



**HAL**  
open science

## Fast Langevin based algorithm for MCMC in high dimensions

Alain Durmus, Roberts O. Gareth, Gilles Vilmart, Konstantinos Zygalakis

### ► To cite this version:

Alain Durmus, Roberts O. Gareth, Gilles Vilmart, Konstantinos Zygalakis. Fast Langevin based algorithm for MCMC in high dimensions. 2015. <hal-01174226>

**HAL Id: hal-01174226**

**<https://hal.science/hal-01174226v1>**

Preprint submitted on 8 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Fast Langevin based algorithm for MCMC in high dimensions

Alain Durmus<sup>1</sup>, Gareth O. Roberts<sup>2</sup>, Gilles Vilmart<sup>3</sup>, and Konstantinos C. Zygalakis<sup>4</sup>

July 8, 2015

## Abstract

We introduce new Gaussian proposals to improve the efficiency of the standard Hastings-Metropolis algorithm in Markov chain Monte Carlo (MCMC) methods, used for the sampling from a target distribution in large dimension  $d$ . The improved complexity is  $\mathcal{O}(d^{1/5})$  compared to the complexity  $\mathcal{O}(d^{1/3})$  of the standard approach. We prove an asymptotic diffusion limit theorem and show that the relative efficiency of the algorithm can be characterised by its overall acceptance rate (with asymptotical value 0.704), independently of the target distribution. Numerical experiments confirm our theoretical findings.

*Keywords:* Hasting-Metropolis algorithm, Markov Chain Monte-Carlo method, modified equations.

*AMS subject classification (2010):* 65C40, 65C05, 60H10

## 1 Introduction

Consider a probability measure  $\pi$  on  $\mathbb{R}^d$  with density again denoted  $\pi$  with respect to the Lebesgue measure, and the Langevin diffusion  $\{x_t, t \geq 0\}$  associated with, solution of the following stochastic differential equation in  $\mathbb{R}^d$ :

$$dx_t = (1/2)\Sigma\nabla \log \pi(x_t)dt + \Sigma^{1/2}dW_t, \quad (1)$$

where  $\{W_t, t \geq 0\}$  is a standard  $d$ -dimensional Brownian motion, and  $\Sigma$  is a given positive definite self-adjoint matrix. Under appropriate assumptions [9] on  $\pi$  one can show that the dynamic generated by (1) is ergodic and has  $\pi$  as unique invariant distribution. This is a key property of (1) and taking advantage of it permits to sample from the invariant distribution  $\pi$ . In particular, if one could solve (1) analytically and then take time  $t$  to infinity then it would be possible to generate samples from  $\pi$ . However, there exist a limited number of cases [11] where such an analytical formula exists. A standard approach is to discretize (1) using a one step integrator, and sample using the averages over the numerical trajectories. The drawback of this approach is that it introduces a bias because the numerical invariant measure does not coincide in general with the exact one  $\pi$  [24, 13, 1]. In addition, the discretisation might fail to be ergodic [22], even though (1) is geometrically ergodic, but numerical ergodicity is however is not needed in practice [16].

---

<sup>1</sup>LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

<sup>2</sup>Dept of Statistics University of Warwick, Coventry, CV4 7AL, UK. Gareth.O.Roberts@warwick.ac.uk

<sup>3</sup>Université de Genève, Section de mathématiques, 2-4 rue du Lièvre, CP 64, 1211 Genève 4, Switzerland. Gilles.Vilmart@unige.ch

<sup>4</sup>Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, UK. k.zygalakis@soton.ac.uk

An alternative way of sampling from  $\pi$  exactly, *i.e.* that does not face the bias issue introduced by discretizing (1), is by using the Metropolis-Hastings algorithm [10]. The idea is to construct a Markov chain  $\{x_j, j \in \mathbb{N}\}$ , where at each step  $j \in \mathbb{N}$ , given  $x_j$ , a new sample proposal  $y$  is generated from the Markov chain with transition kernel  $q(x_j, \cdot)$ . This proposal is then accepted ( $x_{j+1} = y$ ) with probability  $\alpha(x_j, y)$  and rejected ( $x_{j+1} = x_j$ ) otherwise. Setting

$$\alpha(x, y) = \min \left( 1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right), \quad (2)$$

then the resulting Markov chain for the sequence  $\{x_j, j \in \mathbb{N}\}$  is  $\pi$ -invariant and will, for large  $j$  generate samples from  $\pi$  under mild ergodic hypothesis [12, 18].

One crucial component for the efficiency of this algorithm is the value of the accept-reject probability  $\alpha(x, y)$ , which depends on the choice of the transition kernel  $q(x, y)$ . The simplest proposals are random walks for which  $q(x, y)$  is the transition kernel associated with the proposal

$$y = x + \sqrt{h}\Sigma^{1/2}\xi, \quad (3)$$

where  $\xi$  is a standard Gaussian random variable in  $\mathbb{R}^d$ , and leads to the well known Random Walk Metropolis Algorithm (RMW). This proposal is very simple to implement, but it suffers from (relatively) high rejection rate, due to the fact that it utilises no information about  $\pi$  to construct appropriate candidate moves.

Another family of proposal commonly used is based on the Euler-Maruyama discretisation of (1), for which  $q(x, y)$  is the transition kernel associated with the proposal

$$y = x + (h/2)\Sigma\nabla \log \pi(x) + \sqrt{h}\Sigma^{1/2}\xi, \quad (4)$$

where  $\xi$  is again a standard Gaussian random variable in  $\mathbb{R}^d$ . This algorithm is also known as the Metropolis Adjusted Langevin Algorithm (MALA), and it is well-established that it has an improved complexity scaling and a better convergence behaviour than the RWM algorithm in general. This method directs the proposed moves towards areas of high probability for the distribution  $\pi$  thanks to the presence of the  $\Sigma\nabla \log \pi(x)$  term. There is now a growing literature on gradient-based MCMC algorithms, as exemplified through the two papers [7, 4] and the references therein. A natural question is if one can improve on the behaviour of MALA by incorporating more information about the properties of  $\pi$  in their proposal. A first attempt would be to use as proposal a one-step integrator with high weak order for (1), as suggested in the discussion of [7]. Although this appears not sufficient, we shall show that, by slightly modifying this approach and not focusing the the weak order itself, we are able to construct a new proposal with better convergence and complexity scaling properties than MALA. We mention that an analogous proposal is presented independently in [6] in a different context to improve the strong order of convergence of MALA.

The paper is organised as follows. In Section 2 we provide a heuristic for the choice of time-step  $h$  used in the proposal as a function of the dimension  $d$  of the target and present three different proposals that have better complexity scaling properties than RWM and MALA. In Section 3, we present and prove our main scaling results for the introduced methods. Section 4 investigates the ergodic properties of the different proposals for a wide variety of target densities  $\pi$ . Finally, in Section 5 we present numerical results that illustrate our theoretical findings.

## 2 Preliminaries

In this section we discuss some key issues regarding the complexity scaling of MCMC algorithms in high dimensions. In particular, in Section 2.1 we discuss some issues relating with the computational complexity of MCMC methods in high dimensions, while in Section 2.2 we present a useful heuristic for understanding the scaling properties of a given MCMC proposal, and based on this heuristic formally derive a new proposal with desirable scaling properties.

### 2.1 Computational Complexity

Here we discuss a heuristic approach for selecting the time-step  $h$  in all proposals mentioned above, in order to optimise the acceptance probability as the dimension of the space  $d \rightarrow \infty$ . In particular, we choose the time-step  $h$  proportional to an inverse power of the dimension  $d$  such that

$$h \propto d^{-\gamma} . \quad (5)$$

This implies that the proposal  $y$  is now a function of: (i) the current state  $x$ ; (ii) the parameter  $\gamma$  through the time-step scaling above; and (iii) the independent noises  $\xi$  which appear in all the considered proposals. Thus  $y = y(x, \xi; \gamma)$ . Ideally one wants  $\gamma$  to be as small as possible so the chain makes large steps and decorrelation amongst samples will be maximised. At the same time one needs to ensure that the acceptance probability doesn't degenerate to 0 as  $d \rightarrow \infty$ , also to prevent high correlation amongst samples. This naturally leads to the definition of a critical exponent  $\gamma_0$  given by

$$\gamma_0 = \inf_{\gamma_c \geq 0} \left\{ \gamma_c : \liminf_{d \rightarrow \infty} \mathbb{E} [\alpha(x, y)] > 0 , \quad \forall \gamma \in [\gamma_c, \infty) \right\} .$$

The expectation here is with respect to  $x$  distributed according to  $\pi$  and  $y$  chosen from the proposal distribution. In other words, we take the largest possible time-steps, as functions of  $d$ , constrained by asking that the average acceptance probability is bounded away from zero, uniformly in  $d$ . The time-step restriction (5) can be interpreted as a kind of Courant-Friedrichs-Lewy restriction arising in the numerical time-integration of PDEs.

Continuing this analogy, we introduce here the heuristic that the number of steps required to reach stationarity is given by

$$M(d) = d^{\gamma_0} . \quad (6)$$

As we will discuss below, this heuristic can be given by a firm foundation in a number of cases. One intuitive explanation of why this holds is that the Markov chain arising from the Metropolis-Hastings method approximates a Langevin SDE and reaches stationarity exponentially fast with a time interval of size  $\mathcal{O}(1)$ . Using the fact that the Markov chain travels with time-step  $h$  on the path of this SDE, we obtain the expression for  $M(d)$  above. If we now consider the case of a product measure where

$$\pi(x) = \pi_d(x) = Z_d \prod_{i=1}^d e^{g(x_i)} , \quad (7)$$

and  $Z_d$  is a renormalisation constant, then it is well known [19] that for the RWM  $\gamma_0 = 1$ , while for MALA  $\gamma_0 = 1/3$  [20]. In the next subsection, we recall the main ideas that allows one to obtain these scalings (valid also for some non-product cases), and derive a new proposal

which we will call the fast Metropolis Adjusted Langevin algorithm (fMALA) and which satisfies  $\gamma_0 = 1/5$  in the product case, i.e. it has an improved complexity scaling.

