

# Optimal scaling for the transient phase of Metropolis Hastings algorithms: The longtime behavior

BENJAMIN JOURDAIN\*, TONY LELIÈVRE\*\* and BŁAŻEJ MIASOJEDOW†

*Université Paris-Est, CERMICS, 6 & 8, avenue Blaise Pascal, 77455 Marne-La-Vallée, France.*  
E-mail: \*[jourdain@cermics.enpc.fr](mailto:jourdain@cermics.enpc.fr); \*\*[lelievre@cermics.enpc.fr](mailto:lelievre@cermics.enpc.fr); †[bmia@mimuw.edu.pl](mailto:bmia@mimuw.edu.pl)

We consider the Random Walk Metropolis algorithm on  $\mathbb{R}^n$  with Gaussian proposals, and when the target probability measure is the  $n$ -fold product of a one-dimensional law. It is well known (see Roberts *et al.* (*Ann. Appl. Probab.* 7 (1997) 110–120)) that, in the limit  $n \rightarrow \infty$ , starting at equilibrium and for an appropriate scaling of the variance and of the timescale as a function of the dimension  $n$ , a diffusive limit is obtained for each component of the Markov chain. In Jourdain *et al.* (Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit (2012) Preprint), we generalize this result when the initial distribution is not the target probability measure. The obtained diffusive limit is the solution to a stochastic differential equation nonlinear in the sense of McKean. In the present paper, we prove convergence to equilibrium for this equation. We discuss practical counterparts in order to optimize the variance of the proposal distribution to accelerate convergence to equilibrium. Our analysis confirms the interest of the constant acceptance rate strategy (with acceptance rate between 1/4 and 1/3) first suggested in Roberts *et al.* (*Ann. Appl. Probab.* 7 (1997) 110–120).

We also address scaling of the Metropolis-Adjusted Langevin Algorithm. When starting at equilibrium, a diffusive limit for an optimal scaling of the variance is obtained in Roberts and Rosenthal (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* 60 (1998) 255–268). In the transient case, we obtain formally that the optimal variance scales very differently in  $n$  depending on the sign of a moment of the distribution, which vanishes at equilibrium. This suggest that it is difficult to derive practical recommendations for MALA from such asymptotic results.

*Keywords:* diffusion limits; MALA; optimal scaling; propagation of chaos; random walk Metropolis

## 1. Introduction

Many Markov Chain Monte Carlo (MCMC) methods are based on the Metropolis–Hastings algorithm, see [13, 16]. To set up the notation, let us recall this well-known sampling technique. Let us consider a target probability distribution on  $\mathbb{R}^n$  with density  $p$ . Starting from an initial random variable  $X_0$ , the Metropolis–Hastings algorithm generates iteratively a Markov chain  $(X_k)_{k \geq 0}$  in two steps. At time  $k$ , given  $X_k$ , a candidate  $Y_{k+1}$  is sampled using a proposal distribution with density  $q(X_k, y)$ . Then, the proposal  $Y_{k+1}$  is accepted with probability  $\alpha(X_k, Y_{k+1})$ , where

$$\alpha(x, y) = 1 \wedge \frac{p(y)q(y, x)}{p(x)q(x, y)}.$$

Here and in the following, we use the standard notation  $a \wedge b = \min(a, b)$ . If the proposed value is accepted, then  $X_{k+1} = Y_{k+1}$  otherwise  $X_{k+1} = X_k$ . The Markov Chain  $(X_k)_{k \geq 0}$  is by construction reversible with respect to the target density  $p$ , and thus admits  $p$  as an invariant distribution. The efficiency of this algorithm highly depends on the choice of the proposal distribution  $q$ . One common choice is a Gaussian proposal centered at point  $x \in \mathbb{R}^n$  with variance  $\sigma^2 \text{Id}_{n \times n}$ :

$$q(x, y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{|x-y|^2}{2\sigma^2}\right).$$

Since the proposal is symmetric ( $q(x, y) = q(y, x)$ ), the acceptance probability reduces to

$$\alpha(x, y) = 1 \wedge \frac{p(y)}{p(x)}. \quad (1)$$

Metropolis–Hastings algorithms with symmetric kernels are called random walk Metropolis (RWM). Another popular choice yields the so called Metropolis adjusted Langevin algorithm (MALA). In this case, the proposal distribution is a Gaussian random variable with variance  $\sigma^2 \text{Id}_{n \times n}$  and centered at point  $x + \frac{\sigma^2}{2} \nabla \ln(p(x))$  (in particular, it is not symmetric). It corresponds to one step of a time-discretization with timestep  $\sigma^2$  of the (overdamped) Langevin dynamics:  $dX_t = dB_t + \frac{1}{2} \nabla \ln p(X_t) dt$  which is ergodic with respect to  $p(x) dx$  (here,  $B_t$  is a standard  $n$ -dimensional Brownian motion).

In both cases (RWM and MALA), the variance  $\sigma^2$  remains to be chosen. It should be sufficiently large to ensure a good exploration of the state space, but not too large otherwise the rejection rate becomes typically very high since the proposed moves fall in low probability regions, in particular in high dimension. It is expected that the higher the dimension, the smaller the variance of the proposal should be. The first theoretical results to optimize the choice of  $\sigma^2$  in terms of the dimension  $n$  can be found in [22]. The authors study the RWM algorithm under two fundamental (and somewhat restrictive) assumptions: (i) the target probability distribution is the  $n$ -fold tensor product of a one-dimensional density:

$$p(x) = \prod_{i=1}^n \exp(-V(x_i)), \quad (2)$$

where  $x = (x_1, \dots, x_n)$  and  $\int_{\mathbb{R}} \exp(-V) = 1$ , and (ii) the initial distribution is the target probability (what we refer to as *the stationarity assumption* in the following):

$$X_0^n \sim p(x) dx.$$

The superscript  $n$  in the Markov chain  $(X_k^n)_{k \geq 0}$  explicitly indicates the dependency on the dimension  $n$ . Then, under additional regularity assumption on  $V$ , the authors prove that for a proper scaling of the variance as a function of the dimension, namely

$$\sigma_n^2 = \frac{\ell^2}{n},$$

where  $\ell$  is a fixed constant, the Markov process  $(X_{[nt]}^{1,n})_{t \geq 0}$  (where  $X_k^{1,n} \in \mathbb{R}$  denotes the first component of  $X_k^n \in \mathbb{R}^n$ ) converges in law to a diffusion process:

$$dX_t = \sqrt{h(I, \ell)} dB_t - h(I, \ell) \frac{1}{2} V'(X_t) dt, \tag{3}$$

where

$$h(I, \ell) = 2\ell^2 \Phi\left(-\frac{\ell\sqrt{I}}{2}\right) \quad \text{and} \quad I = \int_{\mathbb{R}} (V')^2 \exp(-V). \tag{4}$$

Here and in the following,  $[\cdot]$  denotes the integer part (for  $y \in \mathbb{R}$ ,  $[y] \in \mathbb{Z}$  and  $[y] \leq y < [y] + 1$ ) and  $\Phi$  is the cumulative distribution function of the normal distribution ( $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-y^2/2) dy$ ). The scalings of the variance and of the time as a function of the dimension are indications on how to make the RWM algorithm efficient in high dimension. Moreover, a practical counterpart of this result is that  $\ell$  should be chosen such that  $h(I, \ell)$  is maximum (the optimal value of  $\ell$  is  $\frac{c_0}{\sqrt{I}}$  with  $c_0 \simeq 2.38$ ), in order to optimize the time scaling in (3). This optimal value of  $\ell$  corresponds equivalently to a constant average acceptance rate, with approximate value 0.234: for this choice of  $\ell$ , in the limit  $n$  large,

$$\iint \alpha(x, y) p(x) q(x, y) dx dy \simeq 0.234.$$

Notice that the optimal average acceptance rate does not depend on  $I$ , and is thus the same whatever the target probability. Thus, the practical way to choose  $\sigma^2$  is to scale it in such a way that the average acceptance rate is roughly 1/4. Similar results have been obtained for the MALA algorithm in [23]. In this case, the scaling for the variance is  $\sigma_n^2 = \frac{\ell^2}{n^{1/3}}$ , the time scaling is  $(X_{[n^{1/3}t]}^n)_{t \geq 0}$  and the optimal average acceptance rate is 0.574.

There exists several extensions of such results for various Metropolis–Hastings algorithms, see [6–8,10,17,18,23,24], and some of them relax in particular the first main assumption mentioned above about the separated form of the target distribution, see [4,5,9,11]. Extensions to infinite dimensional settings have also been explored, see [9,15,21].

All these results assume stationarity: the initial measure is the target probability. To the best of the authors’ knowledge, the only works which deal with a nonstationary case are [12] where the RWM and the MALA algorithms are analyzed in the Gaussian case and [20]. In the latter paper, the target measure is assumed to be absolutely continuous with respect to the law of an infinite dimensional Gaussian random field and this measure is approximated in a space of dimension  $n$  where the MCMC algorithm is performed. The authors consider a modified RWM algorithm (called preconditioned Crank–Nicolson walk) started at a deterministic initial condition and prove that when  $\sigma_n$  tends to 0 as  $n$  tends to  $\infty$  (with no restriction on the rate of convergence of  $\sigma_n$  to 0), the rescaled algorithm converges to a stochastic partial differential equation, started at the same initial condition.

The aim of this article is to discuss extensions of the previous results for the RWM and the MALA algorithms, *without assuming stationarity*. The main findings are the following.

Concerning the RWM algorithm, in the companion paper [14], we prove that, using the same scaling for the variance and the time as in the stationary case (namely  $\sigma_n^2 = \frac{\ell^2}{n}$  and considering  $(X_{[nt]}^{1,n})_{t \geq 0}$ ), one obtains in the limit  $n$  goes to infinity a diffusion process nonlinear in the sense

of McKean (see Equation (7) below). This is discussed in Section 2. Contrary to (3), this diffusion process cannot be obtained from the simple Langevin dynamics  $dX_t = dB_t - \frac{V'(X_t)}{2} dt$  by a deterministic time-change and its long-time behavior is not obvious. In Section 3, we first prove that its unique stationary distribution is  $e^{-V(x)} dx$ . Assuming that this measure satisfies a logarithmic Sobolev inequality, we prove that the Kullback–Leibler divergence of the marginal distribution at time  $t$  with respect to  $e^{-V(x)} dx$  converges to 0 at an exponential rate. In Section 4, we discuss optimizing strategies which take into account the transient phase. Roughly speaking, the usual strategy which consists in choosing  $\ell$  (recall that  $\sigma_n^2 = \frac{\ell^2}{n}$ ) such that the average acceptance rate is constant (with value between 1/4 and 1/3) seems to be a very good strategy. This is numerically illustrated in Section 5.

Concerning the MALA algorithm, the situation is much more complicated. The scaling to be used seems to depend on the sign of an average quantity (see Section 6.1.3). In particular, the scaling  $\sigma_n^2 = \frac{\ell^2}{n^{1/3}}$  which has been identified in [23] under the stationary assumption is not adapted to the transient phase. It seems difficult to draw any practical recommendation from this analysis. This is explained with details in Section 6.

## 2. Scaling limit for the RWM algorithm

In this section, we state the diffusion limit for the RWM algorithm, and explain formally why this result holds. A rigorous proof can be found in [14].

### 2.1. The RWM algorithm and the convergence result

We consider a Random Walk Metropolis algorithm using Gaussian proposal with variance  $\sigma_n^2 = \frac{\ell^2}{n}$ , and with target  $p$  defined by (2). The Markov chain generated using this algorithm writes:

$$X_{k+1}^{i,n} = X_k^{i,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^i 1_{\mathcal{A}_{k+1}}, \quad 1 \leq i \leq n \tag{5}$$

with

$$\mathcal{A}_{k+1} = \left\{ U_{k+1} \leq e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + (\ell/\sqrt{n})G_{k+1}^i))} \right\},$$

where  $(G_k^i)_{i,k \geq 1}$  is a sequence of independent and identically distributed (i.i.d.) normal random variables independent from a sequence  $(U_k)_{k \geq 1}$  of i.i.d. random variables with uniform law on  $[0, 1]$ . We assume that the initial positions  $(X_0^{1,n}, \dots, X_0^{n,n})$  are exchangeable (namely the law of the vector is invariant under permutation of the indices) and independent from all the previous random variables. Exchangeability is preserved by the dynamics: for all  $k \geq 1$ ,  $(X_k^{1,n}, \dots, X_k^{n,n})$  are exchangeable. We denote by  $\mathcal{F}_k^n$  the sigma field generated by  $(X_0^{1,n}, \dots, X_0^{n,n})$  and  $(G_l^1, \dots, G_l^n, U_l)_{1 \leq l \leq k}$ .

For  $t > 0$  and  $i \in \{1, \dots, n\}$ , let

$$\begin{aligned} Y_t^{i,n} &= (\lceil nt \rceil - nt) X_{\lfloor nt \rfloor}^{i,n} + (nt - \lfloor nt \rfloor) X_{\lceil nt \rceil}^{i,n} \\ &= X_{\lfloor nt \rfloor}^{i,n} + (nt - \lfloor nt \rfloor) \frac{\ell}{\sqrt{n}} G_{\lceil nt \rceil}^i 1_{\mathcal{A}_{\lceil nt \rceil}} \end{aligned}$$

be the linear interpolation of the Markov chain obtained by rescaling time (the characteristic time scale is  $1/n$ ). This is the classical diffusive time-scale for a random walk, since the variance of each move is of order  $1/n$ .

Let us define the notion of convergence (namely the propagation of chaos) that will be useful to study the convergence of the interacting particle system  $((Y_t^{1,n}, \dots, Y_t^{n,n})_{t \geq 0})_{n \geq 1}$  in the limit  $n$  goes to infinity.

**Definition 1.** Let  $E$  be a separable metric space. A sequence  $(\chi_1^n, \dots, \chi_n^n)_{n \geq 1}$  of exchangeable  $E^n$ -valued random variables is said to be  $\nu$ -chaotic where  $\nu$  is a probability measure on  $E$  if for fixed  $j \in \mathbb{N}^*$ , the law of  $(\chi_1^n, \dots, \chi_j^n)$  converges in distribution to  $\nu^{\otimes j}$  as  $n$  goes to  $\infty$ .

According to [25], Proposition 2.2, the  $\nu$ -chaoticity is equivalent to a law of large numbers result, namely the convergence in probability of the empirical measures  $\mu^n = \frac{1}{n} \sum_{i=1}^n \delta_{\chi_i^n}$  to the constant  $\nu$  when the space of probability measures on  $E$  is endowed with the weak convergence topology.

We are now in position to state the convergence result for the RWM algorithm, taken from [14]. Here and in the following, the bracket notation refers to the duality bracket for probability measures on  $\mathbb{R}$ : for  $\mu$  a probability measure and  $\phi$  a bounded measurable function,

$$\langle \mu, \phi \rangle = \int \phi \, d\mu.$$

**Theorem 1.** Let  $m$  be a probability measure on  $\mathbb{R}$  such that  $\langle m, (V')^4 \rangle < +\infty$ . Let us also assume that

$$V \text{ is a } C^3 \text{ function on } \mathbb{R} \text{ with bounded second and third order derivatives.} \tag{6}$$

If the initial positions  $(X_0^{1,n}, \dots, X_0^{n,n})_{n \geq 1}$  are  $m$ -chaotic and such that

$$\sup_n \mathbb{E}[(V'(X_0^{1,n}))^4] < +\infty,$$

then the processes  $((Y_t^{1,n}, \dots, Y_t^{n,n})_{t \geq 0})_{n \geq 1}$  are  $P$ -chaotic where  $P$  denotes the law (on the space  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$  of continuous function with values in  $\mathbb{R}$ ) of the unique solution to the stochastic differential equation nonlinear in the sense of McKean

$$\begin{aligned} dX_t &= \Gamma^{1/2} (\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell) dB_t \\ &\quad - \mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell) V'(X_t) dt, \end{aligned} \tag{7}$$

where  $(B_t)_{t \geq 1}$  is a Brownian motion independent from the initial position  $X_0$  distributed according to  $m$ . The functions  $\Gamma$  and  $\mathcal{G}$  are, respectively, defined by: for  $\ell \in (0, +\infty)$ ,  $a \in [0, +\infty]$  and  $b \in \mathbb{R}$ ,

$$\Gamma(a, b, \ell) = \begin{cases} \ell^2 \Phi\left(-\frac{\ell b}{2\sqrt{a}}\right) + \ell^2 e^{\ell^2(a-b)/2} \Phi\left(\ell\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)\right), & \text{if } a \in (0, +\infty), \\ \frac{\ell^2}{2}, & \text{if } a = +\infty, \\ \ell^2 e^{-\ell^2 b^+/2}, & \text{if } a = 0, \end{cases} \quad (8)$$

where  $b^+ = \max(b, 0)$ , and

$$\mathcal{G}(a, b, \ell) = \begin{cases} \ell^2 e^{\ell^2(a-b)/2} \Phi\left(\ell\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)\right), & \text{if } a \in (0, +\infty), \\ 0, & \text{if } a = +\infty, \\ 1_{\{b>0\}} \ell^2 e^{-\ell^2 b/2}, & \text{if } a = 0. \end{cases} \quad (9)$$

Notice that the assumption on  $(X_0^{1,n}, \dots, X_0^{n,n})_{n \geq 1}$  is for example satisfied when the random variables  $X_0^{1,n}, \dots, X_0^{n,n}$  are i.i.d. according to some probability measure  $m$  on  $\mathbb{R}$ .

This convergence result generalizes the previous result by Roberts *et al.* [22] where the same diffusive limit is obtained under the restrictive assumption that the vector of initial positions  $(X_0^{1,n}, \dots, X_0^{n,n})$  is distributed according to the target distribution  $p(x) dx$ . In this case,  $(X_t)_{t \geq 0}$  indeed solves the stochastic differential equation (3)–(4) with time-homogeneous coefficients (here, we use the fact that  $\Gamma(I, I, \ell) = 2\mathcal{G}(I, I, \ell) = h(I, \ell)$  where  $I = \int_{\mathbb{R}} (V'(x))^2 e^{-V(x)} dx = \int_{\mathbb{R}} V''(x) e^{-V(x)} dx < \infty$ , see [14], Lemma 1). Moreover, by taking  $V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi)$ , this theorem also yields similar results as [12], where the authors consider a nonstationary case, but restrict their analysis to the evolution of  $k \mapsto \frac{1}{n} \sum_{i=1}^n (X_k^{i,n})^2$  for Gaussian targets.

In addition to the previous convergence result, we are able to identify the limiting average acceptance rate.

**Proposition 1.** *Under the assumptions of Theorem 1, the function*

$$t \mapsto \mathbb{E} \left| \mathbb{P}(\mathcal{A}_{\lfloor nt \rfloor + 1} | \mathcal{F}_{\lfloor nt \rfloor}^n) - \frac{1}{\ell^2} \Gamma(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell) \right|$$

converges locally uniformly to 0 and in particular, the average acceptance rate  $t \mapsto \mathbb{P}(\mathcal{A}_{\lfloor nt \rfloor + 1})$  converges locally uniformly to  $t \mapsto \text{acc}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell)$  where for any  $a \geq 0$  and  $b \in \mathbb{R}$ ,

$$\text{acc}(a, b, \ell) = \frac{\Gamma(a, b, \ell)}{\ell^2}. \quad (10)$$

### 2.2. A formal derivation of the limiting process (7)

Let us introduce the infinitesimal generator  $L_\mu$  associated to (7):

$$L_\mu \varphi(x) = \frac{1}{2} \Gamma(\langle \mu, (V')^2 \rangle, \langle \mu, V'' \rangle, \ell) \varphi''(x) - \mathcal{G}(\langle \mu, (V')^2 \rangle, \langle \mu, V'' \rangle, \ell) V'(x) \varphi'(x). \tag{11}$$

For a probability measure  $\mu$  on  $\mathbb{R}$ ,  $\langle \mu, V'' \rangle$  is well defined by boundedness of  $V''$  (see (6)), and  $\langle \mu, (V')^2 \rangle$  is also well defined in  $[0, +\infty]$ .

The relationship between (7) and (11) is the following: if  $X_t$  satisfies (7), then for any smooth test function  $\varphi$ ,  $\varphi(X_t) - \int_0^t L_{P_s} \varphi(X_s) ds$  is a martingale, where  $P_t$  denotes the law of  $X_t$ : for any  $s < t$ ,

$$\mathbb{E} \left( \varphi(X_t) - \int_s^t L_{P_r} \varphi(X_r) dr \mid \mathcal{F}_s \right) = \varphi(X_s). \tag{12}$$

Actually, as explained in [14], Section 3.1, the martingale representation of the solution is a weak formulation of (7): solutions to (12) are solutions in distribution to (7).

Let us now present formally how (7) is derived. First, let us explain how the scaling of  $\sigma_n$  as a function of  $n$  is chosen. The idea (see [24]) is to choose  $\sigma_n$  in such a way that the limiting acceptance rate (when  $n \rightarrow \infty$ ) is neither zero nor one. In the first case, this would mean that the variance of the proposal is too large, so that all proposed moves are rejected. In the second case, the variance of the proposal is too small, and the rate of convergence to equilibrium is thus not optimal. In particular, it is easy to check that  $\sigma_n$  should go to zero as  $n$  goes to infinity. Now, notice that the limiting acceptance rate is:

$$\begin{aligned} \mathbb{E}(1_{\mathcal{A}_{k+1}} \mid \mathcal{F}_k^n) &= \mathbb{E}(e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + \sigma_n G_{k+1}^i))} \wedge 1 \mid \mathcal{F}_k^n) \\ &= \mathbb{E}(e^{-\sum_{i=1}^n (V'(X_k^{i,n}) \sigma_n G_{k+1}^i + V''(X_k^{i,n}) (\sigma_n^2 / 2))} \wedge 1 \mid \mathcal{F}_k^n) + \mathcal{O}(n \sigma_n^3) + \mathcal{O}(\sqrt{n} \sigma_n^2) \\ &= \exp\left(\frac{a_n - b_n}{2}\right) \Phi\left(\frac{b_n}{2\sqrt{a_n}} - \sqrt{a_n}\right) + \Phi\left(-\frac{b_n}{2\sqrt{a_n}}\right) \\ &\quad + \mathcal{O}(n \sigma_n^3) + \mathcal{O}(\sqrt{n} \sigma_n^2), \end{aligned} \tag{13}$$

where  $a_n = \sigma_n^2 \sum_{i=1}^n (V'(X_k^{i,n}))^2$  and  $b_n = \sigma_n^2 \sum_{i=1}^n V''(X_k^{i,n})$ . The formula (13) is obtained by explicit computations (see [22], Proposition 2.4). From this expression, assuming a propagation of chaos (law of large numbers) on the random variables  $(X_k^{i,n})_{1 \leq i \leq n}$ , one can check that the correct scaling for the variance is  $\sigma_n^2 = \frac{\ell^2}{n}$  in order to obtain a nontrivial limiting acceptance rate (see [14], Section 2.3).

