^N$; one has

$$\begin{aligned}
\psi(U^N(t)) - \psi(U^N(0)) &= \sum_{p=0}^{n-1} \psi(U^N(t_{p+1}^N)) - \psi(U^N(t_p^N)) \\
&= \sum_{p=0}^{n-1} \left\langle \psi'(U^N(t_p^N))\,,\, U^N(t_{p+1}^N) - U^N(t_n^p) \right\rangle \\
&\quad + \frac{1}{2} \sum_{p=0}^{n-1} \left\langle \psi''(U^N(t_p^N))\,,\, [U^N(t_{p+1}^N) - U^N(t_n^p)]^{\otimes 2} \right\rangle \\
&\quad + \psi(U^N(t)) - \psi(U^N(t_n^N)) + r_1^N(t)
\end{aligned}$$

Note that ψ' is a linear form and ψ'' a quadratic form. The symbol $\otimes$ refers to the standard tensor product of two vectors: $a \otimes b$ is the matrix with $(a \otimes b)_i j = a_i b_j$, and when A and B are two matrices of same size, $\langle A\,,\, B\rangle$ is $\sum A_{ij} B_{ij}$.

Using the decomposition (56), it is easy to show that there exists a constant $D = D(T, R)$ such that

$$d_R^N |U^N(t_{p+1}^N) - U^N(t_n^p)| \leq D.\sqrt{\gamma_{p+1}^N} \sup_{t \leq T} |U^N(t)|\,.$$

This implies that the remainder $|r_1^N|$ is smaller than $D'.\sum_{p\leq n}(\gamma_{p+1}^N)^{3/2}\sup_{t\leq T}|U^N(t)|$, which is smaller than $D'\sqrt{\gamma_1^N}\sup_{t\leq T}|U^N(t)|$. Using (62), we can incorporate it into $r^N(t)$. Similarly, $\psi(U^N(t)) - \psi(U^N(t_n^N))$ is smaller than $\sqrt{\gamma_1^N}\sup_{t\leq T}|U^N(t)|$ and thus can also be incorporated into r^N. We now study the other terms.

8.4.1. *Gradient term.* Consider inequality (56). If in the upper bound, we replace

$$\frac{\sqrt{\gamma_{p+1}^N} - \sqrt{\gamma_p^N}}{\sqrt{\gamma_p^N}} U^N(t_p)$$

by $(\delta_1(b)/2a)\gamma_p^N U^N(t_p^N)$ the error is smaller than $\omega^N \gamma_p^N U^N(t_p^N)$. If we replace

$$\gamma_{p+1}^N \frac{h(\theta_p^N) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))}{\sqrt{\gamma_{p+1}^N}}$$

by $\gamma_{p+1}^N |U(t_p^N)|$, it is easily shown that the error is less than $\omega^N \gamma_{p+1}^N(\sup_{t\leq T}|U(t)| + \sup_{t\leq T}|U(t)|^2)$. Finally, the term

$$\frac{1}{\sqrt{\gamma_{p+1}^N}} \int_{t_p^N}^{t_{p+1}^N} [h(\overline{\theta}(s, \overline{\theta}_0)) - h(\overline{\theta}(t_p^N, \overline{\theta}_0))] ds$$

is smaller than $D(\gamma_{p+1}^N)^{3/2}$. Therefore, (56) can be rewritten as

$$U^N(t_{p+1}^N) - U(t_p^N) = \gamma_{p+1}^N \left(h'(\overline{\theta}(t_p^N, \overline{\theta}_0)) + \frac{\delta_1(b)}{2} I \right) U(t_p^N) \tag{64}$$

$$+ \sqrt{\gamma_{p+1}^N}\xi_{p+1}^N + \gamma_{p+1}^N r_{p+1}^N\,, \tag{65}$$

with $|r_{p+1}^N| < \omega^N(1 + \sup_{t\leq T}|U(t)| + \sup_{t\leq T}|U(t)|^2)$ and $\omega^N \to 0$.

Thus,

$$\sum_{p=0}^{n-1} \left\langle \psi'(U^N(t_p^N)) \,,\, U^N(t_{p+1}^N) - U^N(t_n^p) \right\rangle = \int_0^t \left\langle \psi'(U^N(u)) \,,\, \left(h'(\overline{\theta}(u, \overline{\theta}_0)) + \frac{\delta_1(b)}{2} I \right) U^N(u) \right\rangle du + \sum_{p=0}^{n-1} \left\langle \psi'[U^n(t_p^N)] \,,\, \xi_{p+1}^N \right\rangle + r_2^N(t)$$

We must include in $r_2^N(t)$ one term coming from the extension of the integral to t instead of t_n^N, the error being bounded by $\omega^N \sup_{t \leq T} |U^N(t)|$, and another term coming from the sum of the $\gamma_{p+1}^N r_{p+1}^N$, which is smaller than $T\omega^N(1 + \sup_{t\leq T} |U(t)| + \sup_{t\leq T})$. Thus r_2^N can be incorporated into r^N.