## 2.2 Formal derivation

Here we explain the main idea that can be used for proving the scaling of a Gaussian<sup>1</sup> proposal in high dimensions. In particular, the proposal  $y$  is now of the form

$$y = \mu(x, h) + S(x, h)\xi, \quad (8)$$

where  $\xi \sim \mathcal{N}(0, I_d)$  is a standard  $d$  dimensional Gaussian random variable in  $\mathbb{R}^d$ . Note that in the case of RWM,

$$\mu(x, h) = x, \quad S(x, h) = \sqrt{h}\Sigma^{1/2},$$

while in the case of MALA

$$\mu(x, h) = x + (h/2)\Sigma\nabla \log \pi(x), \quad S(x, h) = \sqrt{h}\Sigma^{1/2}.$$

The Gaussianity of the proposal implies that the accept-reject probability  $\alpha(x, y)$  can be written in the form

$$\alpha(x, y) = \min\{1, \exp(R_d(x, y))\},$$

for some function  $R_d(x, y)$  which depends on the Gaussian proposal (8). Now using the fact that  $y$  is related to  $x$  according to (8), and that  $R_d(x, x) = 0$  together with appropriate smoothness properties on the function  $g(x)$ , one can expand  $R_d$  in powers of  $\sqrt{h}$  using a Taylor expansion

$$R_d(x, y) = \sum_{i=1}^k \sum_{j=1}^d h^{i/2} C_{ij}(x, \xi) + h^{(k+1)/2} R_{k+1}(x, h^*, \xi). \quad (9)$$

It turns out [2] that the scaling associated with each proposal relates directly with how many of the  $C_{ij}$  terms are zero in (9). This simplifies if we further assume that  $\Sigma = I_d$  in (1) and that  $\pi$  satisfies (7), because one has

$$C_{ij}(x, \xi) = C_i(x_j, \xi_j),$$

and thus each term in (9) can be written as

$$\frac{\sqrt{h^i d}}{\sqrt{d}} \sum_{j=1}^d C_i(x_j, \xi_j).$$

We then see that if one has  $C_i = 0, i = 1, \dots, m$  then this implies that  $\gamma_0 = 1/(m+1)$ . Indeed, this value of  $\gamma_0$  yields  $h^{m+1}d = 1$  and the leading order term in (9) becomes

$$\frac{1}{\sqrt{d}} \sum_{j=1}^d C_{m+1}(x_j, \xi_j).$$

To understand the behaviour for large  $d$ , we typically assume conditions to ensure that the above term has an appropriate (weak) limit. It turns out that  $m+1$  is generally an odd integer

---

<sup>1</sup>We stress that Gaussianity here is not necessary but it greatly simplifies the calculations.

for known proposals, and the above expression is frequently approximated by a central limit theorem. The second dominant term in (9) turns out to be  $C_{2(m+1)}$ , although to turn this into a rigorous proof one also needs to be able to control the appropriate number of higher order terms, from  $m + 1$  to  $2(m + 1)$ , as well as the remainder term in the above Taylor expansion.

### 2.3 Classes of proposals with $\gamma_0 = 1/5$

We introduce new Gaussian proposals for which  $\gamma_0 = 1/5$  in (6). We start by presenting the simplest method, and then present two variations of it, motivated by the desire to obtain robust and stable ergodic properties (geometric ergodicity). The underlying calculations that show  $C_i = 0, i = 1, \dots, m$  with  $m = 4$  and  $\gamma_0 = 1/5$  for these methods are contained in the supplementary materials in the form of a Mathematica file. Recall that  $f(x) = \Sigma \nabla \log \pi(x)$ . In the sequel, we denote by  $Df$  and  $D^2f$  the Jacobian ( $d \times d$ -matrix) and the Hessian ( $d \times d^2$ -matrix) of  $f$  respectively. Thus  $(Df(x))_{i,j} = \frac{\partial f_i(x)}{\partial x_j}$  and

$$D^2f(x) = [\mathbf{H}_1(x) \quad \dots \quad \mathbf{H}_d(x)] \quad , \quad \text{where } \{\mathbf{H}_i(x)\}_{j,k} = \frac{\partial^2 f_i(x)}{\partial x_k \partial x_j}.$$

Finally for all  $x \in \mathbb{R}^d$ ,  $\{\Sigma : D^2f(x)\} \in \mathbb{R}^d$  is defined by for  $i = 1, \dots, d$ :

$$\{\Sigma : D^2f(x)\}_i = \text{trace}(\Sigma^T \mathbf{H}_i(x)) \quad .$$

Notice that for  $\Sigma = \mathbf{I}_d$ , the above quantity reduces to the Laplacian and we have  $\{\Sigma : D^2f(x)\}_i = \Delta f_i$ .

#### 2.3.1 Fast Metropolis-Adjusted Langevin Algorithm (fMALA).

We define

$$\mu^{\text{fM}}(x, h) = x + \frac{h}{2} f(x) - \frac{h^2}{24} (Df(x)f(x) + \{\Sigma : D^2f(x)\}) \quad , \quad (10a)$$

$$S^{\text{fM}}(x, h) = \left( \sqrt{h} \mathbf{I}_d + \frac{\sqrt{h^3}}{12} Df(x) \right) \Sigma^{1/2} \quad . \quad (10b)$$

We will refer to (8) when  $\mu, S$  are given by (10) as the fast Unadjusted Langevin Algorithm (fULA) when viewed as a numerical method for (1) and as the fast Metropolis-Adjusted Langevin Algorithm (fMALA) when used as a proposal in the Metropolis-Hastings framework.

**Remark 2.1.** *It is interesting to note that when compared with Unadjusted Langevin Algorithm (ULA), fULA has the same order of weak convergence one, if applied as a one-step integrator for (1). One could obtain a second order weak method by changing the constants in front of the higher order coefficients, but in fact the corresponding method would not have better scaling properties than MALA when used in the Metropolis-Hastings framework. This observation answers negatively in part one of the questions in the discussion of [7] about the potential use of higher order integrators for the Langevin equation within the Metropolis-Hastings framework.*

### 2.3.2 Modified Ozaki-Metropolis algorithm (mOMA)

One of the problems related to the MALA proposal is that it fails to be geometrically ergodic for a wide range of functions  $\pi$  [22]. This issue was addressed in [21] where a modification of MALA based on the Ozaki discretisation [17] of (1) was proposed and studied. In the same spirit as in [21] we propose here a modification of fMALA, defined by

$$\mu^{\text{mO}}(x, h) = x + \mathcal{T}_1(Df(x), h, 1)f(x) - \frac{h^2}{6}Df(x)f(x) - \frac{h^2}{24}\{\Sigma : D^2f(x)\} \quad (11a)$$

$$S^{\text{mO}}(x, h)(S^{\text{mO}}(x, h))^T = \left( \mathcal{T}_1(Df(x), 2h, 1) - \frac{h^2}{3}Df(x) \right) \Sigma. \quad (11b)$$

where

$$\mathcal{T}_1(M, h, a) = (aM)^{-1}(e^{(ah/2)M} - I_d) \quad (12)$$

for all<sup>2</sup>  $M \in \mathbb{R}^{d \times d}$ ,  $h > 0$ ,  $a \in \mathbb{R}$ .

The Markov chain defined by (11) will be referred to as the modified unadjusted Ozaki algorithm (mUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the modified Ozaki Metropolis algorithm (mOMA). Note that  $t \mapsto (e^{ht} - 1)/t - (1/3)h^2t$  is positive on  $\mathbb{R}$  for all  $h > 0$ , therefore  $S^{\text{mO}}(x, h)(S^{\text{mO}}(x, h))^T$  is positive definite for all  $x \in \mathbb{R}^d$  and  $h > 0$ .

**Remark 2.2.** *In regions where  $\|\nabla \log \pi(x)\|$  is much greater than  $\|x\|$ , we need in practice to take  $h$  very small (of order  $\|x\| / \|\nabla \log \pi(x)\|$ ) for MALA to exit these regions. However such a choice of  $h$  depends on  $x$  and cannot be used directly. Finding such a value of  $h$  can therefore be hard to find theoretically as well as computationally. This issue can be tackled by multiplying  $f = \nabla \log \pi(x)$  by  $\mathcal{T}_1(Df(x), h, a)$  in (11a). Indeed under some mild conditions, in that case, we can obtain an algorithm with good mixing properties for all  $h > 0$ ; see [21, Theorem 4.1]. mOMA faces similar problems due to the term  $Df(x)f(x)$ .*

### 2.3.3 Generalised Boosted Ozaki-Metropolis Algorithm (gbOMA).

Having discussed the possible limitations of mOMA in Remark 2.2 we generalise here the approach in [21] to deal with the complexities arising to the presence of the  $Df(x)f(x)$  term. In particular we now define

$$\mu^{\text{gbO}}(x, h) = x + \mathcal{T}_1(Df(x), h, a_1)f(x) - \frac{1}{3}\mathcal{T}_3(Df(x), h, a_3)\{\Sigma : D^2f(x)\} + \left( \frac{a_1}{2} + \frac{1}{6} \right) \mathcal{T}_2(Df(x), h, a_2)f(x), \quad (13a)$$

$$S^{\text{gbO}}(x, h)(S^{\text{gbO}}(x, h))^T = \left\{ \mathcal{T}_1(Df(x), 2h, a_4) + \left( \frac{a_4}{2} - \frac{1}{6} \right) \mathcal{T}_2(Df(x), 2h, a_5) \right\} \Sigma. \quad (13b)$$

where  $a_i$ ,  $i = 1, \dots, 5$  are positive parameters,  $\mathcal{T}_1$  is given by (12) and

$$\mathcal{T}_2(M, h, a) = (aM)^{-1}(e^{-(ah^2/4)M^2} - I_d) \quad (14)$$

$$\mathcal{T}_3(M, h, a) = (aM)^{-2}(e^{(ah/2)M} - I_d - (ah/2)M) \quad (15)$$

<sup>2</sup>Notice that the matrix functionals in (12),(14),(15) remain valid if matrix  $aM$  is not invertible, using the appropriate power series for the matrix exponentials.

with  $M \in \mathbb{R}^{d \times d}$ ,  $h > 0$ ,  $a \in \mathbb{R}$  and  $I_d$  is the identity matrix. The Markov chain defined by (13) will be referred to as the generalised boosted unadjusted Ozaki algorithm (gbUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the generalised boosted Ozaki Metropolis algorithm (gbOMA).

We notice that  $S^{\text{gbO}}(S^{\text{gbO}})^T$  in (13b) is not a positive definite matrix in general and so we impose the following condition on the constants  $a_4, a_5$  to guarantee it.

**Assumption 1.** *The function  $t \mapsto (e^{a_4 t} - 1)/(a_4 t) + (a_4/2 - (1/6))(e^{-a_5 t^2} - 1)/(a_5 t)$  is positive on  $\mathbb{R}$ .*

However for  $a_4 = a_5 = 1$ , this assumption is satisfied, and choosing  $a_i = 1$  for all  $i = 1, \dots, 5$ , (13) leads to a well defined proposal, which will be referred to as the boosted Unadjusted Ozaki Algorithm (bUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the boosted Ozaki Metropolis Algorithm (bOMA). We will see in Section 4 that bOMA has nicer ergodic properties than fMALA.

### 3 Main scaling results

In this section, we present optimal scaling results for fMALA and gbOMA introduced in Section 2. We recall from the discussion in Section 2 that the time stepsize  $h$  depends on the dimension and is given as  $h_d = \ell d^{-1/5}$ , with  $\ell > 0$ . Finally, we prove our results for the case of a target distributions of the product form given by (7), we take  $\Sigma = I_d$ , and we make the following assumptions on  $g$ .