Using this scaling, we observe that, for a test function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} &\mathbb{E}(\varphi(X_{k+1}^{1,n}) \mid \mathcal{F}_k^n) \\ &= \mathbb{E} \left( \varphi \left( X_k^{1,n} + \frac{\ell}{\sqrt{n}} G_{k+1}^1 1_{\mathcal{A}_{k+1}} \right) \mid \mathcal{F}_k^n \right) \end{aligned}$$

$$\begin{aligned}
 &= \varphi(X_k^{1,n}) + \varphi'(X_k^{1,n}) \frac{\ell}{\sqrt{n}} \mathbb{E}(G_{k+1}^1 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) \\
 &\quad + \frac{\ell^2}{2n} \varphi''(X_k^{1,n}) \mathbb{E}((G_{k+1}^1)^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) + \mathcal{O}(n^{-3/2}).
 \end{aligned} \tag{14}$$

We compute (by conditioning with respect to  $G_{k+1}^1$ ):

$$\begin{aligned}
 &\mathbb{E}(G_{k+1}^1 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) \\
 &= \mathbb{E}(G_{k+1}^1 e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + (\ell/\sqrt{n})G_{k+1}^i))} \wedge 1 | \mathcal{F}_k^n) \\
 &= \mathbb{E}(G_{k+1}^1 e^{-\sum_{i=1}^n (V'(X_k^{i,n})(\ell/\sqrt{n})G_{k+1}^i + V''(X_k^{i,n})(\ell^2/(2n)))} \wedge 1 | \mathcal{F}_k^n) + \mathcal{O}(n^{-1}) \\
 &= -V'(X_k^{1,n}) \frac{1}{\ell\sqrt{n}} \mathcal{G}(\langle v_k^n, (V')^2 \rangle, \langle v_k^n, V'' \rangle, \ell) + \mathcal{O}(n^{-1}),
 \end{aligned} \tag{15}$$

where

$$v_k^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}$$

denotes the empirical distribution associated to the interacting particle system. The equation (15) is again a consequence of explicit computations (see [14], Equation (A.3)), and the fact that the remainder is of order  $n^{-1}$  requires a detailed analysis (see [14], Proposition 7). Likewise, for the diffusion term, we get

$$\begin{aligned}
 &\mathbb{E}((G_{k+1}^1)^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) \\
 &= \mathbb{E}((G_{k+1}^1)^2 e^{\sum_{i=1}^n (V(X_k^{i,n}) - V(X_k^{i,n} + (\ell/\sqrt{n})G_{k+1}^i))} \wedge 1 | \mathcal{F}_k^n) \\
 &= \mathbb{E}((G_{k+1}^1)^2 e^{-\sum_{i=1}^n (V'(X_k^{i,n})(\ell/\sqrt{n})G_{k+1}^i + V''(X_k^{i,n})(\ell^2/(2n)))} \wedge 1 | \mathcal{F}_k^n) + \mathcal{O}(n^{-1}) \\
 &= \frac{1}{\ell^2} \Gamma(\langle v_k^n, (V')^2 \rangle, \langle v_k^n, V'' \rangle, \ell) + \mathcal{O}(n^{-1}).
 \end{aligned} \tag{16}$$

To obtain (16), we again used an explicit computation (see [14], Equation (A.5)).

By plugging (15) and (16) into (14), we see that the correct scaling in time is to consider  $Y_t^{i,n}$  such that  $Y_{k/n}^{i,n} = X_k^{i,n}$  (diffusive timescale), and we get:

$$\begin{aligned}
 \mathbb{E}(\varphi(Y_{(k+1)/n}^{1,n}) | \mathcal{F}_k^n) &= \varphi(Y_{k/n}^{1,n}) - \varphi'(Y_{k/n}^{1,n}) \frac{1}{n} V'(Y_{k/n}^{1,n}) \mathcal{G}(\langle v_k^n, (V')^2 \rangle, \langle v_k^n, V'' \rangle, \ell) \\
 &\quad + \frac{1}{2n} \varphi''(Y_{k/n}^{1,n}) \Gamma(\langle v_k^n, (V')^2 \rangle, \langle v_k^n, V'' \rangle, \ell) + \mathcal{O}(n^{-3/2}) \\
 &= \varphi(Y_{k/n}^{1,n}) + \frac{1}{n} (L_{v_k^n} \varphi)(Y_{k/n}^{1,n}) + \mathcal{O}(n^{-3/2}),
 \end{aligned}$$



where  $L_\mu$  is defined by (11). This can be seen as a discrete-in-time version (over a timestep of size  $1/n$ ) of the martingale property (12). Thus, by sending  $n$  to infinity, assuming that  $v_k^n$  converges to the law of  $Y_{k/n}^1$ , we expect  $Y_t^{1,n}$  to converge to a solution to (7). For a rigorous proof, we refer to [14].

### 3. Longtime convergence for the RWM nonlinear dynamics

We would like to study the limiting dynamics (7) obtained for the RWM algorithm, that we recall for convenience

$$dX_t = \Gamma^{1/2}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell) dB_t - \mathcal{G}(\mathbb{E}[(V'(X_t))^2], \mathbb{E}[V''(X_t)], \ell) V'(X_t) dt,$$

where  $\Gamma$  and  $\mathcal{G}$  are, respectively, defined by (8) and (9). The associated Fokker–Planck equation is ( $\psi_t$  denotes the density of the random variable  $X_t$ ):

$$\begin{cases} \partial_t \psi_t = \partial_x (\mathcal{G}(a[\psi_t], b[\psi_t], \ell) V' \psi_t + \Gamma(a[\psi_t], b[\psi_t], \ell) \partial_x \psi_t / 2), \\ \text{where } a[\psi_t] = \int_{\mathbb{R}} (V')^2 \psi_t \quad \text{and} \quad b[\psi_t] = \int_{\mathbb{R}} V'' \psi_t. \end{cases} \tag{17}$$

Let us denote  $\psi_\infty = \exp(-V)$ . Notice that  $a[\psi_\infty] = b[\psi_\infty]$  and  $\mathcal{G}(a, a, \ell) = \Gamma(a, a, \ell)/2$ . We thus expect  $\psi_\infty$  to be the longtime limit of  $\psi_t$ .

#### 3.1. Stationary solution

We start the analysis of the limiting process by checking that the solution of (7) has the expected stationary distribution.

**Proposition 2.** *There exists a unique stationary distribution  $\mu$  for the process  $X_t$  defined by (7). In addition, this distribution is absolutely continuous with respect to the Lebesgue measure, with density  $\psi_\infty(x) = \exp(-V(x))$ .*

Before proving Proposition 2, we need some preliminary facts the proof of which is postponed to Appendix A.

**Lemma 1.** *Defining the function sign by:*

$$\text{sign}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -1, & \text{if } x < 0, \end{cases} \tag{18}$$

one has

$$\text{sign}(\Gamma(a, b, \ell) - 2\mathcal{G}(a, b, \ell)) = \text{sign}(a - b). \tag{19}$$

Moreover, the function  $F$  defined for  $a \geq 0, b \in \mathbb{R}$  and  $\ell > 0$  by

$$F(a, b, \ell) = \begin{cases} \frac{b\Gamma(a, b, \ell) - 2a\mathcal{G}(a, b, \ell)}{b - a}, & \text{if } a \neq b, \\ 2\ell^2 \left( \left( 1 + \frac{\ell^2 a}{4} \right) \Phi \left( -\frac{\ell\sqrt{a}}{2} \right) - \frac{\ell\sqrt{a}}{2\sqrt{2\pi}} \exp \left( -\frac{\ell^2 a}{8} \right) \right), & \text{if } a = b, \end{cases} \tag{20}$$

is a continuous function satisfying

$$\forall \ell > 0, \forall M \in (0, +\infty), \inf_{(a,b) \in [0,M] \times [-M,M]} F(a, b, \ell) > 0. \tag{21}$$

**Proof of Proposition 2.** Let  $c = \int_{\mathbb{R}} (V'(x))^2 \psi_{\infty}(x) dx$ . Since  $V''$  is bounded then one can check that  $c = \int_{\mathbb{R}} V''(x) \psi_{\infty}(x) dx < \infty$  (see [14], Lemma 1). By (19), we get that  $\Gamma(c, c, \ell) = 2\mathcal{G}(c, c, \ell)$ . Let us define the Langevin diffusion

$$d\tilde{X}_t = \sqrt{2\mathcal{G}(c, c, \ell)} dB_t - \mathcal{G}(c, c, \ell) V'(\tilde{X}_t) dt$$

with  $X_0$  distributed according to the density  $\psi_{\infty}$ . It is well known that for any  $t \geq 0$  the density of  $\tilde{X}_t$  is  $\psi_{\infty}$  and therefore  $c = \mathbb{E}[(V'(\tilde{X}_t))^2] = \mathbb{E}[V''(\tilde{X}_t)]$ . Then it is clear that the process  $\tilde{X}_t$  satisfies (7). Hence,  $\psi_{\infty}(x) dx$  is a stationary probability distribution for the stochastic differential equation (7).

Let us now prove the uniqueness of the invariant measure. Assume that there exists another stationary probability measure with density  $p_{\infty}$  (the fact that the stationary measure admits a density is standard, since the diffusion term is bounded from below). Assume  $\int_{\mathbb{R}} V'^2 p_{\infty} = +\infty$ . Since  $\mathcal{G}(+\infty, b, \ell) = 0$  and  $\Gamma(+\infty, b, \ell) = \frac{\ell^2}{2}$ , the stochastic differential equation (7) with  $X_0$  distributed according to the density  $p_{\infty}$  reduces in this case to  $dX_t = \frac{\ell}{\sqrt{2}} dB_t$  which does not admit a stationary distribution. Thus, necessarily, we have

$$\int_{\mathbb{R}} V'^2 p_{\infty} < \infty.$$

Let us denote  $a = \int_{\mathbb{R}} V'^2 p_{\infty}$  and  $b = \int_{\mathbb{R}} V'' p_{\infty}$ . Then, Equation (7) with  $X_0$  distributed according to the density  $p_{\infty}$  reduces to

$$dX_t = \Gamma^{1/2}(a, b, \ell) dB_t - \mathcal{G}(a, b, \ell) V'(X_t) dt.$$

The stationary distribution thus writes

$$p_{\infty} \propto \exp \left( -\frac{2\mathcal{G}(a, b, \ell)}{\Gamma(a, b, \ell)} V \right).$$

By integration by parts, we obtain that

$$b\Gamma(a, b, \ell) = 2a\mathcal{G}(a, b, \ell).$$

Hence, by definition of  $F$  and (21), we obtain  $a = b$  and by (19) we get that  $\frac{2\mathcal{G}(a,b,\ell)}{\Gamma(a,b,\ell)} = 1$ . In conclusion,  $p_{\infty} = \exp(-V) = \psi_{\infty}$ . □

### 3.2. Longtime convergence

It is actually possible to prove that, for fixed  $\ell > 0$ , the law of  $X_t$  solution to (7) converges exponentially fast to the equilibrium density  $\psi_\infty$ . The proof is based on entropy estimates, using the Fokker–Planck equation (17), and requires the notion of logarithmic Sobolev inequality.

**Definition 2.** *The probability measure  $\nu$  satisfies a logarithmic Sobolev inequality with constant  $\rho > 0$  (in short LSI( $\rho$ )) if and only if, for any probability measure  $\mu$  absolutely continuous with respect to  $\nu$ ,*

$$H(\mu|\nu) \leq \frac{1}{2\rho} I(\mu|\nu), \tag{22}$$

where  $H(\mu|\nu) = \int \ln(\frac{d\mu}{d\nu}) d\mu$  is the Kullback–Leibler divergence (or relative entropy) of  $\mu$  with respect to  $\nu$  and  $I(\mu|\nu) = \int |\nabla \ln(\frac{d\mu}{d\nu})|^2 d\mu$  is the Fisher information of  $\mu$  with respect to  $\nu$ .

With a slight abuse of notation, we will denote in the following  $H(\psi|\phi)$  and  $I(\psi|\phi)$  the Kullback–Leibler divergence and the Fisher information associated with the continuous probability distributions  $\psi(x) dx$  and  $\phi(x) dx$ . We recall that, by the Csiszar–Kullback inequality (see, for instance, [2], Théorème 8.2.7, page 139), for any probability densities  $\psi$  and  $\phi$ ,

$$\int_{\mathbb{R}} |\psi - \phi| \leq \sqrt{2H(\psi|\phi)}, \tag{23}$$

so that  $H(\psi|\phi)$  may be seen as a measure of the “distance” between  $\psi$  and  $\phi$ .

**Theorem 2.** *Let us assume (6), and that  $X_0$  admits a density  $\psi_0$  such that  $\mathbb{E}[(V'(X_0))^2] < +\infty$  and  $H(\psi_0|\psi_\infty) < \infty$ . Then, for all  $t \geq 0$ ,*

$$\frac{d}{dt} H(\psi_t|\psi_\infty) \leq -\frac{F(a[\psi_t], b[\psi_t], \ell)}{2} I(\psi_t|\psi_\infty), \tag{24}$$

and the function  $t \mapsto H(\psi_t|\psi_\infty)$  is decreasing.

Let us assume moreover that  $\psi_\infty = e^{-V}$  satisfies a LSI( $\rho$ ). Then there exists a positive and nonincreasing function  $\lambda : [0, +\infty) \rightarrow (0, +\infty)$  such that  $\forall t \geq 0$

$$H(\psi_t|\psi_\infty) \leq e^{-t\lambda(H(\psi_0|\psi_\infty))} H(\psi_0|\psi_\infty). \tag{25}$$

Equation (25) shows that  $\psi_t$  converges exponentially fast to  $\psi_\infty$ .

**Remark 1.** Roughly speaking,  $\psi_\infty$  satisfies a LSI if  $V$  grows sufficiently fast at infinity. For example, according to [2], Théorème 6.4.3, a sufficient condition for  $\psi_\infty$  to satisfy a LSI, is that  $|V'|$  does not vanish outside of some compact subset of  $\mathbb{R}$  and

$$\lim_{|x| \rightarrow \infty} \frac{V''(x)}{(V'(x))^2} = 0 \quad \text{and} \quad \limsup_{|x| \rightarrow \infty} \frac{|V(x) + \ln |V'(x)||}{(V'(x))^2} < +\infty.$$

In the Gaussian case  $V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi)$ ,  $\psi_\infty(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$  satisfies LSI(1).

**Proof of Theorem 2.** By simple computation, we have (for notational convenience, we write  $a, b$  for  $a[\psi_t], b[\psi_t]$ ):

$$\begin{aligned}
\frac{d}{dt} \int_{\mathbb{R}} \psi_t \ln(\psi_t/\psi_\infty) &= \int_{\mathbb{R}} \partial_t \psi_t \ln \psi_t + \int_{\mathbb{R}} V \partial_t \psi_t \\
&= \int_{\mathbb{R}} \partial_x (\mathcal{G}(a, b, \ell) V' \psi_t + \Gamma(a, b, \ell) \partial_x \psi_t / 2) \ln \psi_t \\
&\quad + \int_{\mathbb{R}} V \partial_x (\mathcal{G}(a, b, \ell) V' \psi_t + \Gamma(a, b, \ell) \partial_x \psi_t / 2) \\
&= -\mathcal{G}(a, b, \ell) \int_{\mathbb{R}} V' \partial_x \psi_t - (\Gamma(a, b, \ell) / 2) \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t \\
&\quad - \mathcal{G}(a, b, \ell) \int_{\mathbb{R}} (V')^2 \psi_t - (\Gamma(a, b, \ell) / 2) \int_{\mathbb{R}} V' \partial_x \psi_t \\
&= \mathcal{G}(a, b, \ell) b - (\Gamma(a, b, \ell) / 2) \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t \\
&\quad - \mathcal{G}(a, b, \ell) a + \Gamma(a, b, \ell) b / 2.
\end{aligned} \tag{26}$$

On the other hand, we have

$$\begin{aligned}
\int_{\mathbb{R}} (\partial_x \ln(\psi_t/\psi_\infty))^2 \psi_t &= \int_{\mathbb{R}} (\partial_x \ln \psi_t + V')^2 \psi_t \\
&= \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t + 2 \int_{\mathbb{R}} (\partial_x \ln \psi_t) V' \psi_t + \int_{\mathbb{R}} (V')^2 \psi_t \\
&= \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t - 2b + a.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
&\frac{d}{dt} \int_{\mathbb{R}} \psi_t \ln(\psi_t/\psi_\infty) \\
&= \mathcal{G}(a, b, \ell) b - (\Gamma(a, b, \ell) / 2) \left[ \int_{\mathbb{R}} (\partial_x \ln(\psi_t/\psi_\infty))^2 \psi_t + 2b - a \right] \\
&\quad - \mathcal{G}(a, b, \ell) a + \Gamma(a, b, \ell) b / 2 \\
&= -(\Gamma(a, b, \ell) / 2) \int_{\mathbb{R}} (\partial_x \ln(\psi_t/\psi_\infty))^2 \psi_t + (b - a) (\mathcal{G}(a, b, \ell) - \Gamma(a, b, \ell) / 2) \\
&= (\Gamma(a, b, \ell) / 2) \left[ - \int_{\mathbb{R}} (\partial_x \ln(\psi_t/\psi_\infty))^2 \psi_t + 1_{\{b \neq a\}} (b - a)^2 \frac{\mathcal{G}(a, b, \ell) - \Gamma(a, b, \ell) / 2}{(b - a) \Gamma(a, b, \ell) / 2} \right],
\end{aligned}$$

where the ratio  $\frac{\mathcal{G}(a,b,\ell) - \Gamma(a,b,\ell)/2}{(b-a)\Gamma(a,b,\ell)/2}$  is nonnegative by (19). We remark that

$$\begin{aligned} (a - b)^2 &= \left( \int_{\mathbb{R}} (V')^2 \psi_t - \int_{\mathbb{R}} V'' \psi_t \right)^2 = \left( \int_{\mathbb{R}} V' (V' \psi_t + \partial_x \psi_t) \right)^2 \\ &= \left( \int_{\mathbb{R}} V' \partial_x \ln(\psi_t / e^{-V}) \psi_t \right)^2 \leq a \int_{\mathbb{R}} (\partial_x \ln(\psi_t / \psi_\infty))^2 \psi_t. \end{aligned}$$

Using the function  $F$  defined in (20), we deduce that

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}} \psi_t \ln(\psi_t / \psi_\infty) &\leq (\Gamma(a, b, \ell)/2) \left[ -1 + a \frac{\mathcal{G}(a, b, \ell) - \Gamma(a, b, \ell)/2}{(b - a)\Gamma(a, b, \ell)/2} \right] \int_{\mathbb{R}} (\partial_x \ln(\psi_t / \psi_\infty))^2 \psi_t \\ &\leq -\frac{F(a, b, \ell)}{2} \int_{\mathbb{R}} (\partial_x \ln(\psi_t / \psi_\infty))^2 \psi_t, \end{aligned}$$

which is (24). Since by Lemma 1,  $F$  is positive, we deduce that

$$\frac{d}{dt} \int_{\mathbb{R}} \psi_t \ln(\psi_t / \psi_\infty) \leq 0.$$

Let us now assume that  $\psi_\infty$  satisfies a logarithmic-Sobolev inequality (22) with parameter  $\rho$ . We thus have, from (24),

$$\frac{d}{dt} H(\psi_t | \psi_\infty) \leq -\rho F(a[\psi_t], b[\psi_t], \ell) H(\psi_t | \psi_\infty). \tag{27}$$

Thus, to obtain exponential convergence, in view of Lemma 1 and since  $b[\psi_t] \in [-\|V''\|_\infty, \|V''\|_\infty]$ , we need a (uniform-in-time) upper bound on  $\int_{\mathbb{R}} (V')^2 \psi_t$ , to get a (uniform-in-time) positive lower bound on  $F(a[\psi_t], b[\psi_t], \ell)$ . This is the aim of the next paragraph.