The term

$$\int_0^t \left\langle \psi'(U^N(u)) \,,\, \left(h'(\overline{\theta}(u, \overline{\theta}_0)) + \frac{\delta_1(b)}{2} I \right) U^N(u) \right\rangle du \tag{66}$$

constitutes the first part of

$$\int_0^t L_u \psi(U^N(u)) du \,.$$

We carry on the study to get the second part.

We apply lemma 13 to

$$\sum_{p=0}^{n-1} \left\langle \psi'[U^n(t_p^N)] \,,\, \xi_{p+1}^N \right\rangle$$

with $\psi_p^N(\,.\,) = \left\langle \psi'[U^n(t_p^N)] \,,\, . \right\rangle$ and $u_\theta = \rho_\theta$, as given in equation (26). This yields that the sum can be replaced by a martingale (to incorporate into M^N), a remainder which can be incorporated into r^N and the term

$$\sum_{p=0}^{n-1} \frac{\sqrt{\gamma_p^N}}{k_{p-1}^N} \left\langle \psi'(U^N(t_p^N)) - \psi'(U^N(t_{p-1}^N)) \,,\, P_{\theta_{p-1}} \rho_{\theta_{p-1}}(X^p) \right\rangle .$$

In this term, we can write

$$\psi'(U^N(t_p^N)) - \psi'(U^N(t_{p-1}^N)) = \psi''(U^N(t_{p-1}^N)).[U^N(t_p^N) - U^N(t_{p-1}^N)] + r_3^N$$

where $r_3^N \leq D \left| U^N(t_p^N) - U^N(t_{p-1}^N) \right|^2$. From decomposition (56), the only term which may bring a significant contribution is $\sqrt{\gamma_p^N} \xi_p^N$, since all the other are smaller than $\gamma_p^N \max_{t \leq T} |U^N(t)|$. So letting the remainder apart, we get

$$\sum_{p=0}^{n-1} \frac{\gamma_p^N}{k_{p-1}^N} \left\langle \psi''_{p-1} \,,\, \xi_p^N \otimes \hat{\rho}_{p-1}^p \right\rangle , \tag{67}$$

where we have introduced the notation $\psi''_p = \psi''(U^N(t_p^N))$,

$$\hat{\rho}_p^{p'} = P_{\theta_p^N} \rho_{\theta_p}(X_{p'}^N) \,.$$

Although we have dropped the exponent N to simplify the notation, these quantities do depend on N. We shall also set

$$\rho_p^{p'} = \rho_{\theta_p}(X^{p'}) \,,$$

and $\rho_p^{p',q}$, $\hat{\rho}_p^{p',q}$ when $X^{p'}$ is replaced by $Y^{p',q}$.

8.4.2. *Terms of second order.* We now consider

$$\frac{1}{2}\sum_{p=0}^{n-1}\left\langle \psi''_{p-1}\,,\,[U^N(t_p^N)-U^N(t_n^{p-1})]^{\otimes 2}\right\rangle .$$

Inequality (56) can be written

$$U^N(t_{p+1}^N)-U(t_p^N)=\sqrt{\gamma_{p+1}^N}\xi_{p+1}^N+\gamma_{p+1}^N K_{p+1}^N$$

with $K_{p+1}^N \le \sup_{t\le T}|U^N(t)|$. Thus, we have

$$[U^N(t_{p+1}^N)-U(t_p^N)]^{\otimes 2}=\gamma_{p+1}^N\,[\xi_{p+1}^N]^{\otimes 2}+\gamma_{p+1}^N\tilde{r}_{p+1}^N \tag{68}$$

with

$$d_R^N\langle \psi''_{p-1}\,,\,\tilde{r}_{p+1}^N\rangle \le D(T,R)\sqrt{\gamma_{p+1}^N}\sup_{t\le T}|U^N(t)|^2$$

We thus can incorporate $\sum \tilde{r}_{p+1}^N$ into r^N, the remainder being

$$\frac{1}{2}\sum_{p=0}^{n-1}\gamma_p^N\left\langle \psi''_{p-1}\,,\,[\xi_p^N]^{\otimes 2}\right\rangle \tag{69}$$