**Assumption 2.** *We assume*

1.  $g \in C^{10}(\mathbb{R})$  and  $g''$  is bounded on  $\mathbb{R}$ .
2. The derivatives of  $g$  up to order 10 have at most a polynomial growth, i.e. there exists constants  $C, \kappa$  such that

$$|g^{(i)}(t)| \leq C(1 + |t|^\kappa), \quad t \in \mathbb{R}, i = 1, \dots, 10.$$

3. for all  $k \in \mathbb{N}$ ,

$$\int_{\mathbb{R}} t^k e^{g(t)} dt < +\infty.$$

#### 3.1 Optimal scaling of fMALA

The Markov chain produced by fMALA, with target density  $\pi_d$ , will be denoted by  $\{X_k^{d, \text{fM}}, k \in \mathbb{N}\}$ . Let  $q_d^{\text{fM}}$  be the transition density function associated to the proposal of fMALA relatively to  $\pi_d$ . In a similar manner, we denote by  $\alpha_d^{\text{fM}}$  the acceptance ratio. Now we introduce the jump process build on  $\{X_k^{d, \text{fM}}, k \in \mathbb{N}\}$ , which allows us to compare this Markov chain to a continuous-time process. Let  $\{J_t, t \in \mathbb{R}_+\}$  be a Poisson process with rate  $d^{1/5}$ , and let  $\Gamma^{d, \text{fM}} = \{\Gamma_t^{d, \text{fM}}, t \in \mathbb{R}_+\}$  be the  $d$ -dimensional jump process defined by  $\Gamma_t^{d, \text{fM}} = X_{J_t}^{d, \text{fM}}$ . We denote by

$$a_d^{\text{fM}}(\ell) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \pi_d(x) q_d^{\text{fM}}(x, y) \alpha_d^{\text{fM}}(x, y) dx dy$$

the mean under  $\pi_d$  of the acceptance ratio.

**Theorem 3.1.** *Assume Assumption 2. Then*

$$\lim_{d \rightarrow +\infty} a_d^{\text{fM}}(\ell) = a^{\text{fM}}(\ell),$$

where  $a^{\text{fM}}(\ell) = 2\Phi(-K^{\text{fM}}\ell^5/2)$  with  $\Phi(t) = (1/(2\pi)) \int_{-\infty}^t e^{-s^2/2} ds$  and the expression of  $K^{\text{fM}}$  is given in Appendix D.

**Theorem 3.2.** *Assume Assumption 2. Let  $\{Y_t^{d,\text{fM}} = \Gamma_{t,1}^{d,\text{fM}}, t \in \mathbb{R}_+\}$  be the process corresponding to the first component of  $\Gamma^{d,\text{fM}}$ . Then,  $\{Y_t^{d,\text{fM}}, d \in \mathbb{N}^*\}$  converges weakly (in the Skorokhod topology) to the solution  $\{Y_t^{\text{fM}}, t \in \mathbb{R}_+\}$  of the Langevin equation defined by:*

$$dY_t^{\text{fM}} = (h^{\text{fM}}(\ell))^{(1/2)} dB_t + (1/2)h^{\text{fM}}(\ell) \nabla \log \pi_1(Y_t^{\text{fM}}) dt, \quad (16)$$

where  $h^{\text{fM}}(\ell) = 2\ell^2\Phi(-K^{\text{fM}}\ell^5/2)$  is the speed of the limiting diffusion. Furthermore,  $h^{\text{fM}}(\ell)$  is maximised at the unique value of  $\ell$  for which  $a^{\text{fM}}(\ell) = 0.704343$ .

*Proof.* The proof of these two theorems are in Appendix A. □

### 3.2 Scaling results for gbOMA

As in the case of fMALA, we assume  $\pi_d$  is of the form (7) and we take  $\Sigma = I_d$ ,  $h_d = \ell d^{-1/5}$ . The Metropolis-adjusted Markov chain based on gbOMA, with target density  $\pi_d$ , is denoted  $\{X_k^{d,\text{gbO}}, k \in \mathbb{N}\}$ . We will denote by  $q_d^{\text{gbO}}$  the transition density associated to the proposals defined by gbOMA with respect to  $\pi_d$ . In a similar manner, the acceptance ratio relatively to  $\pi_d$  and gbOMA will be denoted by  $\alpha_d^{\text{gbO}}$ . We build, as in Section 3.1 the jump process build on  $\{X_k^{d,\text{gbO}}, k \in \mathbb{N}\}$ . Let  $\{J_t, t \in \mathbb{R}_+\}$  be a Poisson process with rate  $d^{1/5}$ , and let  $\Gamma^{d,\text{gbO}} = \{\Gamma_t^{d,\text{gbO}}, t \in \mathbb{R}_+\}$  be the  $d$ -dimensional jump process defined by  $\Gamma_t^{d,\text{gbO}} = X_{J_t}^{d,\text{gbO}}$ . We denote also

$$a_d^{\text{gbO}}(\ell) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \pi_d(x) q_d^{\text{gbO}}(x, y) \alpha_d^{\text{gbO}}(x, y) dx dy$$

the mean under  $\pi_d$  of the acceptance rate of the algorithm which produces  $\{X_k^{d,\text{gbO}}, k \in \mathbb{N}\}$ .

**Theorem 3.3.** *Assume Assumptions 1 and 2. Then*

$$\lim_{n \rightarrow +\infty} a_d^{\text{gbO}}(\ell) = a^{\text{gbO}}(\ell),$$

where  $a^{\text{gbO}}(\ell) = 2\Phi(-K^{\text{gbO}}\ell^5/2)$  with  $\Phi(t) = (1/(2\pi)) \int_{-\infty}^t e^{-s^2/2} ds$  and  $K^{\text{gbO}}$  are given in Appendix D.

**Theorem 3.4.** *Assume Assumptions 1 and 2. Let  $\{G_t^{d,\text{gbO}} = \Gamma_{t,1}^{d,\text{gbO}}, t \in \mathbb{R}_+\}$  be the process corresponding to the first component of  $\Gamma^{d,\text{gbO}}$ . Then,  $\{G_t^{d,\text{gbO}}, d \in \mathbb{N}^*\}$  converges weakly (in the Skorokhod topology) to the solution  $\{G_t^{\text{gbO}}, t \in \mathbb{R}_+\}$  of the Langevin equation defined by:*

$$dG_t^{\text{gbO}} = (h^{\text{gbO}}(\ell))^{(1/2)} dB_t + (1/2)h^{\text{gbO}}(\ell) \nabla \log \pi_c(G_t^{\text{gbO}}) dt,$$

where  $h^{\text{gbO}}(\ell) = 2\ell^2\Phi(-K^{\text{gbO}}\ell^5/2)$  is the speed of the limiting diffusion. Furthermore,  $h^{\text{gbO}}(\ell)$  is maximised at the unique value of  $\ell$  for which  $a^{\text{gbO}}(\ell) = 0.704343$ .

*Proof.* Note that under Assumption 2-1, at fixed  $a > 0$ , using the regularity properties of  $(x, h) \mapsto \mathcal{T}_i(x, h, a)$  on  $\mathbb{R}^2$  for  $i = 1, \dots, 3$ , there exists an open interval  $I$ , which contains 0, and  $M_0 \geq 0$  such that for all  $x \in \mathbb{R}$ ,  $k = 1, \dots, 11$ , and  $i = 1, \dots, 3$

$$\left| \frac{\partial^k (\mathcal{T}_i(g''(x), h, a))}{\partial h^k} \right| \leq M_0 \quad \forall h \in I .$$

And there exists  $m_0 > 0$  such that for all  $h \in I$ , all  $d \in \mathbb{N}^*$ , and for all  $x \in \mathbb{R}^d$ ,

$$\left\| S^{\text{gbO}}(x, h) \right\| > m_0$$

Using these two results, the proof of both theorems follows the same lines than Theorems 3.1 and 3.2, which can be found in Appendix A.  $\square$

## 4 Geometric ergodicity results for high order Langevin schemes

Having established the scaling behaviour of the different proposals in the previous section, we now proceed with establishing geometric ergodicity results for our new Metropolis algorithms. Furthermore, for completeness, we study the behaviour of the corresponding unadjusted proposal. For simplicity, we will take in the following  $\Sigma = I_d$  and we limit our study of gbOMA to the one of bOMA, which is given by:

$$\begin{aligned} y^{\text{bO}} &= \mu^{\text{bO}}(x, h) + S^{\text{bO}}(x, h) \xi , \\ \mu^{\text{bO}}(x, h) &= x + \mathcal{T}_1(Df(x), h, 1)f(x) + (2/3)\mathcal{T}_2(Df(x), h, 1)f(x) \\ &\quad - (1/3)\mathcal{T}_3(Df(x), h, 1)\{\Sigma : D^2f(x)\} , \\ S^{\text{bO}}(x, h)(S^{\text{bO}}(x, h))^T &= \mathcal{T}_1(Df(x), 2h, 1) + (1/3)\mathcal{T}_2(Df(x), 2h, 1) , \end{aligned} \tag{17}$$

where  $\mathcal{T}_1$ ,  $\mathcal{T}_2$  and  $\mathcal{T}_3$  are respectively defined by (12), (14) and (15). First, let us begin with some definitions. For a signed measure  $\nu$  on  $\mathbb{R}^d$ , we define the total variation norm of  $\nu$  by

$$\|\nu\|_{\text{TV}} = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A)| ,$$

where  $\mathcal{B}(\mathbb{R}^d)$  is the Borel  $\sigma$ -algebra of  $\mathbb{R}^d$ . Let  $P$  be a Markov kernel with invariant measure  $\pi$ . For a given measurable function  $V : \mathbb{R}^d \rightarrow [1, +\infty)$ , we will say that  $P$  is  $V$ -geometrically ergodic if there exist  $C \geq 0$  and  $\rho \in [0, 1)$  such that for all  $x \in \mathbb{R}^d$  and  $n \geq 0$

$$\|P^n(x, \cdot) - \pi\|_V \leq C\rho^n V(x) ,$$

where for  $\nu$  a signed measure on  $\mathbb{R}^d$ , the  $V$ -norm  $\|\cdot\|_V$  is defined by

$$\|\nu\| = \sup_{\{f : |f| \leq V\}} \int_{\mathbb{R}^d} f(x)\nu(dx) .$$

We refer the reader to [15] for the definitions of small sets,  $\varphi$ -irreducibility and transience. Let  $P$  be a Markov kernel on  $\mathbb{R}^d$ , Leb $^d$ -irreducible, where Leb $^d$  is the Lebesgue measure on  $\mathbb{R}^d$ , and aperiodic and  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  be a measurable function. In order to establish that  $P$  is  $V$ -geometric ergodicity, a sufficient and necessary condition is given by a geometrical