First, notice that by [14], Lemma 1 and Lemma 3,  $\int_{\mathbb{R}} (V')^2 (\psi_t + \psi_\infty) < +\infty$ . Now, according to [19], Theorem 1, since  $\psi_\infty$  satisfies a LSI( $\rho$ ),  $\psi_\infty$  also satisfies the transport inequality: for any probability density  $\varphi$  on  $\mathbb{R}$ ,

$$W_2^2(\varphi, \psi_\infty) = \inf_{\gamma(dx, dy) < \varphi(x) dx, \psi_\infty(y) dy} \int_{\mathbb{R}^2} (x - y)^2 \gamma(dx, dy) \leq \frac{2}{\rho} \int_{\mathbb{R}} \varphi \ln(\varphi / \psi_\infty) = \frac{2}{\rho} H(\varphi | \psi_\infty),$$

where, in the definition of the quadratic Wasserstein distance  $W_2$ , the infimum is taken over all coupling measures  $\gamma$  on  $\mathbb{R}^2$  with marginals  $\varphi(x) dx$  and  $\psi_\infty(y) dy$ . Moreover, for a coupling measure  $\gamma$  between the probability measures  $\psi_t(x) dx$  and  $\psi_\infty(y) dy$ , we have, using Cauchy–Schwarz inequality,

$$\begin{aligned} \left| \int_{\mathbb{R}} (V')^2 (\psi_t - \psi_\infty) \right| &= \left| \int_{\mathbb{R}^2} (V'(x) + V'(y))(V'(x) - V'(y)) \gamma(dx, dy) \right| \\ &\leq \left( 2 \int_{\mathbb{R}} (V')^2 (\psi_t + \psi_\infty) \|V''\|_\infty^2 \int_{\mathbb{R}^2} (x - y)^2 \gamma(dx, dy) \right)^{1/2}. \end{aligned}$$

By taking the infimum over all coupling measures between  $\psi_t(x) dx$  and  $\psi_\infty(y) dy$ , using the above transport inequality and the monotonicity of the relative entropy with respect to  $t$ , we deduce that

$$\begin{aligned} \left| \int_{\mathbb{R}} (V')^2 (\psi_t - \psi_\infty) \right| &\leq \left( \frac{4}{\rho} \|V''\|_\infty^2 H(\psi_t | \psi_\infty) \int_{\mathbb{R}} (V')^2 (\psi_t + \psi_\infty) \right)^{1/2} \\ &\leq \left( \frac{4}{\rho} \|V''\|_\infty^2 H(\psi_0 | \psi_\infty) \left( \left| \int_{\mathbb{R}} (V')^2 (\psi_t - \psi_\infty) \right| + 2 \int_{\mathbb{R}} (V')^2 \psi_\infty \right) \right)^{1/2}. \end{aligned}$$

Setting  $c = \frac{4}{\rho} \|V''\|_\infty^2 H(\psi_0 | \psi_\infty)$  and  $d = \frac{8}{\rho} \|V''\|_\infty^2 \int_{\mathbb{R}} (V')^2 \psi_\infty$ , one concludes that  $\left| \int_{\mathbb{R}} (V')^2 (\psi_t - \psi_\infty) \right| \leq \frac{c + \sqrt{c^2 + 4d}}{2}$  so that

$$\forall t \geq 0, \quad \int_{\mathbb{R}} (V')^2 \psi_t \leq \int_{\mathbb{R}} (V')^2 \psi_\infty + \frac{c + \sqrt{c^2 + 4d}}{2}.$$

By definition of  $a[\psi_t]$ , this yields an upper bound on  $a[\psi_t]$  which depends on  $H(\psi_0 | \psi_\infty)$ . Now, since  $b[\psi_t] \in [-\|V''\|_\infty, \|V''\|_\infty]$ , (21) implies that  $t \mapsto F(a[\psi_t], b[\psi_t], \ell)$  is bounded from below by a positive and nonincreasing function of  $H(\psi_0 | \psi_\infty) = \int_{\mathbb{R}} \psi_0 \ln(\psi_0 / \psi_\infty)$ . We conclude that there exists a positive and nonincreasing function  $\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+^*$  such that

$$\frac{d}{dt} H(\psi_t | \psi_\infty) \leq -\lambda(H(\psi_0 | \psi_\infty)) H(\psi_t | \psi_\infty),$$

which yields (25). □

## 4. Optimization strategies for the RWM algorithm

In this section, we discuss how to choose the constant  $\ell$  in the scaling  $\sigma_n^2 = \frac{\ell^2}{n}$  in order to optimize the convergence to equilibrium, using the nonlinear diffusion limit (7).

As a preliminary remark, notice that we will restrict the discussion to cases when

$$b[\psi_t] = \mathbb{E}(V''(X_t)) > 0. \quad (28)$$

Indeed, points where  $V''$  is negative correspond to neighborhood of local maxima of the potential  $V$ , which are visited with very low probability over large time intervals by the dynamics (7). Moreover, we observe from (26) that if  $b[\psi_t] \leq 0$ , then, since  $\Gamma$  and  $\mathcal{G}$  are nonnegative functions and  $a[\psi_t] \geq 0$ ,  $\frac{d}{dt} H(\psi_t | \psi_\infty) \leq -\frac{\Gamma(a[\psi_t], b[\psi_t], \ell)}{2} \int_{\mathbb{R}} (\partial_x \ln \psi_t)^2 \psi_t$  so that, since  $\lim_{\ell \rightarrow \infty} \Gamma(a, b, \ell) = +\infty$  (when  $b \leq 0$ ),  $\ell$  should be chosen as large as possible in order to leave the concave region.

In the following, we thus assume (28).

### 4.1. Maximization of the exponential rate of convergence

In view of the inequalities (24) and (27), it seems natural to try to choose  $\ell$  maximizing (for given values  $(a, b) = (a[\psi_t], b[\psi_t])$ )

$$\ell \mapsto F(a, b, \ell),$$

in order to maximize the exponential rate of convergence to zero of  $H(\psi_t|\psi_\infty)$ . In view of (20), for  $a \neq b$ , this is equivalent to maximizing  $\ell \mapsto |b\Gamma(a, b, \ell) - 2a\mathcal{G}(a, b, \ell)|$ .

**Remark 2.** We notice that, for  $X_t$  solution to (7),

$$2 \frac{d}{dt} \mathbb{E}(V(X_t)) = b\Gamma(a, b, \ell) - 2a\mathcal{G}(a, b, \ell) \tag{29}$$

with  $(a, b) = (\mathbb{E}(V'(X_t)), \mathbb{E}(V''(X_t)))$ , so that this optimization procedure has a simple interpretation in terms of the evolution of the energy: it amounts to maximizing  $|\frac{d}{dt} \mathbb{E}(V(X_t))|$ , namely making the largest possible moves in terms of energy. This seems quite a reasonable objective.

**Remark 3.** In the Gaussian case (namely when  $V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi)$ ), and assuming that the initial condition is also Gaussian, the density remains Gaussian for all time. Let us denote  $m(t) = \mathbb{E}(X_t)$  its mean and  $s(t) = \mathbb{E}(X_t^2)$  its second order moment, which completely characterize the Gaussian law at time  $t$ . Simple computations, still valid for non-Gaussian initial conditions, yield

$$\begin{cases} \frac{ds}{dt} = \Gamma(s, 1, \ell) - 2s\mathcal{G}(s, 1, \ell) = F(s, 1, \ell)(1 - s), \\ \frac{dm}{dt} = -\mathcal{G}(s, 1, \ell)m, \end{cases} \tag{30}$$

where the first equation corresponds to (29), since  $V'(x) = x$  and  $V''(x) = 1$ . We observe that the optimization procedure in this case amounts to maximizing  $|\frac{ds}{dt}|$ . This accelerates the convergence to the equilibrium value 1 of  $s$ .

Let us denote

$$F_1(s, \ell) = F(s, 1, \ell) = \begin{cases} \ell^2 \exp\left(-\frac{\ell^2}{2}\right), & \text{if } s = 0, \\ 2\ell^2 \left( \left(1 + \frac{\ell^2}{4}\right) \Phi\left(-\frac{\ell}{2}\right) - \frac{\ell}{2\sqrt{2\pi}} \exp\left(-\frac{\ell^2}{8}\right) \right), & \text{if } s = 1, \\ \frac{\ell^2}{1-s} \left( \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + (1-2s) \exp\left(\frac{\ell^2(s-1)}{2}\right) \Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) \right), & \text{if } s \in (0, 1) \cup (1, +\infty), \end{cases}$$

the function to be maximized in the Gaussian case, see Remark 3. We observe that (using the fact that  $b > 0$ ),

$$F(a, b, \ell) = \frac{1}{b} F_1\left(\frac{a}{b}, \ell\sqrt{b}\right) \tag{31}$$

so that the general maximization problem on  $F$  can be reduced to the maximization problem on  $F_1$ . Notice that the function  $F_1$  is  $C^\infty$  on  $\mathbb{R}_+ \times \mathbb{R}_+$ .

**Lemma 2.** *For any  $s \geq 0$ , the function  $\ell \mapsto F_1(s, \ell)$  admits a unique global maximum at a point*

$$\ell^*(s) = \arg \max_{\ell \geq 0} F_1(s, \ell). \tag{32}$$

The proof of this lemma is quite tedious and is given in Appendix B. From Lemma 2 and Equation (31), we deduce that, for  $(a, b) \in \mathbb{R}_+ \times \mathbb{R}_+^*$ , there exists a unique  $\tilde{\ell}^*(a, b)$  such that

$$\tilde{\ell}^*(a, b) = \arg \max_{\ell \geq 0} F(a, b, \ell), \tag{33}$$

and that

$$\tilde{\ell}^*(a, b) = \frac{1}{\sqrt{b}} \ell^*\left(\frac{a}{b}\right). \tag{34}$$

In particular,  $\ell^*(s) = \tilde{\ell}^*(s, 1)$ . Notice that these scaling results show that a constant  $\ell$  strategy is far from optimal in the transient case, since when  $a$  and  $b$  vary, the optimal value  $\tilde{\ell}^*(a, b)$  also varies.

We now consider three regimes: the near equilibrium case  $s \rightarrow 1$  (recall that at equilibrium,  $a = b$  and thus  $s = a/b = 1$ ), and the two situations far from equilibrium  $s \rightarrow 0$  and  $s \rightarrow \infty$  (see Figure 1 for an illustration). In the Gaussian case (see Remark 3),  $s(t) = \mathbb{E}(X_t^2)$  so that these three regimes are easy to understand in terms of second moment.

**Lemma 3.** *We have the following asymptotic behaviors for the function  $\ell^*$ :*

- $s \rightarrow 1$ : *The function  $\ell \mapsto F_1(1, \ell)$  admits a unique maximum at point  $\ell^*(1) \simeq 1.85$ . Moreover,*

$$\lim_{s \rightarrow 1} \ell^*(s) = \ell^*(1), \tag{35}$$

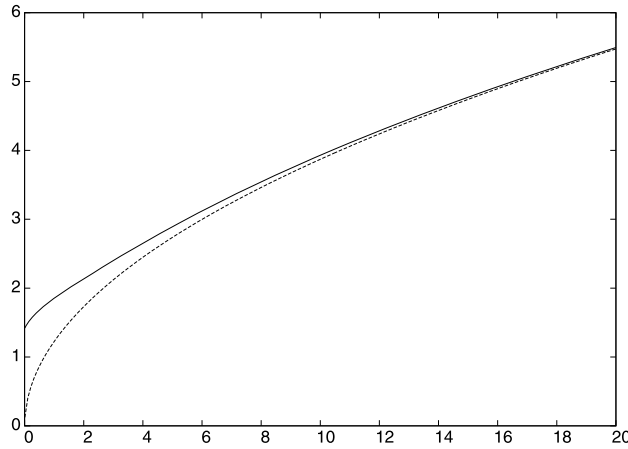
and thus  $\tilde{\ell}^*(a, b) \sim_{a/b \rightarrow 1} \frac{\ell^*(1)}{\sqrt{b}}$ .

- $s \rightarrow 0$ : *The function  $\ell \mapsto F_1(0, \ell)$  admits a unique maximum at point  $\ell^*(0) = \sqrt{2}$ . Moreover,*

$$\lim_{s \rightarrow 0} \ell^*(s) = \ell^*(0) = \sqrt{2}, \tag{36}$$

and thus  $\tilde{\ell}^*(a, b) \sim_{a/b \rightarrow 0} \frac{\sqrt{2}}{\sqrt{b}}$ .





**Figure 1.** Solid line: the function  $s \mapsto \ell^*(s)$ . Dashed line: the function:  $s \mapsto x^* \sqrt{s}$ .

- $s \rightarrow \infty$ : Let us introduce  $\psi(x) = x \sqrt{\frac{2}{\pi}} e^{-x^2/8} - x^2 \Phi(-\frac{x}{2})$ . The function  $\psi$  admits a unique maximum at point  $x^* \simeq 1.22$ . Moreover,

$$\lim_{s \rightarrow +\infty} \frac{\ell^*(s)}{\sqrt{s}} = x^* \tag{37}$$

so that  $\tilde{\ell}^*(a, b) \sim_{a/b \rightarrow \infty} \frac{x^* \sqrt{a}}{b}$ .

**Proof.** The first two statements for  $s = 1$  and  $s = 0$  are simple consequences of Lemma 2 and the implicit function theorem applied to  $F_1(s, \ell)$ , respectively, at point  $(1, \ell^*(1))$  and  $(0, \ell^*(0))$ , using the fact that  $\frac{\partial^2 F_1}{\partial \ell^2}(1, \ell^*(1)) \neq 0$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(0, \ell^*(0)) \neq 0$  (see Equations (B.3) and (B.1) below).

Let us now consider the case  $s \rightarrow \infty$  and recall the well-known Mill’s inequalities:

$$\forall x < 0, \quad \frac{-x}{\sqrt{2\pi}(1+x^2)} \exp\left(-\frac{x^2}{2}\right) < \Phi(x) < \frac{1}{-x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{38}$$

One has  $\psi'(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/8} - 2x \Phi(-\frac{x}{2})$  so that  $\psi'(0) = \sqrt{\frac{2}{\pi}} > 0$  and, by (38),  $\psi'(x) \sim -\sqrt{\frac{2}{\pi}} e^{-x^2/8} < 0$  as  $x \rightarrow +\infty$ . Moreover, by the lower-bound in (38),

$$(e^{x^2/8} \psi'(x))' = -2 \left(1 + \frac{x^2}{4}\right) e^{x^2/8} \Phi\left(-\frac{x}{2}\right) + \frac{x}{\sqrt{2\pi}} < 0 \quad \text{for } x > 0.$$

Whence the existence and uniqueness of  $x^*$ .

For  $s > 1/2$ , by the upper-bound in (38),  $(2s - 1)e^{\ell^2(s-1)/2} \Phi\left(\frac{\ell(1-2s)}{2\sqrt{s}}\right) \leq \frac{\sqrt{2s}}{\ell\sqrt{\pi}} e^{-\ell^2/(8s)}$  so that, for  $s > 1$ ,  $F_1(s, \ell) \leq \frac{s}{s-1} \psi(\ell/\sqrt{s})$ . For  $\varepsilon > 0$ , one deduces that  $\lim_{s \rightarrow \infty} \sup_{\ell \notin \sqrt{s}[x^* - \varepsilon, x^* + \varepsilon]} F_1(s, \ell) \leq \sup_{x \notin [x^* - \varepsilon, x^* + \varepsilon]} \psi(x)$ . On the other hand,

$$\limsup_{s \rightarrow +\infty} F_1(s, \ell^*(s)) \geq \lim_{s \rightarrow +\infty} F_1(s, x^* \sqrt{s}) = \psi(x^*) > \sup_{x \notin [x^* - \varepsilon, x^* + \varepsilon]} \psi(x).$$

Hence, for  $s$  large enough,  $F_1(s, \ell^*(s)) > \sup_{x \notin [x^* - \varepsilon, x^* + \varepsilon]} \psi(x)$  and  $\frac{\ell^*(s)}{\sqrt{s}} \in [x^* - \varepsilon, x^* + \varepsilon]$ . Since  $\varepsilon > 0$  is arbitrary, this yields (37).  $\square$

## 4.2. Comparison with the constant average acceptance rate strategy

Under the stationarity assumption, it is standard (see [22]) to associate to the optimal value of  $\ell \simeq \frac{2.38}{\sqrt{I}}$  an average acceptance rate (see the [Introduction](#)). Indeed, in this case, there is a one-to-one correspondence between  $\ell$  and the limiting acceptance rate

$$\text{acc}(I, I, \ell) = \frac{\Gamma(I, I, \ell)}{\ell^2} = \frac{h(I, \ell)}{\ell^2} = 2\Phi\left(-\frac{\ell\sqrt{I}}{2}\right).$$

More precisely,  $\ell \simeq \frac{2.38}{\sqrt{I}}$  is equivalent to

$$\text{acc}(I, I, \ell) \simeq 2\Phi\left(-\frac{2.38}{2}\right) \simeq 0.23$$

which does not depend on  $I$ . A natural strategy is thus to adjust the variance in such a way that the average acceptance rate is 23%. In this section, we discuss how to use an equivalent approach in the transient phase. Of course, the interest of the constant average acceptance rate strategy is that it can be implemented using the so-called adaptive scaling Metropolis algorithm (see [1,3]): at iteration  $k$ , the standard deviation  $\sigma$  is chosen equal to  $\exp(\theta_k)$  where  $\theta_k$  is updated using the Robbins–Monro procedure  $\theta_{k+1} = \theta_k + \gamma_{k+1}(\alpha_k - \alpha)$  where  $\alpha_{k+1}$  is the observed acceptance rate (1) at iteration  $k$ ,  $\alpha \in (0, 1)$  is the target acceptance rate and  $(\gamma_k)$  is a deterministic fixed sequence of step sizes.

The first question is: for given values of  $a$  and  $b$ , does an acceptance rate  $\alpha \in (0, 1)$  corresponds in a one-to-one way to a value  $\ell > 0$ ? The average acceptance rate is (see Proposition 1)

$$\text{acc}(a, b, \ell) = \frac{\Gamma(a, b, \ell)}{\ell^2} = \Phi\left(-\frac{\ell b}{2\sqrt{a}}\right) + e^{\ell^2(a-b)/2} \Phi\left(\ell\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)\right).$$

We recall that we only consider the case  $b > 0$ , see the discussion at the beginning of Section 4. (Actually, if  $b \leq 0$ ,  $\text{acc}(a, b, \ell) \geq \Phi\left(-\frac{\ell b}{2\sqrt{a}}\right) \geq 1/2$  for all  $a \geq 0$  and  $\ell > 0$ , so that it is not possible to solve  $\text{acc}(a, b, \ell) = \alpha$  for any values of  $\alpha$ , which is again an indication of the ill-posedness of the optimization procedure when  $b \leq 0$ .)

Now, for  $b > 0$ , observe that

$$\text{acc}(a, b, \ell) = J\left(\frac{a}{b}, \ell\sqrt{b}\right),$$

where

$$J(s, \ell) = \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + e^{\ell^2(s-1)/2} \Phi\left(\ell\left(\frac{1}{2\sqrt{s}} - \sqrt{s}\right)\right). \tag{39}$$

Solving  $\text{acc}(a, b, \ell) = \alpha$  amounts to solving  $J\left(\frac{a}{b}, \ell\sqrt{b}\right) = \alpha$ .

**Lemma 4.** *Let  $s > 0$  be fixed. The function  $\ell \mapsto J(s, \ell)$  is decreasing. Moreover, for all  $\alpha \in (0, 1)$  there exists a unique solution to the equation  $J(s, \ell) = \alpha$ . This solution is denoted  $\ell^\alpha(s)$  in the following.*

**Proof.** Let us first prove that, for a given  $s > 0$ ,  $\ell \mapsto J(s, \ell)$  is strictly decreasing. We compute

$$\frac{\partial J}{\partial \ell}(s, \ell) = -\sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + \ell(s-1) \exp\left(\frac{\ell^2(s-1)}{2}\right) \Phi\left(\frac{\ell}{2\sqrt{s}}(1-2s)\right).$$

The right-hand side is negative for  $s \in (0, 1]$ . For  $s > 1$ , we have, using the upper-bound in (38),

$$\frac{\partial J}{\partial \ell}(s, \ell) \leq -\sqrt{\frac{s}{2\pi}} \frac{1}{(2s-1)} \exp\left(-\frac{\ell^2}{8s}\right) < 0.$$

This shows that  $\ell \mapsto J(s, \ell)$  is strictly decreasing.

It is easy to see that  $J(s, 0) = 1$ . Now, using again the upper-bound in (38) for  $s > 1/2$ , one has

$$J(s, \ell) \leq \Phi\left(\frac{-\ell}{2\sqrt{s}}\right) + 1_{\{s \leq 1/2\}} \exp\left(\frac{\ell^2}{2}(s-1)\right) + 1_{\{s > 1/2\}} \frac{\sqrt{2s}}{\ell(2s-1)\sqrt{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) \tag{40}$$

so that  $\lim_{\ell \rightarrow \infty} J(s, \ell) = 0$ . By continuity and strict monotonicity of  $J$ , we then get that for any  $\alpha \in (0, 1)$  there exists a unique  $\ell^\alpha(s)$  such that  $J(s, \ell^\alpha(s)) = \alpha$ .  $\square$

As a corollary of this lemma, we get that for any  $a \geq 0, b > 0, \alpha \in (0, 1)$ , there exists a unique  $\tilde{\ell}^\alpha(a, b) > 0$  such that

$$\text{acc}(a, b, \tilde{\ell}^\alpha(a, b)) = \alpha$$

and that

$$\tilde{\ell}^\alpha(a, b) = \frac{1}{\sqrt{b}} \ell^\alpha\left(\frac{a}{b}\right). \tag{41}$$

In particular,  $\ell^\alpha(s) = \tilde{\ell}^\alpha(s, 1)$ .

Let us now compare the strategy based on the maximization of the exponential rate of convergence, presented in Section 4.1, with a strategy based on a constant average acceptance rate. By

comparing (34) and (41), we observe that the scalings of  $\tilde{\ell}^*$  and  $\tilde{\ell}^\alpha$  in terms of  $a$  and  $b$  are the same, which is already an indication of the fact that a constant acceptance rate strategy is very natural.

Near equilibrium, namely in the limit  $a/b \rightarrow 1$ , the two strategies are the same if  $\alpha$  is chosen such that  $\ell^\alpha(1) = \ell^*(1)$  which corresponds to

$$\alpha \simeq 0.35. \tag{42}$$

Notice that this value is not far (but different, since we take into account the transient phase around equilibrium) from the acceptance probability 0.23 obtained under the stationarity assumption.

To study the two limits  $s \rightarrow 0$  and  $s \rightarrow \infty$ , we need the following lemma.

**Lemma 5.** *We have the following asymptotic behaviors for the function  $\ell^\alpha$ :*

- $s \rightarrow 0$ : For any  $\alpha \in (0, 1)$ ,

$$\lim_{s \rightarrow 0} \ell^\alpha(s) = \sqrt{-2 \ln(\alpha)}, \tag{43}$$

and thus  $\tilde{\ell}^\alpha(a, b) \sim_{a/b \rightarrow 0} \frac{\sqrt{-2 \ln(\alpha)}}{\sqrt{b}}$ .