8.4.3. *Study of (67) and (69).* Adding these terms, we get

$$\frac{1}{2}\sum_{p=0}^{n-1}\gamma_p^N\left\langle \psi''_{p-1}\,,\,\xi_p^N\otimes(\xi_p^N+\frac{2}{k_{p-1}^N}\hat{\rho}_{p-1}^p)\right\rangle$$

Denote this last term by $G^N/2$. We drop the superscript N in the following computation. We have

$$\begin{aligned}\xi_p &= \frac{1}{k_{p-1}}\sum_{q=1}^{k_{p-1}}(\rho_{p-1}^{p,q}-\hat{\rho}_{p-1}^{p,q})\\ &= \frac{1}{k_{p-1}}\sum_{q=1}^{k_{p-1}}\rho_{p-1}^{p,q}-\hat{\rho}_{p-1}^{p,q-1})+\frac{\hat{\rho}_{p-1}^{p-1}}{k_{p-1}}-\frac{\hat{\rho}_{p-1}^{p}}{k_{p-1}}\end{aligned}$$

so that

$$\begin{aligned}G &= \sum_{p=0}^{n-1}\frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,[\hat{\rho}_{p-1}^{p-1}+\sum_{q=1}^{k_{p-1}}(\rho_{p-1}^{p,q}-\hat{\rho}_{p-1}^{p,q-1})]^{\otimes 2}\right\rangle\\ &\quad -\sum_{p=0}^{n-1}\frac{[\hat{\rho}_{p-1}^p]^{\otimes 2}}{k_{p-1}^2}\\ &= \sum_{p=0}^{n-1}\frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,[\sum_{q=1}^{k_{p-1}}(\rho_{p-1}^{p,q}-\hat{\rho}_{p-1}^{p,q-1})]^{\otimes 2}\right\rangle\\ &\quad +2\sum_{p=0}^{n-1}\frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\hat{\rho}_{p-1}^{p-1}\otimes\sum_{q=1}^{k_{p-1}}(\rho_{p-1}^{p,q}-\hat{\rho}_{p-1}^{p,q-1})\right\rangle\\ &\quad +\sum_{p=0}^{n-1}\frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,(\hat{\rho}_{p-1}^{p-1})^{\otimes 2}\right\rangle-\left\langle \psi''_{p-1}\,,\,(\hat{\rho}_{p-1}^{p})^{\otimes 2}\right\rangle\end{aligned}$$

Denote by V_p the term

$$\left\langle \psi''_p\,,\,[\sum_{q=1}^{k_p}(\rho_p^{p,q}-\hat{\rho}_p^{p,q-1})]^{\otimes 2}\right\rangle$$

We have

$$V_p = \sum_{q,q'=1}^{k_p} \left\langle \psi''_p\,,\,(\rho_p^{p,q} - \hat\rho_p^{p,q-1}) \otimes (\rho_{p-1}^{p,q'} - \hat\rho_p^{p,q'-1})\right\rangle$$

When computing the conditional expectation of V_p given $\mathcal{F}_{p-1}$, only the diagonal terms ($q = q'$) remain, and yield

$$\mathbf{E}(V_p \mid \mathcal{B}_{p_1}) = \sum_q \hat{\hat\rho}_{p-1}^{p,q} - \hat\rho_{p,q-1}^{\otimes 2}$$

where

$$\hat{\hat\rho}_p^{p',q} = P_{\theta_p}[\rho_{\theta_p}^{\otimes 2}](Y^{p',q})\,;$$

Therefore

$$G \;=\; \sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\sum_{q=1}^{k_{p-1}} (\rho_{p-1}^{p,q} - \hat\rho_{p-1}^{p,q-1})^{\otimes 2} - \sum_{q=1}^{k_{p-1}} \hat{\hat\rho}_{p-1}^{p,q} - \hat\rho_{p,q-1}^{\otimes 2}\right\rangle \tag{70}$$

$$+2\sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\hat\rho_{p-1}^{p-1} \otimes \sum_{q=1}^{k_{p-1}} (\rho_{p-1}^{p,q} - \hat\rho_{p-1}^{p,q-1})\right\rangle \tag{71}$$

$$+\sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\sum_{q=1}^{k_{p-1}} \hat{\hat\rho}_{p-1}^{p,q} - \hat\rho_{p,q-1}^{\otimes 2} + (\hat\rho_{p-1}^{p-1})^{\otimes 2} - (\hat\rho_{p-1}^{p})^{\otimes 2}\right\rangle \tag{72}$$

The sums in (70) and (71) are martingales, and the third one is

$$\sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\sum_{q=1}^{k_{p-1}} \hat{\hat\rho}_{p-1}^{p,q} - \hat\rho_{p,q}^{\otimes 2}\right\rangle \tag{73}$$