drift (see [15, Theorem 15.0.1]), namely for some small set  $\mathcal{C}$ , there exist  $\lambda < 1$  and  $b < +\infty$  such that for all  $x \in \mathbb{R}^d$ :

$$PV(x) \leq \lambda V(x) + b\mathbf{1}_{\mathcal{C}}(x) . \quad (18)$$

Note that the different considered proposals belong to the class of Gaussian Markov kernels. Namely, let  $Q$  be a Markov kernel on  $\mathbb{R}^d$ . We say that  $Q$  is a Gaussian Markov kernel if for all  $x \in \mathbb{R}^d$ ,  $Q(x, \cdot)$  is a Gaussian measure, with mean  $\mu(x)$  and covariance matrix  $S(x)S^T(x)$ , where  $x \mapsto \mu(x)$  and  $x \mapsto S(x)$  are measurable functions from  $\mathbb{R}^d$  to respectively  $\mathbb{R}^d$  and  $\mathcal{S}_+^*(\mathbb{R}^d)$ , the set of symmetric positive definite matrices of dimension  $d$ . These two functions will be referred to as the mean value map and the the variance map respectively. The Markov kernel  $Q$  has transition density  $q$  given by:

$$q(x, y) = \frac{1}{(2\pi)^{d/2} |S(x)|} \exp\left(-\frac{1}{2} \langle S(x)^{-2}(y - \mu(x)), (y - \mu(x)) \rangle\right) , \quad (19)$$

where for  $M \in \mathbb{R}^{d \times d}$ ,  $|M|$  denotes the determinant of  $M$ . Geometric ergodicity of Markov Chains with Gaussian Markov kernels and the corresponding Metropolis-Hastings algorithms was the subject of study of [22, 8]. But contrary to [8], we assume for simplicity the following assumption on the functions  $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $S : \mathbb{R}^d \rightarrow \mathcal{S}_+^*(\mathbb{R}^d)$ :

**Assumption 3.** *The functions  $x \mapsto \mu(x)$  and  $x \mapsto S(x)$  are continuous.*

Note that if  $\pi$ , a target probability measure on  $\mathbb{R}^d$ , is absolutely continuous with respect to the Lebesgue measure with density still denoted by  $\pi$ , the following assumption ensures that the various different proposals introduced in this paper satisfy Assumption 3:

**Assumption 4.** *The log-density  $g$  of  $\pi$  belongs to  $C^3(\mathbb{R}^d)$ .*

We proceed in Section 4.1 with presenting and extending where necessary the main results about geometric ergodicity of Metropolis-Hasting algorithms using Gaussian proposals. In Section 4.2, we then introduce two different potential classes on which we apply our result in Section 4.3. Finally in Section 4.4, for completeness, we make the same kind of study but for unadjusted Gaussian Markov kernels on  $\mathbb{R}$ .

#### 4.1 Geometric ergodicity of Hastings-Metropolis algorithm based on Gaussian Markov kernel

We first present an extension of the result given in in [8] for geometric ergodicity of Metropolis-Hastings algorithms based on Gaussian proposal kernels. In particular, let  $Q$  be a Gaussian Markov kernel with mean value map and variance map satisfying Assumption 3. We use such proposal in a Metropolis algorithm with target density  $\pi$  satisfying Assumption 4. Then, the produced Markov kernel  $P$  is given by

$$P(x, dy) = \alpha(x, y)q(x, y)dy + \delta_x(dy) \int_{\mathbb{R}^d} (1 - \alpha(x, y))q(x, y)dy , \quad (20)$$

where  $q$  and  $\alpha$  are resp. given by (19) and (2).

**Assumption 5.** *We assume  $\liminf_{\|x\| \rightarrow +\infty} \int_{\mathbb{R}^d} \alpha(x, y)q(x, y)dy > 0$ .*

Note that this condition is necessary to obtain the geometric ergodicity of a Metropolis-Hastings algorithm by [23, Theorem 5.1]. We shall follow a well-known technique in MCMC theory in demonstrating that Assumption 5 allows us to ensure that geometric ergodicity of the algorithm is inherited from that of the proposal Markov chain itself. Thus, in the following lemma we combine the conditions given by [8], which imply geometric ergodicity of Gaussian Markov kernels, with Assumption 5 to get geometric ergodicity of the resultant Metropolis-Hastings Markov kernels.

**Lemma 4.1.** *Assume Assumptions 3, 5, and there exists  $\tau \in (0, 1)$  such that*

$$\limsup_{\|x\| \rightarrow +\infty} \|\mu(x)\| / \|x\| = \tau, \text{ and } \limsup_{\|x\| \rightarrow +\infty} \|S(x)\| / \|x\| = 0. \quad (21)$$

*Then, the Markov kernel  $P$  given by (20) are  $V$ -geometrically ergodic, where  $V(x) = 1 + \|x\|^2$ .*

*Proof.* The proof is postponed to Appendix B.1.  $\square$

We now provide some conditions which imply that  $P$  is not geometrically ergodic.

**Theorem 4.2.** *Assume Assumptions 3,4, that  $\pi$  is bounded and there exists  $\epsilon > 0$  such that*

$$\liminf_{\|x\| \rightarrow +\infty} \|S(x)^{-1}\mu(x)\| \|x\|^{-1} > \epsilon^{-1}, \quad \liminf_{\|x\| \rightarrow +\infty} \inf_{\|y\|=1} \|S(x)y\| \geq \epsilon, \quad (22)$$

*and*

$$\lim_{\|x\| \rightarrow +\infty} \log(|S(x)S(x)^T|) / \|x\|^2 = 0. \quad (23)$$

*Then,  $P$  is not geometrically ergodic.*

*Proof.* The proof is postponed to Appendix B.2  $\square$

## 4.2 Exponential potentials

We illustrate our results on the following classes of density.

### 4.2.1 The one-dimensional class $\mathcal{E}(\beta, \gamma)$

Let  $\pi$  be a probability density on  $\mathbb{R}$  with respect to the Lebesgue measure. We will say that  $\pi \in \mathcal{E}(\beta, \gamma)$  if  $\pi$  is positive, belongs to  $C^3(\mathbb{R})$  and there exist  $R_\pi, \beta > 0$  such that for all  $x \in \mathbb{R}$ ,  $|x| \geq R_\pi$ ,

$$\pi(x) \propto e^{-\gamma|x|^\beta}.$$

Then for  $|x| \geq R_\pi$ ,  $\log(\pi(x))' = -\gamma\beta x|x|^{\beta-2}$ ,  $\log(\pi(x))'' = -\gamma\beta(\beta-1)|x|^\beta/x^2$  and  $\log(\pi(x))^{(3)} = -\gamma\beta(\beta-1)(\beta-2)|x|^\beta/x^3$ .

### 4.2.2 The multidimensional exponential class $\mathcal{P}_m$

Let  $\pi$  be a probability density on  $\mathbb{R}^d$  with respect to the Lebesgue measure. We will say that  $\pi \in \mathcal{P}_m$  if it is positive, belongs to  $C^3(\mathbb{R}^d)$  and there exists  $R_\pi \geq 0$  such that for all  $x \in \mathbb{R}^d$ ,  $\|x\| \geq R_\pi$ ,

$$\pi(x) \propto e^{-q(x)},$$

where  $q$  is a function of the following form. There exists a homogeneous polynomial  $p$  of degree  $m$  and a three-times continuously differentiable function  $r$  on  $\mathbb{R}^d$  satisfying

$$\|D^2(\nabla r)(x)\|_{\|x\| \rightarrow +\infty} = o(\|x\|^{m-3}), \quad (24)$$

and for all  $x \in \mathbb{R}^d$

$$q(x) = p(x) + r(x).$$

Recall that  $p$  is an homogeneous polynomial of degree  $m$  if for all  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^d$ ,  $p(tx) = t^m p(x)$ . Finally we define  $\mathcal{P}_m^+$ , the set of density  $\pi \in \mathcal{P}_m$  such that the Hessian of  $p$  at  $x$ ,  $\nabla^2 p(x)$  is positive definite for all  $x \neq 0$ .

When  $p$  is an homogeneous polynomial of degree  $m$ , it can be written as

$$p(x) = \sum_{|\mathbf{k}|=m} a_{\mathbf{k}} x^{\mathbf{k}},$$

where  $\mathbf{k} \in \mathbb{N}^d$ ,  $|\mathbf{k}| = \sum_i k_i$  and  $x^{\mathbf{k}} = x_1^{k_1} \cdots x_d^{k_d}$ . Then denoting by  $\vec{n}_x = x/\|x\|$ , it is easy to see that the following relations holds for all  $x \in \mathbb{R}^d$ .

$$p(x) = \|x\|^m p(\vec{n}_x) \quad (25)$$

$$\nabla p(x) = \|x\|^{m-1} \nabla p(\vec{n}_x) \quad (26)$$

$$\nabla^2 p(x) = \|x\|^{m-2} \nabla^2 p(\vec{n}_x) \quad (27)$$

$$D^2(\nabla p)(x) = \|x\|^{m-3} D^2(\nabla p)(\vec{n}_x) \quad (28)$$

$$\langle \nabla p(x), x \rangle = m p(x) \quad (29)$$

$$\nabla^2 p(x)x = (m-1) \nabla p(x) \quad (30)$$

$$\langle \nabla^2 p(x)x, x \rangle = m(m-1) p(x). \quad (31)$$

From (27), it follows that  $\nabla^2 p(x)$  is definite positive for all  $x \in \mathbb{R}^d \setminus 0$  if and only if  $\nabla^2 p(\vec{n})$  is positive definite for all  $\vec{n}$ , with  $\|\vec{n}\| = 1$ . Then,  $p$  belongs to  $\mathcal{P}_m^+$  only if  $m \geq 2$ .

### 4.3 Geometric ergodicity of the proposals: the case of Metropolis-Hastings algorithms

In this section we study the behaviour of our proposals within the Metropolis-Hastings framework. We will split our investigations in two parts: in the first we study fMALA and mOMA; while in the second we have a more detailed look in the properties of bOMA not only for the class  $\mathcal{E}(\beta, \gamma)$ , but also for the polynomial class  $\mathcal{P}_m^+$ .