- $s \rightarrow \infty$ : For any  $\alpha \in (0, \frac{1}{2})$ ,

$$\lim_{s \rightarrow \infty} \frac{\ell^\alpha(s)}{\sqrt{s}} = -2\Phi^{-1}(\alpha), \tag{44}$$

and thus  $\tilde{\ell}^\alpha(a, b) \sim_{a/b \rightarrow \infty} -2\Phi^{-1}(\alpha) \frac{\sqrt{a}}{b}$ .

**Proof.** Let us first consider the case  $s \rightarrow 0$ . Observe that for any given  $\ell \geq 0$ , it holds

$$\lim_{s \rightarrow 0} J(s, \ell) = \exp\left(\frac{-\ell^2}{2}\right). \tag{45}$$

Let  $\varepsilon \in (0, -2 \ln(\alpha))$ . By the monotonicity property of  $\ell \mapsto J(s, \ell)$  stated in Lemma 4,

$$\sup_{\ell \geq \sqrt{-2 \ln(\alpha) + \varepsilon}} J(s, \ell) \leq J(s, \sqrt{-2 \ln(\alpha) + \varepsilon}) \xrightarrow{s \rightarrow 0} \alpha \exp\left(\frac{-\varepsilon}{2}\right).$$

In the same way,  $\liminf_{s \rightarrow 0} \inf_{\ell \leq \sqrt{-2 \ln(\alpha) - \varepsilon}} J(s, \ell) \geq \alpha \exp(\frac{\varepsilon}{2})$ . Therefore, for  $s$  close enough to 0,  $\ell^\alpha(s) \in [\sqrt{-2 \ln(\alpha) - \varepsilon}, \sqrt{-2 \ln(\alpha) + \varepsilon}]$  and (43) holds.

Let us now consider the case  $s \rightarrow \infty$ . Observe that  $J(s, \ell) > \Phi(-\frac{\ell}{2\sqrt{s}})$ . Hence, if

$$\liminf_{s \rightarrow \infty} \frac{\ell^\alpha(s)}{\sqrt{s}} = 0$$

then  $\liminf_{s \rightarrow \infty} J(s, \ell^\alpha(s)) \geq \frac{1}{2}$ . Therefore, there exists a constant  $C_1 > 0$  such that, for large enough  $s$ ,  $\ell^\alpha(s) \geq C_1 \sqrt{s}$ . Using (40) for the upper-bound, we get that, for  $s$  large enough,

$$\Phi\left(-\frac{\ell^\alpha(s)}{2\sqrt{s}}\right) \leq \alpha \leq \Phi\left(-\frac{\ell^\alpha(s)}{2\sqrt{s}}\right) + \frac{\sqrt{2}}{\sqrt{\pi}C_1(2s-1)} \exp\left(-\frac{C_1^2}{8}\right).$$

Therefore,

$$\lim_{s \rightarrow \infty} \Phi\left(-\frac{\ell^\alpha(s)}{2\sqrt{s}}\right) = \alpha.$$

The continuity of  $\Phi^{-1}$  concludes the proof of (44).  $\square$

By comparing (36) with (43), we observe that in the regime  $a/b \rightarrow 0$ , the two strategies are the same if  $\alpha$  is chosen such that  $\sqrt{-2 \ln(\alpha)} = \sqrt{2}$  namely

$$\alpha = e^{-1} \sim 0.37. \quad (46)$$

Finally, by comparing (37) with (44), we observe that in the regime  $a/b \rightarrow \infty$ , the two strategies are the same if  $\alpha$  is chosen such that  $-2\Phi^{-1}(\alpha) = x^*$  namely

$$\alpha \sim 0.27. \quad (47)$$

In view of (42), (46) and (47), the constant average acceptance rate strategy with target value between  $1/4$  and  $1/3$  seems to be a very good strategy, since it is almost equivalent to the optimal exponential rate strategy.

## 5. Numerical experiments

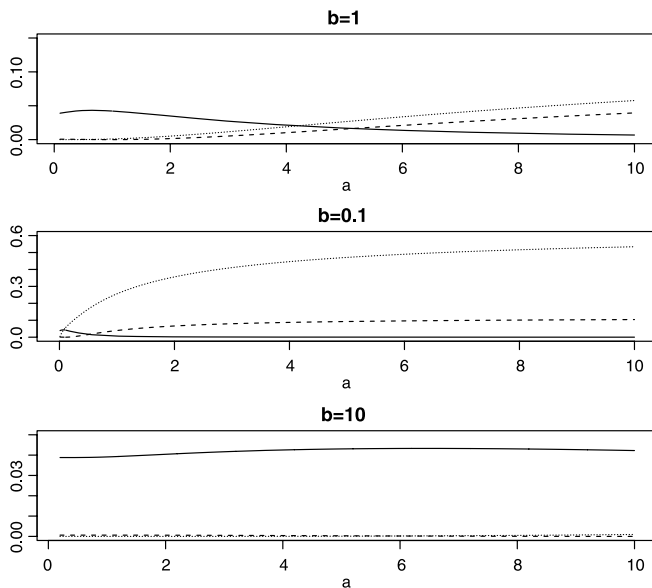
In this section, we present numerical experiments to illustrate results from Section 4.

### 5.1. On the choice of the target average acceptance rate

In this section, we would like to discuss the choice of the average acceptance rate  $\alpha$  in the constant average acceptance rate strategy. As mentioned above, we identified three different values of  $\alpha$  for the constant average acceptance rate to be equivalent to the optimization of the exponential rate of convergence, depending on the regimes:  $\frac{a}{b} \rightarrow 1$  ( $\alpha \simeq 0.35$ );  $\frac{a}{b} \rightarrow 0$  ( $\alpha \simeq 0.37$ );  $\frac{a}{b} \rightarrow \infty$  ( $\alpha \simeq 0.27$ ).

In practice, a value has to be chosen for  $\alpha$ . On Figure 2, we plot as a function of  $a$  and  $b$  the relative loss in terms of exponential rate of convergence, for the constant average acceptance rate strategy compared to the optimization of the exponential rate of convergence:  $\frac{F(a,b,\tilde{\ell}^*(a,b)) - F(a,b,\tilde{\ell}^\alpha(a,b))}{F(a,b,\tilde{\ell}^*(a,b))}$ , for the three values of  $\alpha$  mentioned above.

The main output of these numerical experiments is that the choice  $\alpha \simeq 0.27$  seems to be the most robust, namely the one which leads to an exponential rate of convergence the closest to the optimal one, over the largest range of variation of  $a$  and  $b$ . This confirms the interest of the constant acceptance rate strategy.



**Figure 2.**  $\frac{F(a,b,\tilde{\ell}^*(a,b))-F(a,b,\tilde{\ell}^\alpha(a,b))}{F(a,b,\tilde{\ell}^*(a,b))}$  as function of  $a$  for  $b = 1, 0.1, 10$  and  $\alpha \simeq 0.27$  (solid line),  $\alpha \simeq 0.35$  (dashed line),  $\alpha = e^{-1} \simeq 0.37$  (dotted line).

### 5.2. Gaussian case

Let us first consider the Gaussian target  $V(x) = \frac{1}{2}(x^2 + \ln(2\pi))$  (see Remark 3), with a Gaussian initial condition  $X_0$  such that  $m(0) = \mathbb{E}(X_0)$  and  $s(0) = \mathbb{E}(X_0^2)$ . At time  $t$ , the law of  $X_t$  solution to the limiting stochastic differential equation (7) is Gaussian with mean  $m(t)$  and second moment  $s(t)$ , where  $m$  and  $s$  satisfies (30). The Kullback–Leibler divergence admits an analytical expression in terms of  $m$  and  $s$ :

$$H(\psi_t|\psi_\infty) = \frac{1}{2}(s(t) - \ln(s(t) - m(t)^2) - 1),$$

and its derivative writes

$$\begin{aligned} \frac{d}{dt}H(\psi_t|\psi_\infty) &= \frac{1}{2} \left( \frac{ds}{dt} - \frac{ds/dt - 2m dm/dt}{s - m^2} \right) \\ &= \frac{1}{2} \left( F_1(s, \ell)(1 - s) - \frac{F_1(s, \ell)(1 - s) + 2m\mathcal{G}(s, 1, \ell)}{s - m^2} \right). \end{aligned}$$

In the Gaussian case, it is thus possible, for each time  $t$  (and thus for fixed values of  $m(t)$  and  $s(t)$ ), to minimize  $\frac{d}{dt}H(\psi_t|\psi_\infty)$  in  $\ell$ . This yields the best strategy that we could think of and implement numerically, in terms of the speed of convergence of the Kullback–Leibler divergence

to 0. In the following, let us denote

$$\ell^{\text{ent}}(m, s) = \arg \min_{\ell} \left( F_1(s, \ell)(1 - s) - \frac{F_1(s, \ell)(1 - s) + 2m\mathcal{G}(s, 1, \ell)}{s - m^2} \right).$$

In the numerical experiments, we thus compare four strategies: (i) the constant  $\ell$  strategy, with  $\ell = 2.38$  (which is the optimal value under stationarity assumption since  $I = 1$  in the Gaussian case); (ii) the constant average acceptance rate strategy, using  $\ell^\alpha(a, b)$  (for  $\alpha \simeq 0.27$  and  $\alpha = e^{-1} \simeq 0.37$ ); (iii) the optimal exponential rate of convergence using  $\ell^*(a, b)$ ; (iv) the optimal strategy for the convergence of the entropy using  $\ell^{\text{ent}}(m, s)$ . Notice that in the Gaussian case,  $a = \mathbb{E}(X_t^2) = s(t)$  and  $b = 1$ , so that  $\ell^\alpha$  and  $\ell^*$  are actually functions of  $s$  only. Let us also mention that there are actually two ways to implement (ii): either using a numerical approximation for  $\ell^\alpha(a, 1)$  (and an estimator  $\hat{a}$  of  $a$ ), or using the adaptive scaling Metropolis algorithm mentioned at the beginning of Section 4.2 (see [1,3]).

The dimension is fixed to  $n = 100$ . To assess the convergence, we observe, as a function of the so-called burn-in time  $t_0$ , the convergence to zero of the square biases:

$$\left( \mathbb{E}(\hat{I}_{t_0, T+t_0}^s) - 1 \right)^2 \quad \text{and} \quad \left( \mathbb{E}(\hat{I}_{t_0, T+t_0}^m) \right)^2, \tag{48}$$

where

$$\hat{I}_{t_0, T+t_0}^s = \frac{1}{T} \sum_{k=t_0+1}^{T+t_0} \frac{(X_k^1)^2 + \dots + (X_k^{100})^2}{100} \tag{49}$$

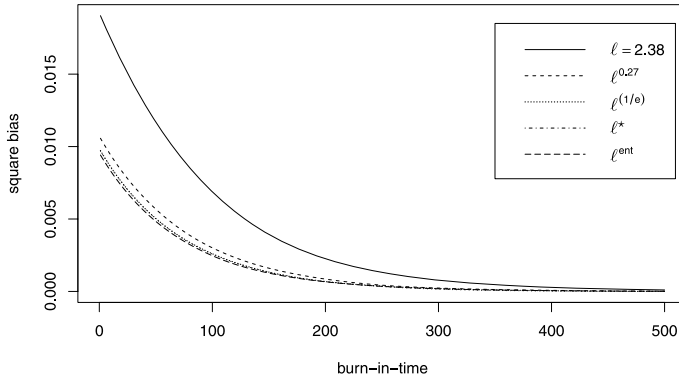
and

$$\hat{I}_{t_0, T+t_0}^m = \frac{1}{T} \sum_{k=t_0+1}^{T+t_0} \frac{X_k^1 + \dots + X_k^{100}}{100}. \tag{50}$$

The expectations in (48) are approximated by empirical averages over 200 independent realizations of  $(X_k^1, \dots, X_k^{100})_{0 \leq k \leq t_0+T}$ . The size of the time window is  $T = 1500$ . When needed, we estimate the values of  $s = a$  and  $m$  using empirical averages over the  $n = 100$  components of the process.

On Figure 3, we first consider the initial condition  $X_0 = (0, \dots, 0)$ . The first moment is thus already at equilibrium, and we only observe the convergence of the second moment. Clearly, the constant  $\ell$  strategy is the worst. Using  $\ell^*$  yields a convergence which is almost the optimal one, obtained for  $\ell = \ell^{\text{ent}}$ . And the constant average rate strategies also lead to excellent results in terms of convergence compared to the optimal scenario, even though it is here implemented using an adaptive scaling Metropolis algorithm.

On Figure 4, we perform similar experiments with the initial condition  $X_0 = (10, \dots, 10)$ . We observe the convergence of the first and second moment. It is clear that the constant  $\ell$  strategy is outperformed by all the other strategies. We notice also that the adaptive scaling Metropolis implementation leads to slightly slower convergences compared to an implementation using  $\ell^\alpha(\hat{a}, 1)$ . This difference could certainly be reduced by optimizing the parameters in the adaptive scaling Metropolis algorithm.



**Figure 3.** Square bias of  $\hat{I}_{t_0, t_0+T}^s$  as a function of the burn-in-time  $t_0$  for various strategies. The initial condition is  $(0, \dots, 0)$ , and the constant acceptance rate strategies are implemented using an adaptive scaling Metropolis algorithm.

In conclusion, we observed that: (i) The constant  $\ell$  strategy is bad; (ii) The constant average acceptance rate strategy (using  $\ell^\alpha$ ) leads to convergence curves which are very close to the ones obtained with the optimal exponential rate of convergence strategy (using  $\ell^*$ ); (iii) the optimal exponential rate of convergence strategy is as good as the most optimal strategy one could design in terms of entropy decay (using  $\ell^{\text{ent}}$ ).

### 5.3. Non-Gaussian case

Let us now consider a non-Gaussian target, and more precisely a double-well potential. In order to satisfy the assumptions of Theorem 1, we consider the function  $V$  given up to a normalizing additive constant by:

$$V(x) = \begin{cases} (x - 1)^2(x + 1)^2, & \text{if } |x| \leq 1, \\ 4x^2 - 8|x| + 4, & \text{otherwise.} \end{cases}$$

Simple calculations yield

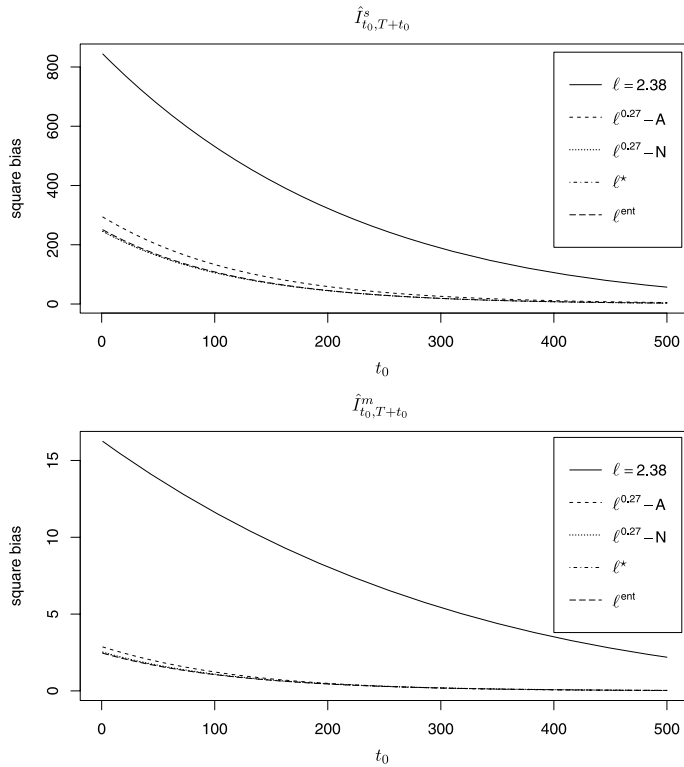
$$V'(x) = \begin{cases} 2(x - 1)(x + 1)^2 + 2(x - 1)^2(x + 1), & \text{if } |x| \leq 1, \\ 8x - 8 \text{ sign}(x), & \text{otherwise,} \end{cases}$$

and

$$V''(x) = \begin{cases} 2(x + 1)^2 + 8(x - 1)(x + 1) + 2(x - 1)^2, & \text{if } |x| \leq 1, \\ 8, & \text{otherwise.} \end{cases}$$

Of course, no analytical expression for the entropy is available in this context, and we thus concentrate on the three following strategies: (i) the constant  $\ell$  strategy; (ii)  $\ell = \ell^\alpha(a, b)$  and (iii)  $\ell = \ell^*(a, b)$ . For the constant  $\ell$  strategy, we use  $\ell = \frac{2.38}{\sqrt{I}} = 1.18$  (where we recall,  $I$  is defined by (4)). When needed,  $a$  and  $b$  are approximated by the estimators over the  $n$  components



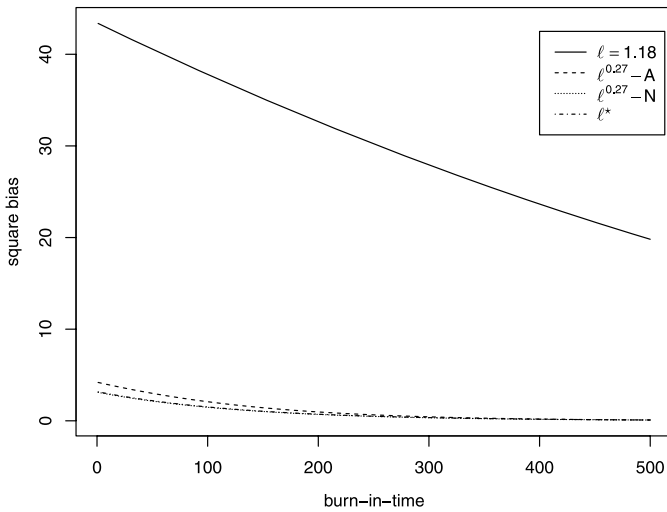


**Figure 4.** Square bias of  $\hat{I}_{t_0, T+t_0}^s$  (top) and  $\hat{I}_{t_0, T+t_0}^m$  (bottom) as a function of the burn-in-time  $t_0$  for various strategies. The initial condition is  $(10, \dots, 10)$ . The notations  $\ell^{0.27} - A$  and  $\ell^{0.27} - N$  refer to the two implementations of the constant average acceptance rate:  $A$  for the adaptive scaling Metropolis algorithm and  $N$  for the numerical approximation of  $\ell^\alpha(a, 1)$ .

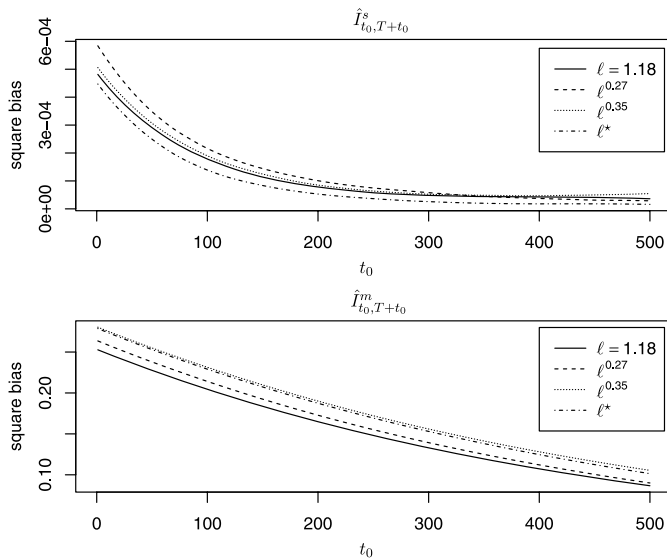
$\hat{a} = \frac{1}{n} \sum_{i=1}^n V'(X_t^i)^2$  and  $\hat{b} = \frac{1}{n} \sum_{i=1}^n V''(X_t^i)$ . The parameters  $n = 100$  and  $T = 1500$  are the same as in the Gaussian case.

Let us first consider as an initial condition  $X_0 = (10, \dots, 10)$ . On Figure 5, we observe the convergence of the first moment to its equilibrium value (namely 0). Again, the constant  $\ell$  strategy appears to be very bad, and the other strategies perform approximately equally well.

Finally, let us consider  $X_0$  distributed according to a Gaussian distribution with mean 1 and variance 0.143Id. The mean and the variance are chosen in such a way that  $a = \mathbb{E}(V'(X_0)^2) = 5.24$  and  $b = \mathbb{E}(V''(X_0)) = 5.24$ . On Figure 6, we observe the convergence of the first and second moments to their equilibrium values (namely 0 and 0.96). For the constant acceptance rate strategy, we compare the results obtained with  $\alpha = 0.35$  and  $\alpha = 0.27$ . Here, it is much more complicated to draw general conclusions from these plots. Basically, all strategies yield comparable results. One could wonder why  $\ell^*$  performs poorly for the first moment. The reason is probably that its bias cannot be encoded into  $a$  and  $b$  which are integrals of even functions with respect to the current marginal distribution.



**Figure 5.** Square bias of  $\hat{I}_{t_0, T+t_0}^m$  for non-Gaussian target as a function of the burn-in-time  $t_0$  for various strategies. The initial condition is  $(10, \dots, 10)$ . The notations  $\ell^{0.27} - A$  and  $\ell^{0.27} - N$  refer to the two implementations of the constant average acceptance rate:  $A$  for the adaptive scaling Metropolis algorithm and  $N$  for the numerical approximation of  $\ell^\alpha(a, b)$ .



**Figure 6.** Square bias of  $\hat{I}_{t_0, T+t_0}^s$  (top) and  $\hat{I}_{t_0, T+t_0}^m$  (bottom) for non-Gaussian target as a function of the burn-in-time  $t_0$  for various strategies and Gaussian initial condition. The constant acceptance rate strategies are implemented using an adaptive scaling Metropolis algorithm.

In conclusion, we observed that the results obtained with the constant acceptance rate strategy (even when it is implemented using an adaptive scaling Metropolis algorithm) are very similar to those obtained with the optimal exponential rate of convergence strategy.

## 6. Scaling limits for the MALA algorithm

The aim of this section is to derive a diffusive limit for the MALA algorithm, following the same reasoning as for the RWM algorithm in Section 2.