We now introduce the matrix $S(\theta)$, and the function w_θ defined by

$$w_\theta() - P_\theta w_\theta() = P_\theta(\rho_\theta^{\otimes 2})() - (P_\theta \rho_\theta)^{\otimes 2}() - S(\theta) \tag{74}$$

Note that $S(\theta)$ is well defined by (74) since the existence of $w(\theta)$ implies that

$$S(\theta) = \pi_\theta(P_\theta(\rho_\theta^{\otimes 2}) - (P_\theta \rho_\theta)^{\otimes 2})$$

Replacing ρ_θ by its expression in (26), one retrieves, after simplification, the expression given in (48)

$$S(\theta) = \pi_\theta(g_\theta^{\otimes 2}) + 2\sum_{p\geq 1} \pi_\theta(g_\theta \otimes P_\theta^k g_\theta) \tag{75}$$

and one can check that $S(\theta)$ is the normalized asymptotic variance of the empirical mean

$$\frac{1}{n}\sum_{l=0}^{n} g_\theta(Z^l)$$

(Z_l) being an outcome of the markov chain with transition P_θ.

Inserting (74) into (73), we obtain (with the same notation $w_p^{p'q}$, ...)

$$\sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}^2}\left\langle \psi''_{p-1}\,,\,\sum_{q=1}^{k_{p-1}} (w_{p-1}^{p,q} - \hat w_{p-1}^{p,q})\right\rangle \tag{76}$$

$$+\;\sum_{p=0}^{n-1} \frac{\gamma_p}{k_{p-1}}\langle \psi''_{p-1}\,,\,S(\theta_{p-1})\rangle \tag{77}$$

The term in (76) can be decomposed as a martingale plus negligible terms as before, and, concerning (77), it is (reintroducing N again)

$$\int_0^t \left\langle \psi''(U^N(s)) \,,\, \frac{1}{k^N(s)} S[\theta^N(s)] \right\rangle ds \tag{78}$$

$k^N(s)$ being the piecewise constant approximation of k_n^N.

Let W^N be the event: $\exists n \geq 0,\ k_n^N \neq k_0$. By lemma 14, $\mathbf{P}(W^N)$ tends to 0 when N tends to infinity. One may write:

$$\begin{aligned}\int_0^t \left\langle \psi''(U^N(s)) \,,\, \frac{1}{k^N(s)} S[\theta^N(s)] \right\rangle ds &= \int_0^t \left\langle \psi''(U^N(s)) \,,\, \frac{1}{k_0} S[\overline{\theta}(s, \overline{\theta}_0)] \right\rangle ds \\ &+ \mathbf{1}_{I^N} \int_0^t \left\langle \psi''(U^N(s)) \,,\, \left(\frac{1}{k^N(s)} - \frac{1}{k_0}\right) S[\overline{\theta}(s, \overline{\theta}_0)] \right\rangle ds \\ &+ \int_0^t \left\langle \psi''(U^N(s)) \,,\, \frac{1}{k_N(s)} \left\{ S[\theta^N(s)] - S[\overline{\theta}(s, \overline{\theta}_0)] \right\} \right\rangle ds\end{aligned}$$

Since

$$\int_0^t \left\langle \psi''(U^N(s)) \,,\, \left(\frac{1}{k^N(s)} - \frac{1}{k_0}\right) S[\overline{\theta}(s, \overline{\theta}_0)] \right\rangle ds$$

is bounded, the second term in the right-hand term may incorporated into r^N. By lemma 11, and because S is continuous (by the same argument as in lemma 3), the third term may also be put into r^N It thus only remains

$$\int_0^t \left\langle \psi''(U^N(s)) \,,\, \frac{1}{k_0} S(\overline{\theta}(s, \overline{\theta}_0)) \right\rangle ds\,. \tag{79}$$

We thus have completed formula (63); formulae (66) and (78) yield the generator:

$$L_t(\psi)(U) = \left\langle \psi'(U) \,,\, h'(\overline{\theta}(t, \overline{\theta}_0))U + \frac{\delta_1(b)}{2a} U \right\rangle + \frac{1}{2} \left\langle \psi''(U) \,,\, \frac{S(\overline{\theta}(t, \overline{\theta}_0))}{k_0} \right\rangle .$$

9. Conclusion.

In this paper, we have obtained almost sure convergence results for Markovian stochastic algorithms under different conditions on the gain and the approximation of the O.D.E. In particular, we have shown that, in order to obtain convergence with little restriction on the size of γ_n, the averaging time k_n has to be adapted to the ergodicity of the Markov chain with transition P_{θ_n}. Such an *adaptive rule* provides an algorithm which reacts to situations in which the current parameter would accidentally reach values far away from the limit by improving the precision of the gradient approximation.