### 4.3.1 Geometric ergodicity of fMALA, mOMA for the class $\mathcal{E}(\beta, \gamma)$

In the case  $\beta \in (0, 2)$ , fMALA and mOMA have their mean map behaving like  $x - \beta\gamma x |x|^{\beta-2} / 2$  at infinity and their variance map bounded from above. This is exactly the behaviour that MALA [22] has for the same values of  $\beta$ , thus one would expect them to behave in the same way. This is indeed the case and thus using the same reasoning as in the proof [22, Theorem 4.3] we deduce that the two algorithms are not geometrically ergodic for  $\beta \in (0, 1)$ . Similarly, the proof in [22, Theorem 4.1] can be used to show that the two algorithms are geometrically ergodic for  $\beta \in [1, 2)$ . Furthermore, for values of  $\beta \geq 2$  we have the following cases

(a) For  $\beta = 2$ ,

- fMALA is geometrically ergodic if  $h\gamma(1 + h\gamma/6) \in (0, 2)$  by [22, Theorem 4.1], and not geometrically ergodic if  $h\gamma(1 + h\gamma/6) > 2$  by Theorem 4.2, since  $\mu^{\text{fM}}$  is equivalent at infinity to  $(1 - h\gamma(1 + h\gamma/6))x$  and  $S^{\text{fM}}(x)$  is constant for  $|x| \geq R_\pi$ .
- Since  $\mu^{\text{mO}}$  is equivalent at infinity to  $(e^{-\gamma h} - 2(h\gamma)^2/3)x$ , we observe that mOMA is geometrically ergodic if  $h\gamma \in (0, 1.22)$  by [22, Theorem 4.1], and not geometrically ergodic if  $h\gamma > 1.23$  by Theorem [23, Theorem 5.1].

(b) For  $\beta > 2$ , fMALA and mOMA are not geometrically ergodic by Theorem 4.2 since the mean value maps of their proposal kernels are equivalent at infinity to  $-C_1 |x|^{2\beta-2} / x$ , their variance map to  $C_2 |x|^{\beta-2}$  for some constants  $C_1, C_2 > 0$ , and the variance maps are bounded from below.

### 4.3.2 Geometric ergodicity of bOMA

In this section, we give some conditions under which bOMA is geometrically ergodic and some examples of density which satisfy such conditions. For a matrix  $M \in \mathbb{R}^{d \times d}$ , we denote  $\lambda_{\min}(M) = \min \text{Sp}(M)$  and  $\lambda_{\max}(M) = \max \text{Sp}(M)$ , where  $\text{Sp}(M)$  is the spectrum of  $M$ . We can observe three different behaviours of the proposal given by (17) when  $x$  is large, which are implied by the behaviour of  $\lambda_{\min}(Df(x))$  and  $\lambda_{\max}(Df(x))$ .

If  $\liminf_{\|x\| \rightarrow +\infty} \lambda_{\min}(Df(x)) = 0$ . Then,  $g(x) = o(\|x\|^2)$  as  $\|x\| \rightarrow \infty$ , and  $y^{\text{bO}}$  tends to be as the MALA proposal at infinity, and we can show that bOMA is geometrically ergodic with the same conditions introduced in [22] for this one.

**Example 4.3.** By [22, Theorem 4.1] bOMA is geometrically ergodic for  $\pi \in \mathcal{E}(\gamma, \beta)$  with  $\beta \in [1, 2)$ .

Now, we focus on the case where  $\limsup_{\|x\| \rightarrow +\infty} \lambda_{\max}(Df(x)) < 0$ . For instance, this condition holds for  $\pi \in \mathcal{E}(\gamma, \beta)$  when  $\beta \geq 2$ . We give conditions similar to the one for geometric convergence of the Ozaki discretisation, given in [8], to check conditions of Lemma 4.1. Although these conditions does not cover all the cases, they seem to apply to interesting ones. Here are our assumptions where we denote by  $\mathbb{S}^d = \{x \in \mathbb{R}^d, \|x\| = 1\}$ , the sphere in  $\mathbb{R}^d$  and  $\vec{n}_x = x / \|x\|$ .

**Assumption 6.** We assume:

1.  $\limsup_{\|x\| \rightarrow +\infty} \lambda_{\max}(Df(x)) < 0$
2.  $\lim_{\|x\| \rightarrow +\infty} Df(x)^{-2} \{\mathbf{I}_d : D^2 f(x)\} = 0$

3.  $Df(x)^{-1}f(x)$  is asymptotically homogeneous to  $x$  when  $\|x\| \rightarrow +\infty$ , i.e. there exists a function  $c : \mathbb{S}^d \rightarrow \mathbb{R}$  such that

$$\lim_{\|x\| \rightarrow +\infty} \left\| \frac{Df(x)^{-1}f(x)}{\|x\|} - c(\vec{n}_x)\vec{n}_x \right\| = 0.$$

The condition 1 in Assumption 6 implies that for all  $x \in \mathbb{R}^d$ ,  $\lambda_{\max}(Df(x)) \leq M_f$ , and guarantees that  $S^{\text{bO}}(x, h)$  is bounded for all  $x \in \mathbb{R}^d$ .

**Lemma 4.4.** *Assume Assumption 4 and 6. There exists  $M_\Sigma \geq 0$  such that for all  $x \in \mathbb{R}^d$   $\|S^{\text{bO}}(x, h)\| \leq M_\Sigma$ .*

*Proof.* Since  $S^{\text{bO}}(x, h)$  is symmetric for all  $x \in \mathbb{R}^d$ , and  $t \mapsto (e^{ht} - 1)/t + (1/3)(e^{-(ht)^2} - 1)/t$  is bounded on  $(-\infty, M]$  for all  $M \in \mathbb{R}$ , we just need to show that there exists  $M_f \geq 0$  such that for all  $x$ ,  $\lambda_{\max}(Df(x)) \leq M_f$ . First, by Assumption 6-(1), there exists  $R \geq 0$ , such that for all  $x$ ,  $\|x\| \geq R$ ,  $\text{Sp}(Df(x)) \subset \mathbb{R}_-$ . In addition by Assumption 4  $x \mapsto Df(x)$  is continuous, and there exists  $M \geq 0$  such that for all  $x$ ,  $\|x\| \leq R$ ,  $\|Df(x)\| \leq M$ .  $\square$

**Theorem 4.5.** *Assume Assumptions 4, 5 and 6. If*

$$0 < \inf_{n \in \mathbb{S}^d} c(n) \leq \sup_{n \in \mathbb{S}^d} c(n) < 6/5, \quad (32)$$

*then bOMA is geometrically ergodic.*

*Proof.* We check that the conditions of Lemma 4.1 hold. By Assumption 4 and (17), Assumption 3 holds, thus it remains to check (21). First, Lemma 4.4 implies that the second equality of (21) is satisfied, and we just need to prove the first equality. By [8, Lemma 3.4], it suffices to prove that

$$\limsup_{\|x\| \rightarrow +\infty} \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle < 0, \quad (33)$$

where  $\eta(x) = \mu^{\text{bO}}(x, h) - x$ . Since  $\limsup_{\|x\| \rightarrow +\infty} \lambda_{\max}(Df(x)) < 0$  we can write  $\mathcal{G}(x) = \mathcal{B}(x)Df(x)^{-1}f(x)$ , where

$$\mathcal{B}(x) = (e^{(h/2)Df(x)} - \text{I}_d) + (2/3)(e^{-(hDf(x)/2)^2} - \text{I}_d),$$

and  $x \mapsto \mathcal{B}(x)$  is bounded on  $\mathbb{R}^d$ . Since  $\mathcal{B}$  is bounded on  $\mathbb{R}^d$ , by Assumption 6-(2)-(3) and (32),

$$\lim_{\|x\| \rightarrow +\infty} \left| \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle - \|\mathcal{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2 \langle \mathcal{B}(x)\vec{n}_x, \vec{n}_x \rangle c(\vec{n}_x) \right| = 0. \quad (34)$$

In addition, if we denote the eigenvalues of  $\mathcal{B}(x)$  by  $\{\lambda_i(x), i = 1, \dots, d\}$  and  $\{e_i(x), i = 1, \dots, d\}$  an orthonormal basis of eigenvectors, we have

$$\|\mathcal{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2 \langle \mathcal{B}(x)\vec{n}_x, \vec{n}_x \rangle c(\vec{n}_x) = \sum_{i=1}^d c(\vec{n}_x)\lambda_i(x) \langle e_i(x), \vec{n}_x \rangle^2 (c(\vec{n}_x)\lambda_i(x) + 2) \quad (35)$$

Since  $\limsup_{\|x\| \rightarrow +\infty} Df(x) < 0$ , for all  $i$  and  $\|x\|$  large enough,  $\lambda_i(x) \in [-5/3, 0)$ . Therefore using (32) we get from (35):

$$\|\mathcal{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2 \langle \mathcal{B}(x)\vec{n}_x, \vec{n}_x \rangle c(\vec{n}_x) < 0.$$

The proof is concluded using this result in (34).  $\square$

### Application to the convergence of bOMA for $\pi \in \mathcal{P}_m^+$

For the proof of the main result of this section, we need the following lemma.

**Lemma 4.6** ([8, Proof of Theorem 4.10]). *Let  $\pi \in \mathcal{P}_m^+$  for  $m \geq 2$ , then  $\pi$  satisfies Assumption 6-(3) with  $c(\vec{n}) = 1/(m-1) \in (0, 6/5)$  for all  $\vec{n} \in \mathbb{S}^d$ .*

**Proposition 4.7.** *Let  $\pi \in \mathcal{P}_m^+$  for  $m \geq 2$ , then bOMA is  $V$ -geometrically ergodic, with  $V(x) = \|x\|^2 + 1$ .*

*Proof.* Let us denote  $\pi \propto \exp(-p(x) - r(x))$ , with  $p$  and  $r$  satisfying the conditions from the definition in Section 4.2.2. We prove that if  $\pi \in \mathcal{P}_m^+$ , Theorem 4.5 can be applied. First, by definition of  $\mathcal{P}_m^+$ , Assumption 4 is satisfied. Furthermore, Assumption 6-(1)-(2) follows from (24), (27), (28) and the condition that  $\nabla^2 p(\vec{n})$  is positive definite for all  $\vec{n} \in \mathbb{S}^d$ . Also by Lemma 4.6, Assumption 6-(3) is satisfied.