The Markov chain generated by the MALA algorithm writes:

$$X_{k+1}^{i,n} = X_k^{i,n} + Z_{k+1}^{i,n} 1_{\mathcal{A}_{k+1}} \tag{51}$$

$$\text{where } Z_{k+1}^{i,n} = \sigma_n G_{k+1}^i - \frac{\sigma_n^2}{2} V'(X_k^{i,n}), \quad 1 \leq i \leq n,$$

$$\text{and } \mathcal{A}_{k+1} = \left\{ U_{k+1} \leq \exp \left\{ \sum_{i=1}^n \left( V(X_k^{i,n}) - V(X_k^{i,n} + Z_{k+1}^{i,n}) \right) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\sigma_n}{2} (V'(X_k^{i,n}) + V'(X_k^{i,n} + Z_{k+1}^{i,n})) \right)^2 \right] \right\} \right\}$$

is the accepting event. Here again,  $(G_k^i)_{i,k \geq 1}$  is a sequence of i.i.d. normal random variables independent from a sequence  $(U_k)_{k \geq 1}$  of i.i.d. random variables uniform on  $[0, 1]$ . In Section 6.1, we formally derive a limiting diffusion process. It appears that the scaling to be used depend on the sign of  $\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n}))$ . This is more rigorously discussed in Section 6.2 for a Gaussian target probability measure.

### 6.1. A formal derivation of the limiting process

#### 6.1.1. Asymptotic analysis and limiting process

We adapt the same strategy as for the RWM algorithm, in Section 2.2. Let us first discuss how to choose the proper scaling for  $\sigma_n$ . Using a Taylor expansion, one obtains:

$$\begin{aligned} & V(X_k^{i,n}) - V(X_k^{i,n} + Z_{k+1}^{i,n}) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\sigma_n}{2} (V'(X_k^{i,n}) + V'(X_k^{i,n} + Z_{k+1}^{i,n})) \right)^2 \right] \\ &= \sigma_n^3 \left( \frac{V^{(3)}(X_k^{i,n})}{12} (G_{k+1}^i)^3 - \frac{V' V''(X_k^{i,n})}{4} G_{k+1}^i \right) \\ &+ \sigma_n^4 \left( \frac{(V')^2 V''(X_k^{i,n})}{8} + \frac{V^{(4)}(X_k^{i,n})}{24} (G_{k+1}^i)^4 - \frac{2V^{(3)} V'(X_k^{i,n}) + (V''(X_k^{i,n}))^2}{8} (G_{k+1}^i)^2 \right) \\ &+ \mathcal{O}(\sigma_n^5). \end{aligned}$$

Setting, as above,  $v_k^n = \frac{1}{n} \sum_{i=1}^n \delta_{X_k^{i,n}}$ , one has by Gaussian computations

$$\begin{aligned} & \mathbb{E} \left( \left( \sum_{i=1}^n \left( \frac{V^{(3)}(X_k^{i,n})}{12} (G_{k+1}^i)^3 - \frac{V'V''(X_k^{i,n})}{4} G_{k+1}^i \right) \right)^2 \right) \\ &= \frac{n}{48} \mathbb{E} \left( (v_k^n, 5(V^{(3)})^2 - 6V'V''V^{(3)} + 3(V'V'')^2) \right) \end{aligned}$$

so that one expects that  $\sum_{i=1}^n \left( \frac{V^{(3)}(X_k^{i,n})}{12} (G_{k+1}^i)^3 - \frac{V'V''(X_k^{i,n})}{4} G_{k+1}^i \right) = \mathcal{O}(\sqrt{n})$  and similarly that  $\sum_{i=1}^n \left( \frac{V^{(4)}(X_k^{i,n})}{24} ((G_{k+1}^i)^4 - 3) - \frac{2V^{(3)}V'(X_k^{i,n}) + (V''(X_k^{i,n}))^2}{8} ((G_{k+1}^i)^2 - 1) \right) = \mathcal{O}(\sqrt{n})$ . If this holds and  $\lim_{n \rightarrow \infty} \sigma_n = 0$ , then

$$\begin{aligned} & \sum_{i=1}^n \left( V(X_k^{i,n}) - V(X_k^{i,n} + Z_{k+1}^{i,n}) \right) \\ &+ \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\sigma_n}{2} (V'(X_k^{i,n}) + V'(X_k^{i,n} + Z_{k+1}^{i,n})) \right)^2 \right] \\ &= \frac{n\sigma_n^4}{8} \langle v_k^n, (V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2 \rangle + \mathcal{O}(\sqrt{n}\sigma_n^3) + \mathcal{O}(n\sigma_n^5). \end{aligned}$$

From this,

$$\mathbb{E}(1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) = e^{(n\sigma_n^4/8) \langle v_k^n, ((V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2) \rangle} \wedge 1 + \mathcal{O}(\sqrt{n}\sigma_n^3) + \mathcal{O}(n\sigma_n^5). \quad (52)$$

Here, we have assumed that  $\langle v_k^n, (V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2 \rangle \neq 0$ . From this formula, we get the correct scaling for the variance, in order to obtain a nontrivial limiting acceptance rate (in accordance with [12], Section 5):

$$\sigma_n^4 = \frac{\ell^4}{n}.$$

Now, following the same reasoning as in Section 2.2, we have: for a test function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ ,

$$\begin{aligned} & \mathbb{E}(\varphi(X_{k+1}^{1,n}) | \mathcal{F}_k^n) \\ &= \mathbb{E}(\varphi(X_k^{1,n} + Z_{k+1}^{1,n} 1_{\mathcal{A}_{k+1}}) | \mathcal{F}_k^n) \\ &= \mathbb{E} \left( \varphi(X_k^{1,n}) + \varphi'(X_k^{1,n}) Z_{k+1}^{1,n} 1_{\mathcal{A}_{k+1}} + \frac{1}{2} \varphi''(X_k^{1,n}) (Z_{k+1}^{1,n})^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n \right) + \mathcal{O}(n^{-3/4}) \\ &= \varphi(X_k^{1,n}) + \varphi'(X_k^{1,n}) \mathbb{E}(Z_{k+1}^{1,n} 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) + \frac{1}{2} \varphi''(X_k^{1,n}) \mathbb{E}((Z_{k+1}^{1,n})^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) + \mathcal{O}(n^{-3/4}). \end{aligned}$$

Using the Lipschitz continuity of  $y \mapsto e^y \wedge 1$ , one may remove the contribution of the  $i$ th coordinate in the acceptance ratio and then introduce it again after using conditional independence to

check that

$$\begin{aligned} \mathbb{E}(G_{k+1}^i 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) &= \mathbb{E}(G_{k+1}^i) \mathbb{E}(1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) + \mathcal{O}(\sigma_n^3) = \mathcal{O}(n^{-3/4}), \\ \mathbb{E}((G_{k+1}^i)^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) &= \mathbb{E}[(G_{k+1}^i)^2] \mathbb{E}(1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n) + \mathcal{O}(\sigma_n^3) = \mathbb{P}(\mathcal{A}_{k+1} | \mathcal{F}_k^n) + \mathcal{O}(n^{-3/4}). \end{aligned}$$

From this, one obtains

$$\begin{aligned} &\mathbb{E}(\varphi(X_{k+1}^{1,n}) | \mathcal{F}_k^n) \\ &= \varphi(X_k^{1,n}) + \varphi'(X_k^{1,n}) \mathbb{E}\left(\left(\sigma_n G_{k+1}^i - \frac{\sigma_n^2}{2} V'(X_k^{1,n})\right) 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n\right) \\ &\quad + \frac{1}{2} \varphi''(X_k^{1,n}) \mathbb{E}\left(\left(\sigma_n G_{k+1}^i - \frac{\sigma_n^2}{2} V'(X_k^{1,n})\right)^2 1_{\mathcal{A}_{k+1}} | \mathcal{F}_k^n\right) + \mathcal{O}(n^{-3/4}) \\ &= \varphi(X_k^{1,n}) + \frac{\ell^2}{2\sqrt{n}} \left(e^{(\ell^4/8)\langle v_k^n, ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2) \rangle} \wedge 1\right) \\ &\quad \times (-V'(X_k^{1,n}) \varphi'(X_k^{1,n}) + \varphi''(X_k^{1,n})) + \mathcal{O}(n^{-3/4}). \end{aligned}$$

The correct scaling in time is thus to consider a piecewise linear process  $Y_t^{1,n}$  such that  $Y_{k/\sqrt{n}}^{1,n} = X_k^{1,n}$  (this is again the standard diffusive timescale), and the expected propagation of chaos limit is solution to the nonlinear stochastic differential equation:

$$\begin{aligned} dX_t &= \sqrt{w(t, \ell)} dB_t - w(t, \ell) \frac{1}{2} V'(X_t) dt, \\ &\text{where } w(t, \ell) = \ell^2 \left(e^{(\ell^4/8)\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_t))} \wedge 1\right). \end{aligned} \tag{53}$$

This equation is obtained by a deterministic (and nonlinear in the sense of McKean) change of time applied to the standard overdamped Langevin stochastic differential equation with reversible density  $e^{-V}$ . Under appropriate assumptions on the potential  $V$ , we believe that a rigorous proof of this result could be done using similar techniques as for the RWM algorithm in [14].

### 6.1.2. Relation to previous results in the literature

These results are related to previous ones in the literature. First, in the Gaussian case  $V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi)$ , one obtains from (53) that  $(\mathbb{E}(X_t^2))_{t \geq 0}$  solves the ordinary differential equation

$$\frac{d}{dt} \mathbb{E}(X_t^2) = \ell^2 \left(e^{(\ell^4/8)\mathbb{E}(X_t^2) - 1} \wedge 1\right) (1 - \mathbb{E}(X_t^2)).$$

We recover here a result from [12], Theorem 2, where it is shown that the process  $(\frac{1}{n} \sum_{i=1}^n (X_{\lfloor nt \rfloor}^{i,n})^2)_{t \geq 0}$ , in the limit  $n \rightarrow \infty$  satisfies this ordinary differential equation.

Second, in the stationary case, namely when  $(X_0^{1,n}, \dots, X_0^{n,n})$  are distributed according to the target density  $p$  defined by (2), the equalities

$$\begin{aligned} \int_{\mathbb{R}} V^{(3)} V'(x) e^{-V(x)} dx &= \int_{\mathbb{R}} V^{(4)}(x) e^{-V(x)} dx, \\ \int_{\mathbb{R}} V''(V')^2(x) e^{-V(x)} dx &= \int_{\mathbb{R}} [V^{(3)} V' + (V'')^2](x) e^{-V(x)} dx \end{aligned}$$

imply that  $\langle v_k^n, ((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2) \rangle = 0$  and this changes the scaling of the limiting acceptance rate in (52). In [23], it is shown that in this case, the correct scaling is  $\sigma_n^2 = \frac{\ell^2}{n^{1/3}}$  and then  $(X_{\lfloor n^{1/3} t \rfloor}^1)_{t \geq 0}$  converges in distribution to the solution  $(X_t)_{t \geq 0}$  of the stochastic differential equation

$$\begin{aligned} dX_t &= \sqrt{z(\ell)} dB_t - z(\ell) \frac{1}{2} V'(X_t) dt, \\ \text{where } z(\ell) &= 2\ell^2 \Phi \left( -\frac{\ell^3 \sqrt{(m, 5(V^{(3)})^2 - 3(V'')^3)/3}}{8} \right), \end{aligned} \tag{54}$$

where  $dm = e^{-V(x)} dx$ .

### 6.1.3. Practical counterparts

The practical counterparts of the convergence results discussed above are the following. We can actually distinguish between three regimes:

- On time intervals such that  $\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n})) < 0$ , then the correct scaling to obtain a diffusive limit is  $\sigma_n^2 = \frac{\ell^2}{n^{1/2}}$  and there exists an optimal value of  $\ell$  to speed up the time scale of the dynamics of  $X_t$ , by maximizing  $w(t, \ell)$  (see Equation (53)).
- On time intervals such that  $\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n})) = 0$ , then the correct scaling to obtain a diffusive limit is  $\sigma_n^2 = \frac{\ell^2}{n^{1/3}}$ , and again, there exists an optimal value of  $\ell$  to speed up the convergence to equilibrium, by maximizing  $z(\ell)$  (see Equation (54) and [23]).
- On time intervals such that  $\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n})) > 0$ , with the scaling  $\sigma_n^2 = \frac{\ell^2}{n^{1/2}}$ , we observe that  $w(t, \ell) = \ell^2$  in (53) so that one should take  $\ell$  as large as possible. This is an indication of the fact that the correct scaling for  $\sigma_n^2$  in this case should be such that  $\sigma_n^2 \gg \frac{\ell^2}{n^{1/2}}$ . Indeed, in the Gaussian case, Proposition 4 below shows that one should take  $\sigma_n$  going to 0 as slowly as possible.

In conclusion, in the MALA case (and contrary to the RWM case), the correct scaling as a function of the dimension is not the same at equilibrium and in the transient phase. Moreover, in the transient phase, the scaling depends on the sign of

$$\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n})).$$

It seems thus difficult to draw any general simple recommendation for practitioners from this analysis. It is likely that the assumption that the target probability is the product of  $n$  one-dimensional laws is too restrictive to understand correctly the scaling  $n \rightarrow \infty$  in this case.

### 6.2. Rigorous results in the Gaussian case and when $\mathbb{E}((X_k^{1,n})^2) > 1$

In this section, we consider the case of a Gaussian target, namely

$$V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi). \tag{55}$$

We thus have

$$(V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2 = x^2 - 1.$$

The aim of this section is to study in details the situation when

$$\mathbb{E}(((V')^2 V'' + V^{(4)} - 2V^{(3)} V' - (V'')^2)(X_k^{1,n})) > 0 \quad \text{namely} \quad \mathbb{E}((X_k^{1,n})^2) > 1.$$

**Proposition 3.** *Let us consider  $(X_k^{i,n})$  solution to (51) for the Gaussian target (55), with a variance independent of  $n$ :*

$$\sigma_n = \ell \in (0, 2).$$

*Let  $m$  be a probability measure on  $\mathbb{R}$  such that  $\langle m, x^2 \rangle > \frac{1}{1-\ell^2/4}$ . We endow the space  $\mathbb{R}^{\mathbb{N}}$  with the product topology. If the initial random variables  $(X_0^{1,n}, \dots, X_0^{n,n})$  are exchangeable and  $m$ -chaotic, then the processes  $(X^{1,n}, \dots, X^{n,n})$  are  $P$ -chaotic where  $P$  denotes the law of the Markov chain*

$$Y_{k+1} = \left(1 - \frac{\ell^2}{2}\right) Y_k + \ell G_{k+1} \tag{56}$$

*with the sequence  $(G_k)$  i.i.d. according to the normal law and independent from the initial position  $Y_0$  distributed according to  $m$ .*

A simple case for which the assumption on the initial condition is satisfied is i.i.d. initial conditions  $(X_0^{i,n})_i$  with law  $m$ .

Notice that  $Y_k$  converges in law to  $\mathcal{N}(0, \frac{1}{1-\ell^2/4})$  as  $k \rightarrow +\infty$ . The asymptotic distribution converges to the target density when  $\ell \rightarrow 0$ . Of course, for fixed  $n$  and  $i \in \{1, \dots, n\}$ ,  $X_k^{i,n}$  converges in law to  $\mathcal{N}(0, 1)$  as  $k \rightarrow +\infty$ . So the limits  $k \rightarrow \infty$  and  $n \rightarrow \infty$  do not commute, meaning that, for large  $n$ , the rate of convergence in distribution of  $(X_k^{i,n})_{k \geq 1}$  to  $\mathcal{N}(0, 1)$  should deteriorate.

**Proof of Proposition 3.** Let  $(Y^{1,n}, \dots, Y^{n,n})$  with  $Y_0^{i,n} = X_0^{i,n}$  and  $Y_{k+1}^{i,n} = (1 - \frac{\ell^2}{2}) Y_k^{i,n} + \ell G_{k+1}^i$  denote the processes obtained when all moves are accepted in the MALA algorithm (51). The proof is divided into two steps. We are first going to prove that the processes  $(Y^{1,n}, \dots, Y^{n,n})$  are

$P$ -chaotic (this would be trivial if the initial conditions  $(X_0^{i,n})_i$  were supposed to be i.i.d.). Then, setting

$$\tilde{\mathcal{A}}_{k+1}^n = \left\{ U_{k+1} \leq \exp \left\{ \sum_{i=1}^n \left( V(Y_k^{i,n}) - V(Y_{k+1}^{i,n}) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\ell}{2} (V'(Y_k^{i,n}) + V'(Y_{k+1}^{i,n})) \right)^2 \right] \right) \right\} \right\},$$

we will check that  $\forall K \in \mathbb{N}^*, \lim_{n \rightarrow \infty} \mathbb{P}(\bigcap_{k=1}^K \tilde{\mathcal{A}}_k^n) = 1$ . Since, on the event  $\bigcap_{k=1}^K \tilde{\mathcal{A}}_k^n, (X_k^{1,n}, \dots, X_k^{n,n})_{0 \leq k \leq K} = (Y_k^{1,n}, \dots, Y_k^{n,n})_{0 \leq k \leq K}$  one obtains the  $P$ -chaoticity of the processes  $(X^{1,n}, \dots, X^{n,n})$  by combining the two steps.

For the first step, notice that for fixed  $j, K \in \mathbb{N}^*$ , the law of  $((Y_k^{1,n}, \dots, Y_k^{j,n}))_{0 \leq k \leq K}$  is

$$m_j^n(dy_0^1, \dots, dy_0^j) \prod_{k=0}^{K-1} (Q(y_k^1, dy_{k+1}^1) \times \dots \times Q(y_k^j, dy_{k+1}^j)),$$

where  $Q(y, dy') = \frac{1}{\ell\sqrt{2\pi}} e^{-(y'-y(1-\ell^2/2))^2/(2\ell^2)} dy'$  and the law  $m_j^n$  of  $(X_0^{1,n}, \dots, X_0^{j,n})$  converges weakly to  $m^{\otimes j}$  as  $n \rightarrow \infty$  (since the initial conditions  $(X_0^{1,n}, \dots, X_0^{n,n})$  are  $m$ -chaotic). Since  $y \mapsto Q(y, dy') \in \mathcal{P}(\mathbb{R})$  is weakly continuous, this law converges weakly to  $\prod_{i=1}^j (m(dy_0^i) \prod_{k=0}^{K-1} Q(y_k^i, dy_{k+1}^i))$  which is the  $j$ -fold product of the image of  $P$  by the canonical restriction to the  $K+1$  first coordinates. Hence, the processes  $(Y^{1,n}, \dots, Y^{n,n})$  are  $P$ -chaotic.

For the second step, let us introduce

$$S_k^n = \frac{8}{n\ell^4} \sum_{i=1}^n \left\{ V(Y_k^{i,n}) - V(Y_{k+1}^{i,n}) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\ell}{2} (V'(Y_k^{i,n}) + V'(Y_{k+1}^{i,n})) \right)^2 \right] \right\}.$$

One has

$$\mathbb{P}(\{\tilde{\mathcal{A}}_{k+1}^n\}^c) \leq \mathbb{P}(S_k^n < 0). \tag{57}$$

Some tedious but simple computations yields (using  $V(x) = \frac{x^2}{2} + \frac{1}{2} \ln(2\pi)$ )

$$\begin{aligned} & V(Y_k^{i,n}) - V(Y_{k+1}^{i,n}) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\ell}{2} (V'(Y_k^{i,n}) + V'(Y_{k+1}^{i,n})) \right)^2 \right] \\ &= \frac{\ell^4}{8} ((Y_k^{i,n})^2 - (G_{k+1}^i)^2) + \left( \frac{\ell^5}{8} - \frac{\ell^3}{4} \right) Y_k^{i,n} G_{k+1}^i - \frac{\ell^6}{32} (Y_k^{i,n})^2 \end{aligned} \tag{58}$$

so that (in law)

$$S_k^n = \left( 1 - \frac{\ell^2}{4} \right) \langle \mu_k^n, y^2 \rangle + \left( \ell - \frac{2}{\ell} \right) \sqrt{\langle \mu_k^n, y^2 \rangle} \frac{\tilde{G}_{k+1}^n}{\sqrt{n}} - \frac{1}{n} \sum_{i=1}^n (G_{k+1}^i)^2$$



with  $\tilde{G}_{k+1}^n = 1_{\{(\mu_k^n, y^2) > 0\}} \frac{\sum_{i=1}^n Y_k^i G_{k+1}^i}{\sqrt{n(\mu_k^n, y^2)}} + 1_{\{(\mu_k^n, y^2) = 0\}} G_{k+1}^1$  a normal random variable independent from  $\mu_k^n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_k^{i,n}}$ . As the exchangeability of the initial condition  $(Y_0^{1,n}, \dots, Y_0^{n,n})$  is preserved by the evolution, the propagation of chaos result obtained in the first step implies (and is actually equivalent to) the convergence in probability of the empirical measures  $\mu^n = \frac{1}{n} \sum_{i=1}^n \delta_{Y^{i,n}} \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$  to  $P$  (see [25], Proposition 2.2). In particular,  $\mu_k^n$  converges in probability to the law  $P_k$  of  $Y_k$ , solution to (56).