We moreover have proved that, for large gains, a diffusion approximation of the tail of the algorithm may be obtained. This provides us with running strategies as well as stopping rules which are needed in practice.

References

[1] A. Benveniste, M. Métivier, P.Priouret: *Adaptive algorithms and stochatic approximations* Springer-Verlag (1990)
Original french edition: *Algorithmes Adaptatifs et Approximations Stochastiques, Théorie et Application. Techniques Stochastiques.* Masson (1987)

[2] J. Besag (1974) : Spatial Interaction and the Statistical Analysis of Lattice Systems. *J. of Roy. Stat. Soc.* B-36 pp 192-236.

[3] P. Billingsley (1968) : *Convergence of probability measures.* Wiley series in probability and mathematical statistics.

[4] C. Bouton: *Approximation gaussienne d'algorithmes stochastiques à dynamique markovienne.* Thésis Univ. Paris 6

[5] R.L. Dobrushin (1968) : The description of a random field by means of conditional probabilities and conditions of its regularity. *Thry. Prob. Appl.*, 13 197-224.

[6] R.L. Dobrushin (1970) : Prescribing a system of random variables by conditional distributions. *Thry. Prob. Appl.*, 15 458-486.

[7] R. M. Dudley (1989) : *Real analysis and Probability* Wadworth publishing company.

[8] Paul Dupuis and Rahul Simha (1991) : On sampling controlled stochastic approximation. *IEEE Trans. Autom. Control*, Vol 36, No 8.

[9] A. Frigessi, F. Martinelli and J. Standler: Computational complexity of Markov chain Monte Carlo methods (preprint).

[10] D. Geman (1991) :*Random Fields and Inverse Problems in Imaging,* In *Proceedings of the Ecole d'été de Saint-Flour*, Lecture Notes in Mathematics 1427, Springer Verlag, New York.

[11] S. and D. Geman (1984): Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images. *IEEE TPAMI.* Vol PAMI-6 pp 721-741.

[12] D. Geman (1990) :*Random Fields and Inverse Problems in Imaging,* In *Proceedings of the Ecole d'été de Saint-Flour*, Lecture Notes in Mathematics 1427, Springer Verlag, New York.

[13] C.J. Geyer and E.A. Thompson : Constrained Monte Carlo Maximum likelihood for dependent data (with discussion) J. of Royal Stat. Soc. B-54, 657-699

[14] S. Ingrassia (1993) : Geometric approach to the estimation of the spectral gap of reversible markov chains. To appear in *Combinatorics, Probability and Computing*

[15] S. Ingrassia (1993) : On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds. To appear in *Annals of Applied Prob.*

[16] H.J. Kushner and D.S. Clark (1978) : *Stochastic approximation for constrained and unconstrained systems.* Applied math. science series, vol 26, Springer.

[17] M. Métivier and P. Priouret (1987) : Théorèmes de Convergence p.s. pour une Classe d'Algorithmes Stochastiques à Pas Décroissants. *Prob. Th. Rel. Fields* 74, pp. 403-428.

[18] H. Robbins and S. Monro (1951) : A Stochastic approximation method. *Annals of Math. Stat.* 22, 400-407.

[19] A. Sokal (1989) : Monte Carlo Methods in Statistical Mechanics. Lecture notes: Cours de troisième cycle de la physique en Suisse Romande (Lausanne).

[20] Stroock and Zegarlinski (1992) : The logarithmic Sobolev inequality for discrete Spin systems on a lattice. *Comm. Math. Phys.* 149, 175-194.

[21] L. Younes : Estimation and annealing for Gibbsian fields. *Annales de l'Institut Henri Poincaré*, vol 2.

[22] L. Younes (1989) : Parameter estimation for imperfectly observed Gibbsian fields. *Prob. Theory and Rel. fields* 82, 625–645.

[23] Maximum likelihood estimation for Gibbs fields. *Spatial Statistics and Imaging: Proceedings of an AMS-IMS-SIAM Joint Summer Research Conference*, A. Possolo (editor), Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California (1991).

[24] L. Younes (1992) : Parameter estimation for imperfectly observed Gibbs fields and some comments on Chalmond's EM Gibssian algorithm. *Proceedings of "stochastic models, statistical methods and algorithms in image analysis"*, Ed. P. Barone and A. Frigessi, Lecture Notes in Statistics, Springer.

CMLA (URA 1611), Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94 235 Cachan CEDEX, E-mail : younes@cmla.ens-cachan.fr