Now we focus on Assumption 5. For ease of notation, in the following we denote  $\mu^{\text{bO}}$  and  $S^{\text{bO}}$  by  $\mu$  and  $S$ , and do not mention the dependence in the parameter  $h$  of  $\mu$  and  $S$  when it does not play any role. Note that

$$\int_{\mathbb{R}^d} \alpha(x, y) q(x, y) dy = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \{1 \wedge \exp \tilde{\alpha}(x, \xi)\} \exp(-\|\xi\|^2/2) d\xi, \quad (36)$$

where

$$\begin{aligned} \tilde{\alpha}(x, \xi) = & -p(\mu(x) + S(x)\xi) + p(x) - r(\mu(x) \\ & + S(x)\xi) + r(x) - \log(|S(\mu(x) + S(x)\xi)|) + \log(|S(x)|) + (1/2)\|\xi\|^2 \\ & - (1/2) \left\langle (\tilde{S}(x, \xi))^{-1} \{x - \mu(\mu(x) + S(x)\xi)\}, x - \mu(\mu(x) + S(x)\xi) \right\rangle, \end{aligned} \quad (37)$$

and  $\tilde{S}(x, \xi) = S(\mu(x) + S(x)\xi)S(\mu(x) + S(x)\xi)^T$ . First, we consider  $m \geq 3$ , then we have the following estimate of the terms in (37) by (24)-(28) and Lemma 4.6:

$$\mu(w) \underset{\|w\| \rightarrow +\infty}{=} \{1 - 5/(3(m-1))\}w + o(\|w\|) \quad (38)$$

$$(S(w)S(w)^T)^{-1} \underset{\|w\| \rightarrow +\infty}{=} \frac{3}{4}m(m-1)\|w\|^{m-2} \nabla^2 p(\vec{n}_w) + o(\|w\|^{m-2}) \quad (39)$$

$$\log(|S(w)|) \underset{\|w\| \rightarrow +\infty}{=} o(\|w\|) \quad (40)$$

Then by (38)-(40), if we define  $\Psi : [3, +\infty) \rightarrow \mathbb{R}$  by

$$m \mapsto 1 - \left\{1 - \frac{5}{3(m-1)}\right\}^m - (3/8)m(m-1) \left\{1 - \left(1 - \frac{5}{3(m-1)}\right)^2\right\}^2 \left\{1 - \frac{5}{3(m-1)}\right\}^{m-2},$$

we get

$$\tilde{\alpha}(x, \xi) \underset{\|x\| \rightarrow +\infty}{=} \|x\|^m p(\vec{n}_x) \Psi(m) + o(\|x\|^m).$$

Since  $\Psi$  is positive on  $[3, +\infty)$ , for all  $\xi \in \mathbb{R}^d$   $\lim_{\|x\| \rightarrow +\infty} \tilde{\alpha}(x, \xi) = +\infty$ . This result, (36) and Fatou's Lemma imply that Assumption 5 is satisfied.

For  $m = 2$ , we can assume  $p(x) = \langle Ax, x \rangle$  with  $A \in \mathcal{S}_+^*(\mathbb{R}^d)$ . Let us denote for  $M$  an invertible matrix of dimension  $p \geq 1$ ,

$$\begin{aligned}\varrho(M) &= (e^{-M} - I_p) + (2/3)(e^{-M^2} - I_p) \\ \varsigma(M) &= (e^{-2M} - I_p) + (1/3)(e^{-4M^2} - I_p).\end{aligned}$$

Then we have the following estimates:

$$\begin{aligned}\tilde{\alpha}(x, \xi) \Big|_{\|x\| \rightarrow +\infty} &= \langle A(\varsigma(hA))^{-1} \{ (2\varrho(hA) + \varrho(hA)^2) x \}, (2\varrho(hA) + \varrho(hA)^2) x \rangle \\ &+ \langle Ax, x \rangle - \langle A \{ (I_d + \varrho(hA))x \}, (I_d + \varrho(hA))x \rangle + o(\|x\|^2).\end{aligned}\quad (41)$$

If we denote the eigenvalues of  $A$  by  $\{\lambda_i, i = 1 \dots d\}$  and  $\{x_i, i = 1, \dots, d\}$  the coordinates of  $x$  in an orthonormal basis of eigenvectors for  $A$ , (41) becomes

$$\tilde{\alpha}(x, \xi) \Big|_{\|x\| \rightarrow +\infty} = \sum_{i=1}^d \Xi(h, \lambda_i) x_i^2 + o(\|x\|^2).\quad (42)$$

where for  $h, \lambda > 0$ ,

$$\Xi(h, \lambda) = \lambda (1 - (\varrho(h\lambda) + 1)^2 + \varsigma(h\lambda)^{-1} (4\varrho(h\lambda)^2 + 4\varrho(h\lambda)^3 + \varrho(h\lambda)^4)).$$

Using that for any  $h, \lambda > 0$ ,  $\Xi(h, \lambda) > 0$  and (42), we have for all  $\xi \in \mathbb{R}^d$ ,  $\lim_{\|x\| \rightarrow +\infty} \tilde{\alpha}(x, \xi) = +\infty$ , and as in the first case Assumption 5 is satisfied.  $\square$

**Remark 4.8.** Using the same reasoning as in Proposition 4.7, one can show that bOMA is geometrically ergodic for  $\pi \in \mathcal{E}(\beta, \gamma)$  with  $\beta \geq 2$ .

We now summarise the behaviour for all the different algorithms for the one dimensional class  $\mathcal{E}(\beta, \gamma)$  in Table 1

Method	$\beta \in [1, 2)$	$\beta = 2$	$\beta > 2$
fMALA (10)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
mOMA (11)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
bOMA (17)	geometrically ergodic	geometrically ergodic	geometrically ergodic

Table 1: Summary of ergodicity results for the Metropolis-Hastings algorithms for the class  $\mathcal{E}(\beta, \gamma)$

#### 4.4 Convergence of Gaussian Markov kernel on $\mathbb{R}$

We now present precise results for the ergodicity of the unadjusted proposals, by extending the results of [22] for the ULA to Gaussian Markov kernels on  $\mathbb{R}$ . Under Assumption 3, it is straightforward to see that  $Q$  is  $\text{Leb}^d$ -irreducible, where  $\text{Leb}^d$  is the Lebesgue measure, aperiodic and all compact set of  $\mathbb{R}^d$  are small; see [8, Theorem 3.1]. We now state our main theorems, which essentially complete [22, Theorem 3.1-3.2]. Since their proof are very similar, they are omitted.

**Theorem 4.9.** Assume Assumption 3, and there exist  $s_\wedge, u_+, u_- \in \mathbb{R}_+^*$  and  $\chi \in \mathbb{R}$  such that:

$$\limsup_{|x| \rightarrow +\infty} S(x) \leq s_\wedge ,$$

$$\lim_{x \rightarrow +\infty} \{\mu(x) - x\} x^{-\chi} = -u_+ , \text{ and } \lim_{x \rightarrow +\infty} \{\mu(x) - x\} |x|^{-\chi} = u_- .$$

- (1) If  $\chi \in [0, 1)$ , then  $Q$  is geometrically ergodic.
- (2) If  $\chi = 1$  and  $(1 - u_+)(1 - u_-) < 1$ , then  $Q$  is geometrically ergodic.
- (3) If  $\chi \in (-1, 0)$ , then  $Q$  is ergodic but not geometrically ergodic.

*Proof.* See the proof of [22, Theorem 3.1] □

**Theorem 4.10.** Assume Assumption 3, and there exist  $s_\vee, u_+, u_- \in \mathbb{R}_+^*$  and  $\chi \in \mathbb{R}$  such that:

$$\liminf_{|x| \rightarrow +\infty} S(x) \geq s_\vee ,$$

$$\lim_{x \rightarrow +\infty} S(x)^{-1} \mu(x) x^{-\chi} = -u_+ , \text{ and } \lim_{x \rightarrow +\infty} S(x)^{-1} \mu(x) |x|^{-\chi} = u_- .$$

- (1) If  $\chi > 1$ , then  $Q$  is transient.
- (2) If  $\chi = 1$  and  $(u_+ \wedge u_-) s_\vee > 1$ , then  $Q$  is transient.

*Proof.* See the proof of [22, Theorem 3.2] □

### Ergodicity of the unadjusted proposals for the class $\mathcal{E}(\beta, \gamma)$

We now apply Theorems 4.9 and 4.10 in order to study the ergodicity of the different unadjusted proposals applied to  $\pi \in \mathcal{E}(\beta, \gamma)$ . In the case  $\beta \in (0, 2)$  all the three algorithms (fULA, mUOA, bUOA) have their mean map behaving like  $x - \beta\gamma x |x|^{\beta-2}/2$  at infinity and their variance map bounded from above. This is exactly the behaviour that ULA [22] has for the same values of  $\beta$ , thus it should not be a surprise that Theorem 4.9 implies that all the three algorithms behaved as the ULA does for the corresponding values, namely being ergodic for  $\beta \in (0, 1)$  and geometrically ergodic for  $\beta \in [1, 2)$ . Furthermore, for values of  $\beta \geq 2$  we have the following cases

(a) For  $\beta = 2$ ,

- fULA is geometrically ergodic if  $h\gamma(1 + h\gamma/6) \in (0, 2)$  by Theorem 4.9-(2), and is transient if  $h\gamma(1 + h\gamma/6) > 2$  by Theorem 4.10-(2), since  $\mu^{\text{fM}}$  is equivalent at infinity to  $(1 - h\gamma(1 + h\gamma/6))x$  and  $S^{\text{fM}}(x)$  is constant for  $|x| \geq R_\pi$ .
- mUOA is geometrically ergodic if  $1 + 2(h\gamma)^2/3 - e^{-\gamma h} \in (0, 2)$  by Theorem 4.9-(2), and is transient if  $1 + 2(h\gamma)^2/3 - e^{-\gamma h} > 2$  by Theorem 4.10-(2), since  $\mu^{\text{mO}}$  is equivalent at infinity to  $(e^{-\gamma h} - 2(h\gamma)^2/3)x$  and  $S^{\text{mO}}(x)$  is constant for  $|x| \geq R_\pi$ .
- bUOA is geometrically ergodic by Theorem 4.9-(2), since  $\mu^{\text{bO}}$  is equivalent at infinity to  $-2x/3$  and  $S^{\text{bO}}(x)$  is constant for  $|x| \geq R_\pi$ .

(b) For  $\beta > 2$ ,

- fULA and mUOA are transient by Theorem 4.10-(1) since their mean value map is equivalent at infinity to  $-C_1 |x|^{2\beta-2}/x$ , and their variance map to  $C_2 |x|^{\beta-2}$  for some constants  $C_1, C_2 > 0$ , and their variance map are bounded from below.
- bUOA is geometrically ergodic by Theorem 4.9-(1) since its mean value map is equivalent at infinity to  $\{1 - 5/(3(\beta - 1))\}x$  and its variance map is bounded from above.