With this law of large numbers, we see that in order to estimate  $\mathbb{P}(S_k^n < 0)$  we need to understand the evolution of  $\langle P_k, y^2 \rangle = \mathbb{E}((Y_k)^2)$  with  $k$ . One has  $\langle P_{k+1}, y^2 \rangle = (1 - \frac{\ell^2}{2})^2 \langle P_k, y^2 \rangle + \ell^2$ , and since  $\langle P_0, y^2 \rangle = \langle m, y^2 \rangle > \frac{1}{1-\ell^2/4}$ , one easily checks by induction that for all  $k \in \mathbb{N}$ ,  $\langle P_k, y^2 \rangle > \frac{1}{1-\ell^2/4}$ . Hence for fixed  $k \in \mathbb{N}$ , there exists  $M < +\infty$  and  $\varepsilon > 0$  such that  $\langle P_k, y^2 \wedge M \rangle \geq \frac{1+3\varepsilon}{1-\ell^2/4}$ . One has

$$\begin{aligned} \mathbb{P}(S_k^n < 0) &\leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n (G_{k+1}^i)^2 > 1 + \varepsilon\right) \\ &\quad + \mathbb{P}\left(\langle \mu_k^n, y^2 \rangle \left(1 + \frac{\ell - 2/\ell}{1 - \ell^2/4} \frac{\tilde{G}_{k+1}^n}{\sqrt{n(\mu_k^n, y^2)}}\right) < \frac{1 + \varepsilon}{1 - \ell^2/4}\right) \\ &\leq \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n (G_{k+1}^i)^2 > 1 + \varepsilon\right) + \mathbb{P}\left(\langle \mu_k^n, y^2 \wedge M \rangle < \frac{1 + 2\varepsilon}{1 - \ell^2/4}\right) \\ &\quad + \mathbb{P}\left(\langle \mu_k^n, y^2 \rangle \geq \frac{1 + 2\varepsilon}{1 - \ell^2/4}, \frac{\ell - 2/\ell}{1 - \ell^2/4} \tilde{G}_{k+1}^n < -\frac{\varepsilon \sqrt{n}}{\sqrt{(1 + 2\varepsilon)(1 - \ell^2/4)}}\right). \end{aligned}$$

The first term of the right-hand side converges to 0 as  $n \rightarrow +\infty$ , since, by the strong law of large numbers,  $\frac{1}{n} \sum_{i=1}^n (G_{k+1}^i)^2$  converges a.s. to 1. The second term converges to 0 since  $\langle \mu_k^n, y^2 \wedge M \rangle$  converges in probability to  $\langle P_k, y^2 \wedge M \rangle \geq \frac{1+3\varepsilon}{1-\ell^2/4}$ . The third term is bounded from above by  $\Phi\left(-\frac{\varepsilon}{|\ell-2/\ell|} \sqrt{\frac{n(1-\ell^2/4)}{(1+2\varepsilon)}}\right)$  and also converges to 0. Hence,  $\mathbb{P}(S_k^n < 0)$  tends to 0 as  $n \rightarrow \infty$  and with (57), one deduces that for fixed  $K \in \mathbb{N}^*$ ,  $\mathbb{P}(\bigcap_{k=1}^K \tilde{\mathcal{A}}_k^n) \geq 1 - \sum_{k=1}^K \mathbb{P}(\{\tilde{\mathcal{A}}_k^n\}^c)$  tends to 1.  $\square$

As is clear from the previous proposition, for a fixed variance  $\sigma_n = \ell$  and if  $\mathbb{E}((X_0^{1,n})^2) > 1$ , then, for sufficiently small  $\ell$  (namely  $\ell < 2\sqrt{1 - 1/\mathbb{E}((X_0^{1,n})^2)}$ ) and in the limit  $n \rightarrow \infty$ , (i) the components  $(X_k^{i,n})_i$  do not interact and evolve independently according to the explicit Euler discretization (56) (with a timestep  $\ell^2$ ) of the Langevin dynamics  $dY_t = dB_t - Y_t/2 dt$  and (ii) the system remains in the region  $\mathbb{E}(X_k^2) > 1$  for all  $k \geq 0$ .

Based on the previous result, it is natural to look for a diffusive limit for a  $\sigma_n$  which goes to zero at an arbitrary rate with respect to  $n$ .

**Proposition 4.** *Let us consider  $(X_k^{i,n})$  solution to (51) for the Gaussian target (55), with a variance  $\sigma_n$  satisfying:*

$$\lim_{n \rightarrow \infty} \sigma_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n\sigma_n^2 = +\infty.$$

*Let  $m$  be a probability measure on  $\mathbb{R}$  such that  $\langle m, x^2 \rangle > 1$  and  $\langle m, x^8 \rangle < +\infty$ . If the initial random variables  $(X_0^{1,n}, \dots, X_0^{n,n})$  are i.i.d. according to  $m$ , then the processes  $((X_{\lfloor t/\sigma_n^2 \rfloor}^{1,n})_{t \geq 0}, \dots, (X_{\lfloor t/\sigma_n^2 \rfloor}^{n,n})_{t \geq 0})$  are  $Q$ -chaotic where  $Q$  denotes the law of the Ornstein–Uhlenbeck process*

$$dY_t = dB_t - \frac{Y_t}{2} dt \tag{59}$$

*with the initial position  $Y_0$  distributed according to  $m$  and independent from the Brownian motion  $(B_t)_{t \geq 0}$ . Moreover, the limiting mean acceptance rate is 1.*

**Remark 4.** For a more general potential  $V$ , if the initial random variables  $(X_0^{1,n}, \dots, X_0^{n,n})$  are exchangeable and  $m$ -chaotic with  $\langle m, (V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2 \rangle > 0$ , one expects the limit in law to be the one of the solution of  $Y_t = Y_0 + B_t - \int_0^t \frac{V'(Y_s)}{2} ds$ . But, unlike in the Gaussian case, it is not clear that  $\mathbb{E}[\{(V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2\}(Y_t)] > 0$  for all  $t \geq 0$ . Therefore, setting  $T = \inf\{t > 0 : \mathbb{E}[\{(V')^2 V'' + V^{(4)} - 2V^{(3)}V' - (V'')^2\}(Y_t)] = 0\}$  with the convention  $\inf \emptyset = +\infty$  and denoting by  $Q^T$  the law of  $(Y_t)_{t \in [0, T]}$ , one actually expects the processes  $((X_{\lfloor t/\sigma_n^2 \rfloor}^{1,n})_{t \in [0, T]}, \dots, (X_{\lfloor t/\sigma_n^2 \rfloor}^{n,n})_{t \in [0, T]})$  to be  $Q^T$ -chaotic.

**Proof of Proposition 4.** As in the proof of Proposition 3, let  $(Y^{1,n}, \dots, Y^{n,n})$  with  $Y_0^{i,n} = X_0^{i,n}$  and  $Y_{k+1}^{i,n} = (1 - \frac{\sigma_n^2}{2})Y_k^{i,n} + \sigma_n G_{k+1}^i$  denote the processes obtained when all moves are accepted in the MALA algorithm (51). The processes  $(Y_{\lfloor t/\sigma_n^2 \rfloor}^{1,n}, \dots, Y_{\lfloor t/\sigma_n^2 \rfloor}^{n,n})$  are independent and identically distributed and their common distribution converges weakly to  $Q$  by the strong convergence analysis of the Euler scheme applied to (59). Hence, to conclude the proof, it is enough to check that for fixed  $T > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}(\cap_{k=1}^{\lfloor T/\sigma_n^2 \rfloor} \tilde{\mathcal{A}}_k^n) = 1$ , where, as in the proof of Proposition 3,

$$\begin{aligned} \tilde{\mathcal{A}}_{k+1}^n = & \left\{ U_{k+1} \leq \exp \left\{ \sum_{i=1}^n \left( V(Y_k^{i,n}) - V(Y_{k+1}^{i,n}) \right. \right. \right. \\ & \left. \left. \left. + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\sigma_n}{2} (V'(Y_k^{i,n}) + V'(Y_{k+1}^{i,n})) \right)^2 \right] \right) \right\} \right\}. \end{aligned}$$

To do so, we use an upper-bound sharper than (57). Let us introduce (using (58)):

$$\begin{aligned} S_k^n &= \frac{8}{n\sigma_n^4} \sum_{i=1}^n \left\{ V(Y_k^{i,n}) - V(Y_{k+1}^{i,n}) + \frac{1}{2} \left[ (G_{k+1}^i)^2 - \left( G_{k+1}^i - \frac{\sigma_n}{2} (V'(Y_k^{i,n}) + V'(Y_{k+1}^{i,n})) \right)^2 \right] \right\} \\ &= \frac{2}{n\sigma_n} \sum_{i=1}^n R_k^{i,n}, \end{aligned}$$

where the random variables

$$\left( R_k^{i,n} = \left( \frac{\sigma_n}{2} - \frac{\sigma_n^3}{8} \right) (Y_k^{i,n})^2 + \left( \frac{\sigma_n^2}{2} - 1 \right) Y_k^{i,n} G_{k+1}^i - \frac{\sigma_n}{2} (G_{k+1}^i)^2 \right)_{1 \leq i \leq n}$$

are independent and identically distributed. Then, we have

$$\begin{aligned} \mathbb{P}(\{\tilde{\mathcal{A}}_{k+1}^n\}^c) &= \mathbb{E}((1 - e^{(n\sigma_n^4/8)S_k^n})^+) \leq -\mathbb{E}\left(\frac{n\sigma_n^4}{8} S_k^n 1_{\{S_k^n \leq 0\}}\right) \\ &= -\frac{\sigma_n^3}{4} \mathbb{E}\left(\sum_{i=1}^n R_k^{i,n} 1_{\{\sum_{i=1}^n R_k^{i,n} \leq 0\}}\right). \end{aligned} \tag{60}$$

We need to estimate the moments of the random variables  $R_k^{i,n}$ . To do so, we assume from now on that  $n$  is large enough so that  $\sigma_n < \sqrt{2}$  and we first estimate the moments of

$$Y_k^{i,n} = \left(1 - \frac{\sigma_n^2}{2}\right)^k X_0^{i,n} + \sigma_n \sum_{j=1}^k \left(1 - \frac{\sigma_n^2}{2}\right)^{k-j} G_j^i.$$

One has, using the fact that  $\sigma_n \sum_{j=1}^k (1 - \frac{\sigma_n^2}{2})^{k-j} G_j^i \sim \mathcal{N}(0, \frac{1-(1-\sigma_n^2/2)^{2k}}{1-\sigma_n^2/4})$ ,

$$\begin{aligned} \mathbb{E}((Y_k^{i,n})^8) &= \left(1 - \frac{\sigma_n^2}{2}\right)^{8k} \langle m, x^8 \rangle + 28 \left(1 - \frac{\sigma_n^2}{2}\right)^{6k} \frac{1 - (1 - \sigma_n^2/2)^{2k}}{1 - \sigma_n^2/4} \langle m, x^6 \rangle \\ &\quad + 210 \left(1 - \frac{\sigma_n^2}{2}\right)^{4k} \left(\frac{1 - (1 - \sigma_n^2/2)^{2k}}{1 - \sigma_n^2/4}\right)^2 \langle m, x^4 \rangle \\ &\quad + 420 \left(1 - \frac{\sigma_n^2}{2}\right)^{2k} \left(\frac{1 - (1 - \sigma_n^2/2)^{2k}}{1 - \sigma_n^2/4}\right)^3 \langle m, x^2 \rangle \\ &\quad + 105 \left(\frac{1 - (1 - \sigma_n^2/2)^{2k}}{1 - \sigma_n^2/4}\right)^4 \\ &\leq \langle m, x^8 \rangle + 56 \langle m, x^6 \rangle + 840 \langle m, x^4 \rangle + 3360 \langle m, x^2 \rangle + 1680. \end{aligned}$$

Therefore,  $\sup_{n:\sigma_n < \sqrt{2}} \sup_{k \geq 0} \mathbb{E}((R_k^{i,n})^4) < +\infty$ . Moreover, for  $n$  large enough so that  $\sigma_n^2 \leq \frac{2(\langle m, x^2 \rangle - 1)}{\langle m, x^2 \rangle}$  (so that we also have  $\sigma_n^2 < 2$ ),

$$\begin{aligned} \left(1 - \frac{\sigma_n^2}{4}\right) \mathbb{E}((Y_k^{i,n})^2) &= 1 + \left(1 - \frac{\sigma_n^2}{2}\right)^{2k} \left(\left(1 - \frac{\sigma_n^2}{4}\right) \langle m, x^2 \rangle - 1\right) \\ &\geq 1 + e^{-(2k\sigma_n^2/(2-\sigma_n^2))} \frac{\langle m, x^2 \rangle - 1}{2} \geq 1 + e^{-\langle m, x^2 \rangle T} \frac{\langle m, x^2 \rangle - 1}{2}, \end{aligned}$$

where the latter inequality holds for  $k \leq \lfloor T/\sigma_n^2 \rfloor$ . From now on, we suppose that  $n$  is large enough so that  $\sigma_n^2 \leq \frac{2\langle m, x^2 \rangle - 1}{\langle m, x^2 \rangle}$  and we fix  $k \leq \lfloor T/\sigma_n^2 \rfloor$ . Setting  $c_T = (e^{-\langle m, x^2 \rangle T})^{\frac{\langle m, x^2 \rangle - 1}{4}}$ , one has

$$\mathbb{E}(R_k^{i,n}) = \frac{\sigma_n}{2} \left( \left( 1 - \frac{\sigma_n^2}{4} \right) \mathbb{E}((Y_k^{i,n})^2) - 1 \right) \geq c_T \sigma_n.$$

Therefore (using in particular the fact that  $\mathbb{E}(R_k^{i,n}) \geq 0$ ),

$$\begin{aligned} -\mathbb{E} \left( \sum_{i=1}^n R_k^{i,n} 1_{\{\sum_{i=1}^n R_k^{i,n} \leq 0\}} \right) &= -n \mathbb{E}(R_k^{1,n}) \mathbb{P} \left( \sum_{i=1}^n R_k^{i,n} \leq 0 \right) \\ &\quad - \mathbb{E} \left( \sum_{i=1}^n (R_k^{i,n} - \mathbb{E}(R_k^{i,n})) 1_{\{\sum_{i=1}^n (R_k^{i,n} - \mathbb{E}(R_k^{i,n})) \leq -n \mathbb{E}(R_k^{1,n})\}} \right) \\ &\leq 0 + \frac{\mathbb{E}((\sum_{i=1}^n (R_k^{i,n} - \mathbb{E}(R_k^{i,n})))^4)}{(n \mathbb{E}(R_k^{1,n}))^3} \\ &= \frac{3n(n-1) \text{Var}^2((R_k^{1,n})) + n \mathbb{E}((R_k^{1,n} - \mathbb{E}(R_k^{1,n}))^4)}{(n \mathbb{E}(R_k^{1,n}))^3} \\ &\leq \frac{(3n^2 + 13n) \mathbb{E}((R_k^{1,n})^4)}{c_T^3 n^3 \sigma_n^3} \leq \frac{C_T}{n \sigma_n^3}, \end{aligned} \quad (61)$$

where  $C_T$  is some constant not depending on  $n$  and  $k$ . With (60), we deduce that

$$\mathbb{P} \left( \bigcup_{k=0}^{\lfloor T/\sigma_n^2 \rfloor - 1} \{\tilde{\mathcal{A}}_{k+1}^n\}^c \right) \leq \sum_{k=0}^{\lfloor T/\sigma_n^2 \rfloor - 1} \mathbb{P}(\{\tilde{\mathcal{A}}_{k+1}^n\}^c) \leq \frac{T C_T}{4n \sigma_n^2}. \quad (62)$$

Since  $\lim_{n \rightarrow \infty} n \sigma_n^2 = +\infty$ , we conclude that  $\lim_{n \rightarrow \infty} \mathbb{P}(\bigcap_{k=1}^{\lfloor T/\sigma_n^2 \rfloor} \tilde{\mathcal{A}}_k^n) = 1$ .  $\square$

**Remark 5.** In the case when  $\lim_{n \rightarrow \infty} n \sigma_n^2 = 0$  and the initial conditions  $(X_0^{1,n}, \dots, X_0^{n,n})$  are i.i.d. according to  $m$  such that  $\langle m, x^4 \rangle < +\infty$ , then, whatever the sign of  $\langle m, x^2 \rangle - 1$ , the processes  $((X_{\lfloor t/\sigma_n^2 \rfloor}^{1,n})_{t \geq 0}, \dots, (X_{\lfloor t/\sigma_n^2 \rfloor}^{n,n})_{t \geq 0})$  are  $Q$ -chaotic where  $Q$  denotes the law of the Ornstein–Uhlenbeck process  $Y_t = Y_0 + B_t - \int_0^t \frac{Y_s}{2} ds$  with the initial position  $Y_0$  distributed according to  $m$  and independent from the Brownian motion  $(B_t)_{t \geq 0}$ .

Indeed, for  $n$  large enough so that  $\sigma_n < \sqrt{2}$ , one may check that  $\sup_{k \geq 0} \mathbb{E}((Y_k^{i,n})^4) < C$  and replace (61) by the estimation

$$-\mathbb{E} \left( \sum_{i=1}^n R_k^{i,n} 1_{\{\sum_{i=1}^n R_k^{i,n} \leq 0\}} \right) \leq n |\mathbb{E}(R_k^{1,n})| + \mathbb{E}^{1/2} \left( \left( \sum_{i=1}^n (R_k^{i,n} - \mathbb{E}(R_k^{i,n})) \right)^2 \right) \leq C n \sigma_n + C \sqrt{n},$$

so that

$$\mathbb{P}\left(\bigcup_{k=0}^{\lfloor T/\sigma_n^2 \rfloor - 1} \{\tilde{A}_{k+1}^n\}^c\right) \leq C_T(n\sigma_n^2 + \sqrt{n}\sigma_n)$$

which converges to zero when  $n$  goes to infinity.

## Appendix A: Proof of Lemma 1

Let us define for  $x \in \mathbb{R}$ ,

$$f(x) = \exp(x^2/2)\Phi(x), \tag{A.1}$$

$$h(x) = xf(x) = x \exp\left(\frac{x^2}{2}\right)\Phi(x). \tag{A.2}$$

The derivative of  $f$  is

$$f'(x) = \frac{1}{\sqrt{2\pi}} + x \exp\left(\frac{x^2}{2}\right)\Phi(x).$$

For  $x \geq 0$ ,  $f'(x) > 0$ . For  $x < 0$ , using the upper-bound in (38), we also obtain  $f'(x) > 0$ . Therefore, the function  $f$  is increasing.

Since  $h'(x) = (1 + x^2) \exp\left(\frac{x^2}{2}\right)\Phi(x) + \frac{x}{\sqrt{2\pi}}$ , it is obvious that  $h'(x) > 0$  for  $x \geq 0$ . For  $x < 0$  this comes from the lower-bound in (38).

By definitions of  $\Gamma$  and  $\mathcal{G}$ , we get

$$\frac{\Gamma(a, b, \ell) - 2\mathcal{G}(a, b, \ell)}{\ell^2} = \Phi\left(\frac{-\ell b}{2\sqrt{a}}\right) - \exp\left(\frac{\ell^2}{2}(a - b)\right)\Phi\left(\frac{\ell b}{2\sqrt{a}} - \ell\sqrt{a}\right). \tag{A.3}$$

Using the identity

$$\exp\left(\frac{\ell^2}{2}(a - b)\right) = \exp\left[\frac{\ell^2}{2}\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right)^2\right] \exp\left(\frac{-\ell^2 b^2}{8a}\right),$$

the right-hand side of (A.3) can be rewritten in terms of  $f$  (defined by (A.1))

$$\frac{\Gamma(a, b, \ell) - 2\mathcal{G}(a, b, \ell)}{\ell^2} = \exp\left(\frac{-\ell^2 b^2}{8a}\right) \left[ f\left(\frac{-\ell b}{2\sqrt{a}}\right) - f\left(\frac{\ell(b - 2a)}{2\sqrt{a}}\right) \right].$$

Now it is clear that

$$\text{sign}(\Gamma(a, b, \ell) - 2\mathcal{G}(a, b, \ell)) = \text{sign}\left[ f\left(\frac{-\ell b}{2\sqrt{a}}\right) - f\left(\frac{\ell(b - 2a)}{2\sqrt{a}}\right) \right].$$

Recall that the function  $f$  is increasing and thus  $\text{sign}(\Gamma(a, b, \ell) - 2\mathcal{G}(a, b, \ell)) = \text{sign}(a - b)$ .

Similarly,

$$F(a, b, \ell) = \begin{cases} 2\ell\sqrt{a} \exp\left(\frac{-\ell^2 b^2}{8a}\right) \frac{h(\ell(b-2a)/(2\sqrt{a})) - h(-\ell b/(2\sqrt{a}))}{b-a}, & \text{if } a \neq b, \\ 2\ell^2 \exp\left(\frac{-\ell^2 a}{8}\right) h'\left(-\frac{\ell\sqrt{a}}{2}\right), & \text{if } a = b. \end{cases}$$

This shows the continuity of  $F$ , and the positivity of  $F$  is a consequence of the positivity of  $h'$ .

Setting for  $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ ,  $\chi(a, b, \ell) = \ell \frac{b-2a}{2\sqrt{a}}$  and  $\zeta(a, b, \ell) = -\frac{\ell b}{2\sqrt{a}}$  and for  $(x, y) \in \mathbb{R}^2$ ,

$$\psi(x, y) = \begin{cases} \frac{-(x+y)(f(x) - f(y))}{(f(x) + f(y))(x - y)}, & \text{if } y \neq x, \\ -\frac{xf'(x)}{f(x)}, & \text{otherwise,} \end{cases}$$

one has

$$F(a, b, \ell) = \begin{cases} \Gamma(0, b, \ell), & \text{if } a = 0, \\ \Gamma(a, b, \ell)(1 - \psi(\chi(a, b, \ell), \zeta(a, b, \ell))), & \text{otherwise.} \end{cases}$$

By [14], Lemma 2, Equation (3.2), the function  $(a, b) \mapsto \Gamma(a, b, \ell)$  is bounded from below by a positive constant on  $[0, +\infty] \times [-M, M]$ . To show (21), it is then sufficient to show that  $\sup_{(a,b) \in (0, M] \times \mathbb{R}} \psi(\chi(a, b, \ell), \zeta(a, b, \ell)) < 1$ .

When  $x > y$ , since  $h$  is increasing,  $yf(y) < xf(x)$  which implies  $-(x+y)(f(x) - f(y)) < (f(x) + f(y))(x - y)$  and therefore  $\psi(x, y) < 1$ . This inequality remains valid for  $y > x$  by symmetry of  $\psi$  and for  $y = x$  since  $xf'(x) + f(x) = h'(x) > 0$ .