The summary of our findings can be found in Table 2

Method	$\beta \in (0, 1)$	$\beta \in [1, 2)$	$\beta = 2$	$\beta > 2$
fULA (10)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
mUOA (11)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
bUOA (17)	ergodic	geometrically ergodic	geometrically ergodic	geometrically ergodic

Table 2: Summary of ergodicity results for the unadjusted proposals for the class  $\mathcal{E}(\beta, \gamma)$

## 5 Numerical illustration of the improved efficiency

In this section, we illustrate our analysis (Section 3.1) of the asymptotic behaviour of fMALA as the dimension  $d$  tends to infinity, and we demonstrate its gain of efficiency as  $d$  increases compared to the standard MALA. Following [20], we define the first-order efficiency of a multidimensional Markov chain  $\{X_k, k \in \mathbb{N}\}$  with first component denoted  $X_k^{(1)}$  as  $\mathbb{E}[(X_{k+1}^{(1)} - X_k^{(1)})^2]$ . In Figure 1, we consider as a test problem the product case (7) where we take the double well potential with  $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$ . We consider the dimensions  $d = 10, 50, 100, 200$ , respectively, and we plot for many time stepsizes  $h = \ell^2 d^{-1/5}$  the first order efficiency (multiplied by the dimension  $d$  because this is the scale which is asymptotically constant as  $d \rightarrow \infty$ ) as a function of the acceptance rate for the standard MALA (white bullets) and the improved version fMALA (black bullets). The expectancies is approximated as the average over  $2 \times 10^5$  iterations of the algorithms and we use the same sets of generated random numbers for both methods. For comparison, we also include (as solid lines) the asymptotic efficiency curve of fMALA as  $d \rightarrow \infty$ , normalised to have the same maximum as fMALA in finite dimension  $d$ . Precisely, it corresponds to the (rescaled) limiting diffusion speed  $h^{\text{fM}}(\ell)$  as a function of  $a^{\text{fM}}(\ell)$  (quantities given respectively in Theorems 3.1 and 3.2). We observe an excellent agreement of the numerical first order efficiency compared to the asymptotic one, especially as  $d$  increases, which corroborated the scaling results of fMALA. We observe for the considered dimensions  $d$  that the optimal acceptance rate maximizing the first-order efficiency remains very close to the limiting value of 0.704 predicted in Theorem 3.2. This numerical experiment shows that the efficiency improvement of fMALA compared to MALA is significant and indeed increases as the dimension  $d$  increases, which confirms the analysis of Section 3.1.

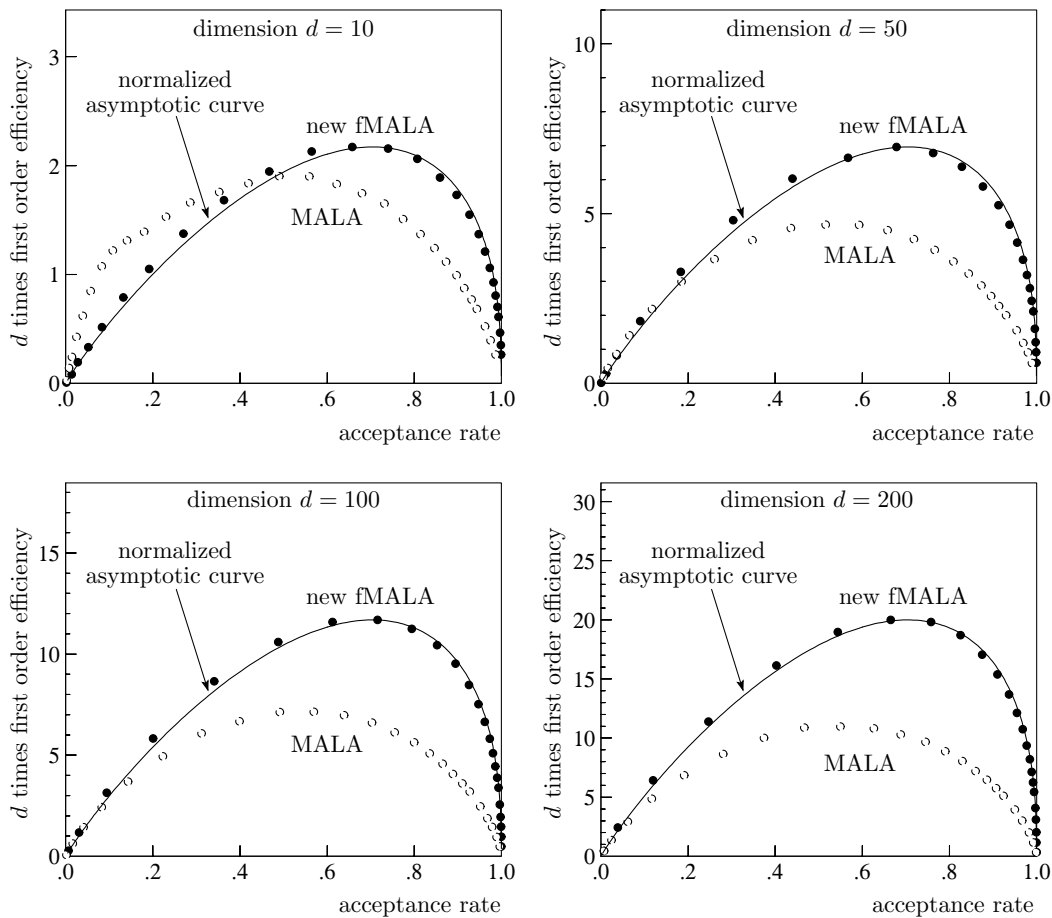


Figure 1: First-order efficiency of the new fMALA and the standard MALA as a function of the overall acceptance rates for the dimensions  $d = 10, 50, 100, 200$ , respectively, for the double well potential with  $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$ . The solid line is the reference asymptotic curve of efficiency for the new fMALA, normalised to have the same maximum value as the finite dimensional fMALA.

**Acknowledgements.** We thank Gabriel Stoltz for fruitful discussions. The research of G. O. R. is partially supported by EPSRC under grants EP/K014463/1 and EP/D002060/1, and that of G. V. by the Swiss National Foundation, Grant No 200020\_144313/1.

## References

- [1] A. Abdulle, G. Vilmart, and K. C. Zygalakis. High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.*, 52(4):1600–1622, 2014.
- [2] A. Beskos and A. Stuart. MCMC methods for sampling function space. In *ICIAM 07—6th International Congress on Industrial and Applied Mathematics*, pages 337–364. Eur. Math. Soc., Zürich, 2009.
- [3] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience

Publication.

- [4] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013.
- [5] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
- [6] M. Fathi and G. Stoltz. Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics. *preprint*, 2015.
- [7] M. Girolami and B. Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.
- [8] N. R. Hansen. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, 9(4):725–743, 2003.
- [9] R. Z. Has'minskiĭ. *Stochastic stability of differential equations*, volume 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980. Translated from the Russian by D. Louvish.
- [10] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.
- [11] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.
- [12] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.
- [13] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, 48(2):552–577, 2010.
- [14] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.
- [15] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
- [16] G. N. Milstein and M. V. Tretyakov. Computing ergodic limits for Langevin equations. *Phys. D*, 229(1):81–95, 2007.
- [17] T. Ozaki. A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach. *Statist. Sinica*, 2(1):113–135, 1992.
- [18] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

- [19] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997.
- [20] G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(1):255–268, 1998.
- [21] G. O. Roberts and O. Stramer. Langevin diffusions and Metropolis-Hastings algorithms. *Methodol. Comput. Appl. Probab.*, 4(4):337–357 (2003), 2002. International Workshop in Applied Probability (Caracas, 2002).
- [22] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.
- [23] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.
- [24] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, 8(4):483–509 (1991), 1990.
- [25] I. Wolfram Research. Mathematica. 2014.

## A Proof of Theorems 3.1 and 3.2

We provide here the proofs of Theorems 3.1 and 3.2 for the analysis of the optimal scaling properties of fMALA. We use tools analogous to that of [19] and [20]. Consider the generator of the jump process  $\Gamma^{d,\text{fM}}$ , defined for  $\psi^d \in C_c^2(\mathbb{R}^d)$ , and  $x \in \mathbb{R}^d$  by

$$A_d^{\text{fM}}\psi^d(x) = d^{1/5}\mathbb{E} \left[ (\psi^d(y) - \psi^d(x))\alpha_d^{\text{fM}}(x, y) \right] ,$$

where  $y$  follows the distribution defined by  $q_d^{\text{fM}}(x, \cdot)$ . Also, consider the generator of the process  $\{G_t, t \geq 0\}$ , solution of (16), defined for  $\psi \in C_c^2(\mathbb{R})$ , and  $x \in \mathbb{R}^d$  by

$$A^{\text{fM}}\psi(x) = (h(\ell)/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) .$$

We check that the assumptions of [5, Corollary 8.7, Chapter 4] are satisfied, which will imply Theorem 3.2. These assumptions consist in showing there exists a sequence of set  $\{F_d \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$  such that for all  $T \geq 0$ :

$$\begin{aligned} \lim_{d \rightarrow +\infty} \mathbb{P} \left[ \Gamma_s^{d,\text{fM}} \in F_d , \forall s \in [0, T] \right] &= 1 \\ \lim_{d \rightarrow +\infty} \sup_{x \in F_d} \left| A_d^{\text{fM}}\psi(x) - A^{\text{fM}}\psi(x) \right| &= 0 , \end{aligned}$$

for all functions  $\psi$  in a core of  $A^{\text{fM}}$ , which strongly separates points. Since  $A^{\text{fM}}$  is an operator on the set of function only depending on the first component, we restrict our study on this class of functions, which belong to  $C_c^\infty(\mathbb{R})$ , since by [5, Theorem 2.1, Chapter 8], this set of functions is a core for  $A^{\text{fM}}$  which strongly separates points. The following lemma is the proper result which was introduced in Section 2.2. For the sequel, let  $\{\xi_i, i \in \mathbb{N}^*\}$  be a sequence of i.i.d. standard one-dimensional Gaussian random variables and  $X$  be a random variable distributed according to  $\pi_1$ . Also, for all  $x \in \mathbb{R}^d$ , denote by  $y^{\text{fM}}$  the proposal of fMALA, defined by (8), (10a) and (10b), started at  $x \in \mathbb{R}^d$ , with parameter  $h_d$  and associated with the  $d$ -dimensionnal Gaussian random variable  $\{\xi_i, i = 1, \dots, d\}$ .