For  $(x, y) \in \mathbb{R}^2$ , with  $x > 0$  and  $-\ell\sqrt{M} \leq x + y \leq 0$ , (so that  $x - y \geq 2x > 0$ ) one has  $0 \leq -\frac{x+y}{x-y} \leq \frac{\ell\sqrt{M}}{2x}$  and  $0 < \frac{f(x)-f(y)}{f(x)+f(y)} < 1$  so that  $\psi(x, y) \leq \frac{\ell\sqrt{M}}{2x}$ . With the symmetry of  $\psi$ , one deduces that  $\sup_{(x,y): -\ell\sqrt{M} \leq x+y \leq 0, x \vee y \geq \ell\sqrt{M}} \psi(x, y) \leq \frac{1}{2}$ . Since  $f$  is  $\mathcal{C}^1$  and positive, one easily checks that  $\psi$  is continuous on  $\mathbb{R}^2$ . As  $\psi < 1$  and  $\{(x, y) : -\ell\sqrt{M} \leq x + y \leq 0, x \vee y \leq \ell\sqrt{M}\}$  is compact, one obtains that  $\sup_{(x,y): -\ell\sqrt{M} \leq x+y \leq 0} \psi(x, y) < 1$ .

As for  $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$ ,  $\chi(a, b, \ell) + \zeta(a, b, \ell) = -\ell\sqrt{a}$ , one concludes that

$$\sup_{(a,b) \in (0, M] \times \mathbb{R}} \psi(\chi(a, b, \ell), \zeta(a, b, \ell)) < 1.$$

## Appendix B: Proof of Lemma 2

Recall first that the function  $(s, \ell) \mapsto F_1(s, \ell)$  is  $\mathcal{C}^\infty$  on  $\mathbb{R}_+ \times \mathbb{R}_+$ . It is easily checked that for any  $s \geq 0$ ,  $F_1(s, 0) = 0$  and  $\lim_{\ell \rightarrow \infty} F_1(s, \ell) = 0$ . With (21) and the continuity of  $\ell \mapsto F_1(s, \ell)$ , one deduces the existence of a point  $\ell^*(s) > 0$  such that  $F_1(s, \ell^*(s)) = \max_{\ell \geq 0} F_1(s, \ell)$ .

When  $s = 0$ ,  $F_1(0, \ell) = \ell^2 \exp(-\frac{\ell^2}{2})$ . This function admits a unique maximum at point  $\ell^*(0) = \sqrt{2}$ . For further use, we observe that

$$\frac{\partial^2 F_1}{\partial \ell^2}(0, \ell^*(0)) \neq 0. \tag{B.1}$$

In the case  $s = 1$ , we compute the derivatives

$$\begin{aligned} \frac{\partial F_1}{\partial \ell}(1, \ell) &= (4\ell + 2\ell^3)\Phi\left(-\frac{\ell}{2}\right) - \frac{4\ell^2}{\sqrt{2\pi}} \exp\left(-\frac{\ell^2}{8}\right), \\ \frac{\partial^2 F_1}{\partial \ell^2}(1, \ell) &= (4 + 6\ell^2)\Phi\left(-\frac{\ell}{2}\right) - \frac{10\ell}{\sqrt{2\pi}} \exp\left(-\frac{\ell^2}{8}\right). \end{aligned}$$

As a consequence, at a critical point of  $\ell \mapsto F_1(1, \ell)$ ,

$$\frac{\partial^2 F_1}{\partial \ell^2}(1, \ell) = (\ell^2 - 6)\Phi\left(-\frac{\ell}{2}\right). \tag{B.2}$$

We deduce that any local maximum belongs to  $(0, \sqrt{6}]$  and any local minimum to  $[\sqrt{6}, +\infty)$ . Since there is a local minimum (resp. maximum) between two distinct local maxima (resp. minima), we conclude that  $\ell \mapsto F_1(1, \ell)$  admits a unique local maximum which is also a global maximum and belongs to  $(0, \sqrt{6}]$  and no local minimum on  $(0, +\infty)$ . For further use, we observe that  $\frac{\partial F_1}{\partial \ell}(1, \sqrt{6}) \neq 0$  and thus (from (B.2))

$$\frac{\partial^2 F_1}{\partial \ell^2}(1, \ell^*(1)) \neq 0. \tag{B.3}$$

Let us now consider the case  $s \in (0, 1) \cup (1, \infty)$ . The partial derivative of  $F_1$  with respect to  $\ell$  is:

$$\frac{\partial F_1}{\partial \ell}(s, \ell) = \left(\frac{2}{\ell} - \ell(1-s)\right)F_1(s, \ell) + \ell^2 \left(-\sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + \ell\Phi\left(-\frac{\ell}{2\sqrt{s}}\right)\right). \tag{B.4}$$

Of course,  $\frac{\partial F_1}{\partial \ell}(s, \ell^*(s)) = 0$ . Then, at any critical point of  $\ell \mapsto F(s, \ell)$ , we have (using the fact that  $\frac{\partial F_1}{\partial \ell}(s, \ell) = 0 \iff \frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = \tilde{\rho}(s, \ell)$  where

$$\tilde{\rho}(s, \ell) = \left(-\frac{2}{\ell^2} - 1 + s\right)F_1(s, \ell) - 2\ell\sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + 3\ell^2\Phi\left(-\frac{\ell}{2\sqrt{s}}\right)$$

so that  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = \rho(s, \ell)$  with (using again  $\frac{\partial F_1}{\partial \ell}(s, \ell) = 0$  to eliminate  $F_1(s, \ell)$ )

$$\begin{aligned} \rho(s, \ell) &= \ell \frac{-\ell^2 + s\ell^2 + 6}{\ell^2 - s\ell^2 - 2} \sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + 2\ell^2 \frac{\ell^2 - s\ell^2 - 4}{\ell^2 - s\ell^2 - 2} \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) \\ &= 2\ell^2 \frac{\ell^2 - s\ell^2 - 4}{\ell^2 - s\ell^2 - 2} \chi(s, \ell), \end{aligned}$$

where

$$\chi(s, \ell) = \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) - \frac{1}{\ell} \frac{\ell^2 - s\ell^2 - 6}{\ell^2 - s\ell^2 - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right).$$

**B.1. The case  $s > 1$**

Let us assume  $s > 1$ . In this section, we will prove that the function  $\ell \mapsto \rho(s, \ell)$  is negative on some interval  $(0, \tilde{\ell})$  and positive on  $(\tilde{\ell}, \infty)$ , which is equivalent to show that  $\ell \mapsto \chi(s, \ell)$  is negative on  $(0, \tilde{\ell})$  and positive on  $(\tilde{\ell}, \infty)$ , since the ratio  $\frac{\ell^2 - s\ell^2 - 4}{\ell^2 - s\ell^2 - 2}$  is positive. This implies that  $\ell \mapsto F(s, \ell)$  has a unique global maximum at point  $\ell^*(s)$ . Indeed, if  $\ell_1^*(s) < \ell_2^*(s)$  are two points in  $\arg \max_{\ell \geq 0} F_1(s, \ell)$ , then,  $\ell_2^*(s) < \tilde{\ell}$  and we reach a contradiction by noticing that there is necessarily a local minimum of  $\ell \mapsto F(s, \ell)$  in the interval  $(\ell_1^*(s), \ell_2^*(s))$ .

We note that

$$\frac{\partial \chi}{\partial \ell}(s, \ell) = \frac{1}{\sqrt{2\pi s}} \frac{1}{\ell^2} \frac{1}{(\ell^2(1-s) - 4)^2} \exp\left(-\frac{\ell^2}{8s}\right) P(s, \ell^2), \tag{B.5}$$

where

$$\begin{aligned} P(s, y) &= -\frac{(1-s)^2}{4} y^3 + (1-s) \left( s(1-s) + \frac{3}{2} \right) y^2 - (2 + 14s(1-s))y + 24s \\ &= -\frac{(1-s)^2}{4} \left( y - \frac{2}{1-s} \right) \left( y^2 - 4 \left( \frac{1}{1-s} + s \right) y + \frac{48s}{1-s} \right). \end{aligned}$$

We will show that  $y \mapsto P(s, y)$  is positive on some interval  $(0, \bar{\ell})$  and negative on  $(\bar{\ell}, \infty)$ . This means that  $\ell \mapsto \chi(s, \ell)$  is increasing on  $(0, \bar{\ell})$  and decreasing on  $(\bar{\ell}, \infty)$ . Since  $\lim_{\ell \rightarrow 0} \chi(s, \ell) = -\infty$ ,  $\lim_{\ell \rightarrow \infty} \chi(s, \ell) = 0$  and  $\ell \mapsto \chi(s, \ell)$  is a  $C^\infty$  function, this implies that  $\ell \mapsto \chi(s, \ell)$  is negative on some interval  $(0, \tilde{\ell})$  and positive on  $(\tilde{\ell}, \infty)$ , which concludes the proof.

Let us now study the polynomial  $y \mapsto P(s, y)$ . Let us introduce

$$Q(s, y) = y^2 - 4 \left( \frac{1}{1-s} + s \right) y + \frac{48s}{1-s}.$$

The discriminant of  $y \mapsto Q(s, y)$  is

$$\Delta(s) = \frac{16}{(1-s)^2} (s^2(1-s)^2 - 10s(1-s) + 1).$$

Since  $s > 1$ , and thus  $s(1-s) < 0$ , then  $\Delta(s) > 0$ . The polynomial  $y \mapsto Q(s, y)$  has two roots:

$$y_+ = 2 \left( \frac{1}{1-s} + s \right) + \frac{2}{|1-s|} (s^2(1-s)^2 - 10s(1-s) + 1)^{1/2}$$



and

$$y_- = 2\left(\frac{1}{1-s} + s\right) - \frac{2}{|1-s|}(s^2(1-s)^2 - 10s(1-s) + 1)^{1/2}.$$

Then,  $Q(s, y) < 0$  if and only if  $y \in (y_-, y_+)$ . The roots of  $y \mapsto P(s, y)$  are  $\{y_-, y_+, y_0\}$  where

$$y_0 = \frac{2}{1-s}.$$

We notice that  $y_- < y_+$  and  $y_+ > y_0$ . We observe that

$$\begin{aligned} y_- < y_0 &\iff 2\left(\frac{1}{1-s} + s\right) - \frac{2}{|1-s|}(s^2(1-s)^2 - 10s(1-s) + 1)^{1/2} < \frac{2}{1-s} \\ &\iff s < \frac{1}{|1-s|}(s^2(1-s)^2 - 10s(1-s) + 1)^{1/2} \\ &\iff s(1-s) < \frac{1}{10}. \end{aligned}$$

Thus, since  $s > 1$ , we have

$$y_- < y_0 < 0 < y_+,$$

and  $y \mapsto P(s, y)$  changes its sign at each of its roots  $\{y_-, y_+, y_0\}$ . Since  $\lim_{y \rightarrow \infty} P(s, y) = -\infty$ , we deduce that  $P(s, y) > 0$  for  $y \in (0, y_+)$  and  $P(s, y) < 0$  for  $y \in (y_+, \infty)$ . This concludes the proof in the case  $s > 1$ .

## B.2. The case $s < 1$

First, we observe that the maximum of  $\ell \mapsto F_1(s, \ell)$  is necessarily in  $(0, \ell_0)$  where

$$\ell_0 = \sqrt{y_0} = \sqrt{\frac{2}{1-s}}.$$

Indeed, if  $\ell^2 \geq \frac{2}{1-s}$ , we have (using the fact that  $F_1 \geq 0$  and the upper bound in the classical inequality (38)):

$$\begin{aligned} \frac{\partial F_1}{\partial \ell}(s, \ell) &= \left(\frac{2}{\ell} - \ell(1-s)\right)F_1(s, \ell) + \ell^2\left(-\sqrt{\frac{2s}{\pi}}\exp\left(-\frac{\ell^2}{8s}\right) + \ell\Phi\left(-\frac{\ell}{2\sqrt{s}}\right)\right) \\ &< \ell^2\left(-\sqrt{\frac{2s}{\pi}}\exp\left(-\frac{\ell^2}{8s}\right) + \sqrt{\frac{2s}{\pi}}\exp\left(-\frac{\ell^2}{8s}\right)\right) \\ &< 0. \end{aligned}$$

This shows in particular that  $\ell^*(s) \in (0, \ell_0)$ . In all what follows, we only study the function  $\ell \mapsto F_1(s, \ell)$  for

$$\ell \in (0, \ell_0).$$

We need to prove that  $\ell \mapsto F_1(s, \ell)$  admits a unique global maximum on  $(0, \ell_0)$ . A sufficient condition is that  $\ell \mapsto \rho(s, \ell)$  is negative for  $\ell < \ell_0$ .

Notice that the function  $\ell \mapsto \chi(s, \ell)$  is  $C^\infty([0, \ell_0))$ , has the same sign as  $\rho(s, \ell)$  and that  $\lim_{\ell \rightarrow 0} \chi(s, \ell) = -\infty$  while

$$\begin{aligned} \chi(s, \ell_0) &= \Phi\left(-\frac{\ell_0}{2\sqrt{s}}\right) - \frac{1}{\ell_0} \frac{\ell_0^2(1-s) - 6}{\ell_0^2(1-s) - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell_0^2}{8s}\right) \\ &= \Phi\left(-\frac{\ell_0}{2\sqrt{s}}\right) - \frac{1}{\ell_0} \sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell_0^2}{8s}\right) \end{aligned} \tag{B.6}$$

which is negative, using the upper bound in the classical inequality (38).

Let us now study the sign of  $\chi(s, \ell)$ . As in the previous case, we first study the sign of  $\frac{\partial \chi}{\partial \ell}$ , namely the sign of  $P$ . We distinguish between two cases.

If  $s(1-s) \geq 5 - \sqrt{24}$ , then  $\Delta(s) \leq 0$ , so that  $Q(s, y) > 0$  for  $y < y_0$ . This implies that  $P(s, y) > 0$  for  $y < y_0$ . Therefore, in view of (B.5),  $\frac{\partial \chi}{\partial \ell}(s, \ell) > 0$  for  $\ell < \ell_0$ . Thus, in this case,  $\ell \mapsto \chi(s, \ell)$  is increasing from 0 to  $\ell_0$ , going from  $-\infty$  to  $\chi(s, \ell_0)$  which is negative. In conclusion,  $\ell \mapsto \chi(s, \ell)$  is negative on  $(0, \ell_0)$ , and  $\ell \mapsto F_1(s, \ell)$  admits a unique global maximum.

Now, if  $s(1-s) < 5 - \sqrt{24}$ ,  $\Delta(s) > 0$ , so that  $y \mapsto Q(s, y)$  has two roots  $y_+ > y_-$ . We recall that  $y_- < y_0 \iff s(1-s) < \frac{1}{10}$  and notice that  $\frac{1}{10} < 5 - \sqrt{24}$ . Let us thus distinguish between two subcases.

If  $s(1-s) \in [\frac{1}{10}, 5 - \sqrt{24})$ , then  $0 < y_0 \leq y_- < y_+$ . The polynomial  $y \mapsto P(s, y)$  changes its sign at each of its roots  $\{y_0, y_-, y_+\}$ , and  $\lim_{y \rightarrow \infty} P(s, y) = -\infty$ . Thus, in this case,  $\ell \mapsto \chi(s, \ell)$  is increasing from 0 to  $\ell_0$ , going from  $-\infty$  to  $\chi(s, \ell_0)$  which is negative. In conclusion,  $\chi(s, \ell)$  is negative, and  $\ell \mapsto F_1(s, \ell)$  admits a unique global maximum.

The last subcase to consider is  $s(1-s) < \frac{1}{10}$ , which is equivalent to

$$s \in (0, s_0) \cup (s_1, 1)$$

with

$$s_0 = \frac{1}{2} \left(1 - \sqrt{\frac{3}{5}}\right) \quad \text{and} \quad s_1 = \frac{1}{2} \left(1 + \sqrt{\frac{3}{5}}\right).$$

In this case,  $0 < y_- < y_0 < y_+$ . Indeed (using the fact that  $s < 1$ ),

$$\begin{aligned} y_- > 0 &\iff (1 + s(1-s)) > (s^2(1-s)^2 - 10s(1-s) + 1)^{1/2} \\ &\iff s(1-s) > 0, \end{aligned}$$

which is true. The polynomial  $y \mapsto P(s, y)$  changes its sign at each of its roots  $\{y_-, y_0, y_+\}$ , and  $\lim_{y \rightarrow \infty} P(s, y) = -\infty$ . Let us denote  $\ell_- = \sqrt{y_-}$ . Thus, in this case,  $\ell \mapsto \chi(s, \ell)$  is increasing from 0 to  $\ell_-$  (going from  $-\infty$  to  $\chi(s, \ell_-)$ ) and then decreasing from  $\ell_-$  to  $\ell_0$  (going from  $\chi(s, \ell_-)$  to  $\chi(s, \ell_0)$ , which is negative). Thus, if  $\chi(s, \ell_-) < 0$ , then  $\chi(s, \ell)$  is negative, and  $\ell \mapsto F_1(s, \ell)$  admits a unique global maximum.

In conclusion,  $\ell \mapsto F_1(s, \ell)$  admits at least one local maximum and at most two local maxima. The function  $\ell \mapsto F_1(s, \ell)$  admits two local maxima  $\ell_1^* < \ell_2^*$  if and only if  $\chi(s, \ell_-) \geq 0$ , in which case  $\ell_1^* < \ell_- < \ell_2^*$ , and  $\frac{\partial F_1}{\partial \ell}(s, \ell_1^*) = \frac{\partial F_1}{\partial \ell}(s, \ell_2^*) = 0$ .

B.2.1. The case  $s \in (0, s_0)$ 

Let us assume the existence of  $\underline{s} \in (0, s_0)$  such that  $\ell \mapsto F_1(\underline{s}, \ell)$  admits two local maxima  $\ell_1^*(\underline{s}) < \ell_2^*(\underline{s})$  and let us show that

$$\exists (s, \ell) \in [\underline{s}, s_0] \times \mathbb{R}_+^*, \quad \frac{\partial F_1}{\partial \ell}(s, \ell) = \frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = 0. \quad (\text{B.7})$$

If  $\frac{\partial^2 F_1}{\partial \ell^2}(\underline{s}, \ell_1^*(\underline{s})) = 0$  or  $\frac{\partial^2 F_1}{\partial \ell^2}(\underline{s}, \ell_2^*(\underline{s})) = 0$ , we are done. Otherwise, we may apply the implicit function theorem to construct for  $i \in \{1, 2\}$  a continuous curve  $\ell_i^*(s)$  on a maximal interval  $[\underline{s}, \bar{s}_i]$  with  $\bar{s}_i > \underline{s}$  such that for  $s \in [\underline{s}, \bar{s}_i]$ ,  $\frac{\partial F_1}{\partial \ell}(s, \ell_i^*(s)) = 0$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell_i^*(s)) < 0$ . In case  $\min(\bar{s}_1, \bar{s}_2) > s_0$ , then, since by the uniqueness part of the implicit function theorem,  $\forall s \in [\underline{s}, \min(\bar{s}_1, \bar{s}_2))$ ,  $\ell_1^*(s) < \ell_2^*(s)$ , we contradict the fact that  $\ell \mapsto F_1(s_0, \ell)$  admits a unique local maximum. Thus, choosing  $i \in \{1, 2\}$  such that  $\bar{s}_i = \min(\bar{s}_1, \bar{s}_2)$ , one has  $\bar{s}_i \leq s_0$ . Since  $\ell_i^*(s) < \ell_0(s) = \sqrt{\frac{2}{1-s}}$ , we may find an increasing sequence  $(s_n)_{n \in \mathbb{N}}$  of elements of  $[\underline{s}, \bar{s}_i]$  converging to  $\bar{s}_i$  and such that  $\ell_i^*(s_n)$  converges to some limit denoted by  $\ell_i^*(\bar{s}_i)$  as  $n \rightarrow \infty$ . By continuity of  $\frac{\partial F_1}{\partial \ell}(s, \ell)$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell)$ , one has  $\frac{\partial F_1}{\partial \ell}(\bar{s}_i, \ell_i^*(\bar{s}_i)) = 0$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}_i, \ell_i^*(\bar{s}_i)) \leq 0$ . Let us now consider  $\ell_{3-i}^*(\bar{s}_i)$ , defined as the limit of a converging subsequence of  $(\ell_{3-i}^*(s_n))_n$  in case  $\bar{s}_{3-i} = \bar{s}_i$ . If  $\ell_1^*(\bar{s}_i) = \ell_2^*(\bar{s}_i)$ , then from the existence of a local minimum  $\ell \in (\ell_1^*(s_n), \ell_2^*(s_n))$  such that  $\frac{\partial^2 F_1}{\partial \ell^2}(s_n, \ell) \geq 0$ , we conclude that  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}_i, \ell_i^*(\bar{s}_i)) = 0$ . If  $\ell_1^*(\bar{s}_i) < \ell_2^*(\bar{s}_i)$  and both  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}_i, \ell_1^*(\bar{s}_i))$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}_i, \ell_2^*(\bar{s}_i))$  are negative, then, using the implicit function theorem, we contradict the maximality of  $\bar{s}_i$ . This concludes the proof of (B.7).

Let us consider a point  $(s, \ell)$  such that  $\frac{\partial F_1}{\partial \ell}(s, \ell) = \frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = 0$ , where  $s \in [0, s_0] \cup [s_1, 1]$  and  $\ell^2 < \frac{2}{1-s}$ . From  $\frac{\partial F_1}{\partial \ell}(s, \ell) = 0$ , we get:

$$F_1(s, \ell) = -\frac{\ell^3}{2 - \ell^2(1-s)} \left( -\sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + \ell \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) \right).$$

From  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = 0$ , which implies  $\chi(s, \ell) = 0$  (since  $\frac{\partial F_1}{\partial \ell}(s, \ell) = 0$ ), we get:

$$\Phi\left(-\frac{\ell}{2\sqrt{s}}\right) = \frac{1}{\ell} \frac{\ell^2(1-s) - 6}{\ell^2(1-s) - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right). \quad (\text{B.8})$$

By combining these two relations, we have

$$\begin{aligned} F_1(s, \ell) &= -\frac{\ell^3}{2 - \ell^2(1-s)} \left( -\sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + \frac{\ell^2(1-s) - 6}{\ell^2(1-s) - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right) \right) \\ &= -\frac{\ell^3}{\ell^2(1-s) - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right). \end{aligned}$$

Finally, using the expression for  $F_1(s, \ell)$ , we get:

$$\Phi\left(-\frac{\ell}{2\sqrt{s}}\right) + (1 - 2s) \exp\left(\frac{\ell^2(s - 1)}{2}\right) \Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) = \frac{\ell(1 - s)}{4 - \ell^2(1 - s)} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right).$$

Using again (B.8), this yields

$$\begin{aligned} & \frac{1}{\ell} \frac{\ell^2(1 - s) - 6}{\ell^2(1 - s) - 4} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + (1 - 2s) \exp\left(\frac{\ell^2(s - 1)}{2}\right) \Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) \\ &= \frac{\ell(1 - s)}{4 - \ell^2(1 - s)} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right), \end{aligned}$$

which implies

$$(1 - 2s) \Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) = \frac{2}{\ell} \frac{\ell^2(1 - s) - 3}{4 - \ell^2(1 - s)} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}(1 - 2s)^2\right). \tag{B.9}$$

We notice that the right-hand side is negative, so that this equation has no solution if  $1 - 2s > 0$ , which leads to a contradiction with (B.7) in the case  $s \in [0, s_0]$ . In conclusion, in the case  $s \in (0, s_0)$ ,  $\ell \mapsto F_1(s, \ell)$  admits only one local maximum at point  $\ell^*(s)$ , which is also a global maximum.

*B.2.2. The case  $s \in (s_1, 1)$*

In the case  $s \in (s_1, 1)$ , we need another argument.

**Lemma B.1.** *Let us consider  $s \in (s_1, 1)$  and  $\ell \in [0, \ell_0(s)]$  such that  $\frac{\partial F_1}{\partial \ell}(s, \ell) = \frac{\partial^2 F_1}{\partial \ell^2}(s, \ell) = 0$ . Then,  $\ell < \ell_-(s)$ .*

**Proof.** We know from the previous computations that  $(s, \ell)$  satisfies (B.9). Using the lower bound in the classical inequality (38), we get

$$\Phi\left(\frac{\ell}{2\sqrt{s}} - \ell\sqrt{s}\right) = \Phi\left(-\frac{\ell(2s - 1)}{2\sqrt{s}}\right) > \frac{\ell(2s - 1)/(2\sqrt{s})}{1 + \ell^2(2s - 1)^2/(4s)} \exp\left(-\frac{\ell^2(2s - 1)^2}{8s}\right) \frac{1}{\sqrt{2\pi}}.$$

From (B.9), we thus obtain (since  $1 - 2s < 0$ )

$$\begin{aligned} & \frac{2}{\ell} \frac{\ell^2(1 - s) - 3}{4 - \ell^2(1 - s)} \sqrt{\frac{s}{2\pi}} \exp\left(-\frac{\ell^2}{8s}(1 - 2s)^2\right) \\ & < (1 - 2s) \frac{\ell(2s - 1)/(2\sqrt{s})}{1 + \ell^2(2s - 1)^2/(4s)} \exp\left(-\frac{\ell^2(2s - 1)^2}{8s}\right) \frac{1}{\sqrt{2\pi}} \end{aligned}$$

which implies

$$\frac{\ell^2(1 - s) - 3}{4 - \ell^2(1 - s)} < -\frac{\ell^2(2s - 1)^2}{4s + \ell^2(2s - 1)^2}$$

and then (since  $\ell^2(1 - s) < 2$ )

$$(\ell^2(1 - s) - 3)(4s + \ell^2(2s - 1)^2) < -\ell^2(2s - 1)^2(4 - \ell^2(1 - s)).$$

This implies that

$$\ell^2 < 12s.$$

On the other hand, it is easy to check that

$$(\ell_-)^2 > 12s.$$

Indeed

$$\begin{aligned} (\ell_-)^2 > 12s &\iff 2\left(\frac{1}{1-s} + s\right) - \frac{2}{|1-s|}(s^2(1-s)^2 - 10s(1-s) + 1)^{1/2} > 12s \\ &\iff 1 - 5s(1-s) > (s^2(1-s)^2 - 10s(1-s) + 1)^{1/2} \\ &\iff 1 - 10s(1-s) + 25s^2(1-s)^2 > s^2(1-s)^2 - 10s(1-s) + 1 \end{aligned}$$

which is obviously true. Thus, (B.9) implies  $\ell < \ell_-$ . □

Let us now assume the existence of  $\underline{s} \in (s_1, 1)$  such that  $\ell \mapsto F_1(\underline{s}, \ell)$  admits two local maxima  $\ell_1^*(\underline{s}) < \ell_2^*(\underline{s})$ . We recall that necessarily,  $\chi(\underline{s}, \ell_-(\underline{s})) \geq 0$  and  $\ell_1^*(\underline{s}) < \ell_-(\underline{s}) < \ell_2^*(\underline{s})$ . Lemma B.3 below shows that  $\frac{d}{ds}\chi(s, \ell_-(s)) > 0$  for  $s \in (s_1, 1)$ . This implies that  $\forall s \in (\underline{s}, 1]$ ,  $\chi(s, \ell_-(s)) > 0$ . Using the implicit function theorem, we can construct, a continuous curve  $\ell_2^*(s)$  on a maximum interval of the form  $s \in [\underline{s}, \bar{s})$  with  $\bar{s} > \underline{s}$  such that for  $s \in [\underline{s}, \bar{s})$ ,  $\frac{\partial F_1}{\partial \ell}(s, \ell_2^*(s)) = 0$ ,  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell_2^*(s)) < 0$  and thus  $\chi(s, \ell_2^*(s)) < 0$ . Due to the respective signs of the continuous function  $\chi(s, \ell)$  on the two continuous curves  $s \mapsto \ell_2^*(s)$  and  $s \mapsto \ell_-(s)$ , these curves cannot intersect on  $[\underline{s}, \min(\bar{s}, 1))$ . Therefore,  $\forall s \in [\underline{s}, \min(\bar{s}, 1))$ ,  $\ell_2^*(s) > \ell_-(s)$ . We now distinguish between three cases.

If  $\bar{s} > 1$ , then  $\ell_2^*(1) \geq \ell_-(1) = \sqrt{12}$  whereas  $\frac{\partial F_1}{\partial \ell}(1, \ell_2^*(1)) = 0$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(1, \ell_2^*(1)) < 0$  so that we contradict (B.2).

If  $\bar{s} < 1$ , then since  $\ell_2^*(s) < \ell_0(s) = \sqrt{\frac{2}{1-s}}$ , we may find an increasing sequence  $(s_n)_{n \in \mathbb{N}}$  of elements of  $[\underline{s}, \bar{s})$  converging to  $\bar{s}$  and such that  $\ell_2^*(s_n)$  converges to some limit denoted by  $\ell_2^*(\bar{s})$  and which belongs to  $[\ell_-(\bar{s}), \ell_0(\bar{s})]$ . By continuity of  $\frac{\partial F_1}{\partial \ell}(s, \ell)$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell)$ , one has  $\frac{\partial F_1}{\partial \ell}(\bar{s}, \ell_2^*(\bar{s})) = 0$ ,  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}, \ell_2^*(\bar{s})) \leq 0$  and thus  $\chi(\bar{s}, \ell_2^*(\bar{s})) \leq 0$ . This implies that  $\ell_2^*(\bar{s}) > \ell_-(\bar{s})$  since  $\chi(\bar{s}, \ell_-(\bar{s})) > 0$ . In turn, this implies, by Lemma B.1, that  $\frac{\partial^2 F_1}{\partial \ell^2}(\bar{s}, \ell_2^*(\bar{s})) < 0$ . Combining the implicit function theorem with the uniqueness of local maxima of  $\ell \mapsto F_1(s, \ell)$  for  $\ell \geq \ell_-(s)$ , we contradict the maximality of  $\bar{s}$ .

Let us finally consider the case  $\bar{s} = 1$ . We are going to check that  $\frac{\partial F_1}{\partial \ell}(s, \ell)$  is negative for  $\ell$  large uniformly in  $s \in (\frac{1}{2}, 1)$  (see Lemma B.2) so that  $\ell_2^*(s)$  remains bounded in the limit  $s \rightarrow 1$ . This implies that we may find an increasing sequence  $(s_n)_{n \in \mathbb{N}}$  of elements of  $[\underline{s}, 1)$  converging to 1 and such that  $\ell_2^*(s_n)$  converges to some limit denoted by  $\ell_2^*(1) \geq \ell_-(1) = \sqrt{12}$ . By continuity

of  $\frac{\partial F_1}{\partial \ell}(s, \ell)$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(s, \ell)$ , one has  $\frac{\partial F_1}{\partial \ell}(1, \ell_2^*(1)) = 0$  and  $\frac{\partial^2 F_1}{\partial \ell^2}(1, \ell_2^*(1)) \leq 0$  but this contradicts (B.2), and concludes the proof of Lemma 2.

**Lemma B.2.** *There exists  $L > 0$  and  $\alpha < 0$  such that, for all  $\ell > L$  and for all  $s \in (s_1, 1)$ ,  $\frac{\partial F_1}{\partial \ell}(s, \ell) < \alpha$ .*

**Proof.** Let  $s \in (s_1, 1)$ . By (B.4) and nonnegativity of  $F_1$ , one has

$$\begin{aligned} \frac{\partial F_1}{\partial \ell}(s, \ell) &\leq \frac{2}{\ell} F_1(s, \ell) + \ell^2 \left( -\sqrt{\frac{2s}{\pi}} \exp\left(-\frac{\ell^2}{8s}\right) + \ell \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) \right) \\ &= \frac{2\ell}{1-s} \left( 1 + \frac{\ell^2}{2}(1-s) + (1-2s) \exp\left(\frac{\ell^2(s-1)}{2}\right) \right) \Phi\left(-\frac{\ell}{2\sqrt{s}}\right) \\ &\quad - \frac{2\ell(2s-1)}{1-s} \exp\left(\frac{\ell^2(s-1)}{2}\right) \int_{\ell(2s-1)/(2\sqrt{s})}^{\ell/(2\sqrt{s})} \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \\ &\quad - \frac{2\ell^2\sqrt{s}}{\sqrt{2\pi}} \exp\left(-\frac{\ell^2}{8s}\right). \end{aligned}$$

Using two integrations by parts, one obtains

$$\begin{aligned} \int_{\ell(2s-1)/(2\sqrt{s})}^{\ell/(2\sqrt{s})} e^{-x^2/2} dx &\geq \left( \frac{2\sqrt{s}}{\ell(2s-1)} - \frac{8s^{3/2}}{\ell^3(2s-1)^3} \right) \exp\left(-\frac{\ell^2(2s-1)^2}{8s}\right) \\ &\quad - \left( \frac{2\sqrt{s}}{\ell} - \frac{8s^{3/2}}{\ell^3} \right) \exp\left(-\frac{\ell^2}{8s}\right) \end{aligned}$$

and

$$\Phi\left(-\frac{\ell}{2\sqrt{s}}\right) = \frac{1}{\sqrt{2\pi}} \left( \frac{2\sqrt{s}}{\ell} - \frac{8s^{3/2}}{\ell^3} + \mathcal{O}\left(\frac{1}{\ell^5}\right) \right) \exp\left(-\frac{\ell^2}{8s}\right),$$

with the term  $\mathcal{O}\left(\frac{1}{\ell^5}\right)$  uniform in  $s \in (s_1, 1)$ . Using the fact that

$$\begin{aligned} \frac{2\ell}{1-s} \left( 1 + \frac{\ell^2}{2}(1-s) + (1-2s) \exp\left(\frac{\ell^2(s-1)}{2}\right) \right) \\ = \ell^3 + 4\ell \exp\left(\frac{\ell^2(s-1)}{2}\right) + \frac{2\ell}{1-s} \left( 1 - \exp\left(\frac{\ell^2(s-1)}{2}\right) \right) \end{aligned}$$

we get, since  $s < 1$ ,

$$0 \leq \frac{2\ell}{1-s} \left( 1 + \frac{\ell^2}{2}(1-s) + (1-2s) \exp\left(\frac{\ell^2(s-1)}{2}\right) \right) \leq 2\ell^3 + 4\ell.$$

Thus, we get

$$\begin{aligned}
 & \sqrt{2\pi} \exp\left(\frac{\ell^2}{8s}\right) \frac{\partial F_1}{\partial \ell}(s, \ell) \\
 & \leq \left( \ell^3 + 4\ell \exp\left(\frac{\ell^2(s-1)}{2}\right) + \frac{2\ell}{1-s} \left(1 - \exp\left(\frac{\ell^2(s-1)}{2}\right)\right) \right) \\
 & \quad \times \left( \frac{2\sqrt{s}}{\ell} - \frac{8s^{3/2}}{\ell^3} + \mathcal{O}\left(\frac{1}{\ell^5}\right) \right) \\
 & \quad - \frac{2\ell(2s-1)}{1-s} \left( \frac{2\sqrt{s}}{\ell(2s-1)} - \frac{8s^{3/2}}{\ell^3(2s-1)^3} \right) \\
 & \quad + \frac{2\ell(2s-1)}{1-s} \exp\left(\frac{\ell^2(s-1)}{2}\right) \left( \frac{2\sqrt{s}}{\ell} - \frac{8s^{3/2}}{\ell^3} \right) - 2\ell^2\sqrt{s} \\
 & = -8s^{3/2} + 8\sqrt{s} \exp\left(\frac{\ell^2(s-1)}{2}\right) + \frac{4\sqrt{s}}{1-s} \left(1 - \exp\left(\frac{\ell^2(s-1)}{2}\right)\right) \\
 & \quad - \frac{16s^{3/2}}{\ell^2(1-s)} \left(1 - \exp\left(\frac{\ell^2(s-1)}{2}\right)\right) \\
 & \quad - \frac{4\sqrt{s}}{1-s} + \frac{16s^{3/2}}{\ell^2(2s-1)^2(1-s)} + \exp\left(\frac{\ell^2(s-1)}{2}\right) \frac{4\sqrt{s}(2s-1)}{1-s} \\
 & \quad - \exp\left(\frac{\ell^2(s-1)}{2}\right) \frac{16(2s-1)s^{3/2}}{\ell^2(1-s)} + \mathcal{O}\left(\frac{1}{\ell^2}\right) \\
 & = -8s^{3/2} + \frac{16s^{3/2}2(1-s)}{\ell^2(1-s)} \exp\left(\frac{\ell^2(s-1)}{2}\right) + \frac{16s^{3/2}4s(1-s)}{\ell^2(2s-1)^2(1-s)} + \mathcal{O}\left(\frac{1}{\ell^2}\right).
 \end{aligned}$$

Therefore, one concludes that

$$\sqrt{2\pi} \exp\left(\frac{\ell^2}{8s}\right) \frac{\partial F_1}{\partial \ell}(s, \ell) \leq -8s^{3/2} + \mathcal{O}\left(\frac{1}{\ell^2}\right),$$

which indeed shows that  $\frac{\partial F_1}{\partial \ell}(s, \ell)$  is negative for  $\ell$  large uniformly in  $s \in (s_1, 1)$ . □

To conclude the proof, we need to prove the following lemma which has been used above.

**Lemma B.3.** *The function  $s \mapsto \frac{d}{ds} \chi(s, \ell_-(s))$  is positive for  $s \in (s_1, 1)$ .*

**Proof.** Let us consider the derivative  $\frac{d}{ds} \chi(s, \ell_-(s))$ . Using the fact that  $\frac{\partial \chi}{\partial \ell}(s, \ell_-(s)) = 0$ , we obtain that

$$\frac{d}{ds} \chi(s, \ell_-(s)) = \frac{\partial \chi}{\partial s}(s, \ell_-(s)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\ell_-^2}{8s}\right) \frac{1}{(\ell_-^2 - s\ell_-^2 - 4)^2 s^{3/2} \ell_-} \xi(s),$$

where

$$\xi(s) = \frac{\ell_-^2(\ell_-^2(1-s)-4)^2}{4} + 2\ell_-^2s^2 - \left(\frac{s}{2} + \frac{\ell_-^2}{8}\right)(\ell_-^2(1-s)-6)(\ell_-^2(1-s)-4),$$

where, here and in the following,  $\ell_-$  should be understood as  $\ell_-(s)$ . Notice that  $\frac{d}{ds}\chi(s, \ell_-(s))$  has the same sign as  $\xi(s)$ . By simple computations, we get:

$$\begin{aligned} \xi(s) &= \frac{\ell_-^6(1-s)^2 - 8\ell_-^4(1-s) + 16\ell_-^2}{4} + 2\ell_-^2s^2 \\ &\quad - \left(\frac{s}{2} + \frac{\ell_-^2}{8}\right)(\ell_-^4(1-s)^2 - 10\ell_-^2(1-s) + 24) \\ &= \frac{\ell_-^6(1-s)^2}{8} - \frac{3}{4}\ell_-^4(1-s) + \ell_-^2 + 2\ell_-^2s^2 - \frac{s}{2}\ell_-^4(1-s)^2 + 5s\ell_-^2(1-s) - 12s. \end{aligned}$$

By using the fact that  $Q(s, \ell_-^2) = 0$ , namely  $\ell_-^4 = 4(\frac{1}{1-s} + s)\ell_-^2 - \frac{48s}{1-s}$  to rewrite the term proportional to  $\ell_-^6$ , we obtain

$$\xi(s) = -s\ell_-^2(1-s) - \frac{1}{4}\ell_-^4(1-s) + \ell_-^2 + 2\ell_-^2s^2 - 12s$$

so that, using again  $\ell_-^4 = 4(\frac{1}{1-s} + s)\ell_-^2 - \frac{48s}{1-s}$  to rewrite the term proportional to  $\ell_-^4$ ,

$$\xi(s) = 2s\ell_-^2(2s - 1)$$

which is positive for  $s \in (s_1, 1)$ . This concludes the proof. □

## Acknowledgements

This work is supported by the French National Research Agency under the grants ANR-08-BLAN-0218 (BigMC), ANR-09-BLAN-0216-01 (MEGAS) and ANR-12-BLAN-Stab.

## References

- [1] Andrieu, C. and Robert, C. (2001). Controlled MCMC for optimal sampling. Working Papers 2001-33, Centre de Recherche en Economie et Statistique. Available at <http://ideas.repec.org/p/crs/wpaper/2001-33.html>.
- [2] Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C. and Scheffer, G. (2000). *Sur les Inégalités de Sobolev Logarithmiques. Panoramas et Synthèses [Panoramas and Syntheses]* **10**. Paris: Société Mathématique de France. With a preface by Dominique Bakry and Michel Ledoux. [MR1845806](#)
- [3] Atchadé, Y.F. and Rosenthal, J.S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* **11** 815–828. [MR2172842](#)



- [4] Bédard, M. (2007). Weak convergence of Metropolis algorithms for non-i.i.d. target distributions. *Ann. Appl. Probab.* **17** 1222–1244. [MR2344305](#)
- [5] Bédard, M. (2008). Optimal acceptance rates for Metropolis algorithms: Moving beyond 0.234. *Stochastic Process. Appl.* **118** 2198–2222. [MR2474348](#)
- [6] Bédard, M., Douc, R. and Moulines, E. (2014). Scaling analysis of delayed rejection MCMC methods. *Methodol. Comput. Appl. Probab.* To appear. Published online: 6 March 2013.
- [7] Bédard, M., Douc, R. and Moulines, E. (2012). Scaling analysis of multiple-try MCMC methods. *Stochastic Process. Appl.* **122** 758–786. [MR2891436](#)
- [8] Beskos, A., Pillai, N., Roberts, G., Sanz-Serna, J.M. and Stuart, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. *Bernoulli* **19** 1501–1534. [MR3129023](#)
- [9] Beskos, A., Roberts, G. and Stuart, A. (2009). Optimal scalings for local Metropolis–Hastings chains on nonproduct targets in high dimensions. *Ann. Appl. Probab.* **19** 863–898. [MR2537193](#)
- [10] Breyer, L.A., Piccioni, M. and Scarlatti, S. (2004). Optimal scaling of MaLa for nonlinear regression. *Ann. Appl. Probab.* **14** 1479–1505. [MR2071431](#)
- [11] Breyer, L.A. and Roberts, G.O. (2000). From Metropolis to diffusions: Gibbs states and optimal scaling. *Stochastic Process. Appl.* **90** 181–206. [MR1794535](#)
- [12] Christensen, O.F., Roberts, G.O. and Rosenthal, J.S. (2005). Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 253–268. [MR2137324](#)
- [13] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.
- [14] Jourdain, B., Lelièvre, T. and Miasojedow, B. (2012). Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit. Preprint. Available at <http://fr.arxiv.org/abs/1210.7639>.
- [15] Mattingly, J.C., Pillai, N.S. and Stuart, A.M. (2012). Diffusion limits of the random walk Metropolis algorithm in high dimensions. *Ann. Appl. Probab.* **22** 881–930. [MR2977981](#)
- [16] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.
- [17] Neal, P. and Roberts, G. (2011). Optimal scaling of random walk Metropolis algorithms with non-Gaussian proposals. *Methodol. Comput. Appl. Probab.* **13** 583–601. [MR2822397](#)
- [18] Neal, P., Roberts, G. and Yuen, W.K. (2012). Optimal scaling of random walk Metropolis algorithms with discontinuous target densities. *Ann. Appl. Probab.* **22** 1880–1927. [MR3025684](#)
- [19] Otto, F. and Villani, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. [MR1760620](#)
- [20] Pillai, N., Stuart, A. and Thiéry, A. (2011). Optimal proposal design for random walk type Metropolis algorithms with Gaussian random field priors. Preprint. Available at <http://arxiv.org/abs/1108.1494>.
- [21] Pillai, N.S., Stuart, A.M. and Thiéry, A.H. (2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Probab.* **22** 2320–2356. [MR3024970](#)
- [22] Roberts, G.O., Gelman, A. and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.* **7** 110–120. [MR1428751](#)
- [23] Roberts, G.O. and Rosenthal, J.S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60** 255–268. [MR1625691](#)
- [24] Roberts, G.O. and Rosenthal, J.S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statist. Sci.* **16** 351–367. [MR1888450](#)
- [25] Sznitman, A.S. (1991). Topics in propagation of chaos. In *École D’Été de Probabilités de Saint-Flour XIX—1989. Lecture Notes in Math.* **1464** 165–251. Berlin: Springer. [MR1108185](#)

Received January 2013