**Lemma A.1.** *Assume Assumption 2. The following Taylor expansion in  $\sqrt{h_d}$  holds: for all  $x \in \mathbb{R}^d$  and  $i \in \{1, \dots, d\}$ ,*

$$\log \left( \frac{\pi(y_i^{\text{fM}})q^{\text{fM}}(y_i^{\text{fM}}, x_i)}{\pi(x_i)q^{\text{fM}}(x_i, y_i^{\text{fM}})} \right) = C_5^{\text{fM}}(x_i, \xi_i)d^{-1/2} + C_6^{\text{fM}}(x_i, \xi_i)d^{-3/5} + C_7^{\text{fM}}(x_i, \xi_i)d^{-7/10} \\ + C_8^{\text{fM}}(x_i, \xi_i)d^{-4/5} + C_9^{\text{fM}}(x_i, \xi_i)d^{-9/10} + C_{10}^{\text{fM}}(x_i, \xi_i)d^{-1} + C_{11}^{\text{fM}}(x_i, \xi_i, h_d), \quad (43)$$

where  $C_5^{\text{fM}}(x_1, \xi_1)$  is given in Appendix C. Furthermore, for  $j = 6, \dots, 10$ ,  $C_j^{\text{fM}}(x_i, \xi_i)$  are polynomials in  $\xi_i$  and derivatives of  $g$  at  $x_i$  and

$$\mathbb{E} \left[ C_j^{\text{fM}}(\mathbf{X}, \xi_1) \right] = 0 \text{ for } j = 5, \dots, 9, \quad (44)$$

$$\mathbb{E} \left[ \left( \mathbb{E} \left[ C_5^{\text{fM}}(\mathbf{X}, \xi_1) | \mathbf{X} \right] \right)^2 \right] = \ell^{10}(K^{\text{fM}})^2 = -2\mathbb{E} \left[ C_{10}^{\text{fM}}(\mathbf{X}, \xi_1) \right]. \quad (45)$$

In addition, there exists a sequence of sets  $\{F_d^1 \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$  such that  $\lim_{d \rightarrow +\infty} d^{1/5}\pi_d((F_d^1)^c) = 0$  and for  $j = 6, \dots, 10$

$$\lim_{d \rightarrow +\infty} d^{-3/5} \sup_{x \in F_d^1} \mathbb{E} \left[ \left| \sum_{i=2}^d C_j(x_i^d, \xi_i) - \mathbb{E} \left[ C_j^{\text{fM}}(\mathbf{X}, \xi_i) \right] \right| \right] = 0, \quad (46)$$

and

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^1} \mathbb{E} \left[ \left| \sum_{i=2}^d C_{11}(x_i^d, \xi_i, h_d) \right| \right] = 0. \quad (47)$$

Finally,

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^1} \mathbb{E} \left[ \left| \zeta^d \right| \right] = 0, \quad (48)$$

with

$$\zeta^d = \sum_{i=2}^d \log \left( \frac{\pi(y_i^{\text{fM}})q^{\text{fM}}(y_i^{\text{fM}}, x_i)}{\pi(x_i)q^{\text{fM}}(x_i, y_i^{\text{fM}})} \right) - \left( \left( d^{-1/2} \sum_{i=2}^d C_5(x_i^d, \xi_i) \right) - \ell^{10}(K^{\text{fM}})^2/2 \right).$$

*Proof.* The Taylor expansion was computed using the computational software Mathematica [25]. Then, since just odd powers of  $\xi_i$  occur in  $C_5, C_7$  and  $C_9$ , we deduce (44) for  $j = 5, 7, 9$ . Furthermore by explicit calculation, the anti-derivative in  $x_1$  of  $e^{g(x_1)}\mathbb{E} \left[ C_j^{\text{fM}}(x_1, \xi_1) \right]$ , for  $j = 6, 8$ , and  $e^{g(x_1)}\mathbb{E} \left[ C_5^{\text{fM}}(x_1, \xi_1)^2 + 2C_{10}^{\text{fM}}(x_1, \xi_1) \right]$  are on the form of some polynomials in the derivatives of  $g$  in  $x_1$  times  $e^{g(x_1)}$ . Therefore, Assumption 2-(3) implies (44) for  $j = 6, 8$  and (45). We now build the sequence of sets  $F_d^1$ , which satisfies the claimed properties.

Denote for  $j = 6, \dots, 10$  and  $x_i \in \mathbb{R}$ ,  $\tilde{C}_j^{\text{fM}}(x_i) = \mathbb{E} \left[ C_j^{\text{fM}}(x_i, \xi_i) \right]$  and  $V_j^{\text{fM}}(x_i) = \text{Var} \left[ C_j^{\text{fM}}(x_i, \xi_i) \right]$ , which are bounded by a polynomial  $P_1$  in  $x_i$  by Assumption 2-(2) since  $C_j^{\text{fM}}(x_i, \xi_i)$  are polynomials in  $\xi_i$  and the derivatives of  $g$  at  $x_i$ . Therefore for all  $k \in \mathbb{N}^*$ ,

$$\mathbb{E} \left[ \left| \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right|^k \right] + \mathbb{E} \left[ \left| V_j^{\text{fM}}(\mathbf{X}) \right|^k \right] < +\infty. \quad (49)$$

Consider for all  $j = 6, \dots, 10$ , the sequence of sets  $F_{d,j}^1 \in \mathbb{R}^d$  defined by  $F_{d,j}^1 = F_{d,j,1}^1 \cap F_{d,j,2}^1$  where

$$F_{d,j,1}^1 = \left\{ x \in \mathbb{R}^d ; \left| \sum_{i=2}^d \tilde{C}_j^{\text{fM}}(x_i) - \mathbb{E} \left[ \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right] \right| \leq d^{23/40} \right\} \quad (50)$$

$$F_{d,j,2}^1 = \left\{ x \in \mathbb{R}^d ; \left| \sum_{i=2}^d V_j^{\text{fM}}(\mathbf{X}) - \mathbb{E} \left[ V_j^{\text{fM}}(\mathbf{X}) \right] \right| \leq d^{23/20} \right\}. \quad (51)$$

Note that  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_{d,j}^1)^c) = 0$  for all  $j = 6 \dots 10$ , is implied by  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_{d,j,1}^1)^c) = 0$  and  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_{d,j,2}^1)^c) = 0$ . Let  $\{\mathbf{X}_i, i \geq 2\}$  be a sequence of i.i.d.random variables with distribution  $\pi_1$ . By definition of  $F_{d,j,1}^1$ , the Markov inequality and independence, we get

$$\begin{aligned} d^{1/5} \pi_d((F_{d,j,1}^1)^c) &\leq d^{-21/10} \mathbb{E} \left[ \left( \sum_{i=2}^d \tilde{C}_j^{\text{fM}}(\mathbf{X}_i) - \mathbb{E} \left[ \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right] \right)^4 \right] \\ &\leq \sum_{i_1, i_2=2}^d \mathbb{E} \left[ \left( \tilde{C}_j^{\text{fM}}(\mathbf{X}_{i_1}) - \mathbb{E} \left[ \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right] \right)^2 \left( \tilde{C}_j^{\text{fM}}(\mathbf{X}_{i_2}) - \mathbb{E} \left[ \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right] \right)^2 \right] \\ &\leq d^{-1/10} \mathbb{E} \left[ \left( \tilde{C}_j^{\text{fM}}(\mathbf{X}) - \mathbb{E} \left[ \tilde{C}_j^{\text{fM}}(\mathbf{X}) \right] \right)^4 \right], \end{aligned} \quad (52)$$

where we have used the Young inequality for the last line. On another hand, using the Chebyshev and Hölder inequality, we get

$$\begin{aligned} d^{1/5} \pi_d((F_{d,j,2}^1)^c) &\leq d^{-21/10} \mathbb{E} \left[ \left( \sum_{i=2}^d V_j^{\text{fM}}(\mathbf{X}_i) - \mathbb{E} \left[ V_j^{\text{fM}}(\mathbf{X}) \right] \right)^2 \right] \\ &\leq d^{-1/10} \mathbb{E} \left[ \left( V_j^{\text{fM}}(\mathbf{X}) - \mathbb{E} \left[ V_j^{\text{fM}}(\mathbf{X}) \right] \right)^2 \right]. \end{aligned} \quad (53)$$

Therefore (49), (52) and (53) imply that  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_{d,j}^1)^c) = 0$  for all  $j = 6, \dots, 10$ . In addition, for all  $x \in F_{d,j}^1$ , by the triangle inequality and the Cauchy-Schwarz inequality we have for all  $j = 6, \dots, 10$

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{i=2}^d C_j^{\text{fM}}(x_i, \xi_i) - \mathbb{E} \left[ C_j^{\text{fM}}(\mathbf{X}, \xi_i) \right] \right| \right] &\leq \left| \sum_{i=2}^d V_j^{\text{fM}}(x_i) - \mathbb{E} \left[ V_j^{\text{fM}}(\mathbf{X}) \right] \right|^{1/2} \\ &\quad + d^{1/2} \mathbb{E} \left[ V_j^{\text{fM}}(\mathbf{X}) \right]^{1/2} + \left| \sum_{i=2}^d \tilde{C}_j^{\text{fM}}(x_i) - \mathbb{E} \left[ C_j^{\text{fM}}(\mathbf{X}, \xi_i) \right] \right|. \end{aligned}$$

Therefore by this inequality, (50) and (51), there exists a constant  $M_1$  such that

$$d^{3/5} \sup_{x \in F_{d,j}^1} \mathbb{E} \left[ \left| \sum_{i=2}^d C_j^{\text{fM}}(x_i, \xi_i) - \mathbb{E} \left[ C_j^{\text{fM}}(\mathbf{X}, \xi_i) \right] \right| \right] \leq d^{-1/40} M_1,$$

and (46) follows. It remains to show (47). By definition,  $C_{11}$  is the remainder in the eleventh order expansion in  $\sigma_d := \sqrt{h_d}$  given by (43) of the function  $\Theta$  defined by  $\Theta(x_i, \xi_i, \sigma_d) =$

$\log(\pi_1(y_i^{\text{fM}})q_1^{\text{fM}}(y_i^{\text{fM}}, x_i)) - \log(\pi_1(x_i)q_1^{\text{fM}}(x_i, y_i^{\text{fM}}))$ . Therefore, by the mean-value form of the remainder, there exists  $u_d \in [0, \sigma_d]$  such that

$$C_{11}(x_i, \xi_i, h_d) = (\sigma_d^{11}/(11!)) \frac{\partial^{11}\Theta}{\partial \sigma_d^{11}}(x_i, \xi_i, u_d) .$$

By Assumption 2-(1) which implies that  $g''$  is bounded, and Assumption 2-(2), for all  $u_d \in [0, \sigma_d]$ , the eleventh derivative of  $\Theta$  with respect to  $\sigma_d$ , taken in  $(x_i, \xi_i, u_d)$ , can be bounded by a positive polynomial in  $(x_i, \xi_i)$  on the form  $P_2(x_i)P_3(\xi_i)$ . Hence, there exists a constant  $M_2$  such that

$$\mathbb{E} [|C_{11}(x_i, \xi_i, h_d)|] \leq M_2 d^{-11/10} P_2(x_i) . \quad (54)$$

And if we define

$$F_{d,11}^1 = \left\{ x \in \mathbb{R}^d ; \left| \sum_{i=2}^d P_2(x_i) - \mathbb{E}[P_2(X)] \right| \leq d \right\} ,$$

then we have by the Chebychev inequality, this definition and (54)

$$\begin{aligned} d^{1/5} \pi_d((F_{d,11}^1)^c) &\leq \text{Var}[P_2(X)] d^{-4/5} \\ \sup_{x \in F_{d,11}^1} \sum_{i=2}^d \mathbb{E} [|C_{11}(x_i, \xi_i, h_d)|] &\leq M_2 (\mathbb{E}[P_2(x)] + 1) d^{-1/10} . \end{aligned}$$

These results, combined with Assumption 2-(3), imply  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_{d,11}^1)^c) = 0$  and (47). Finally,  $F_d^1 = \bigcap_{j=6}^{11} F_{d,j}^1$  satisfies the claimed properties of the Lemma, and (48) directly follows from all the previous results.  $\square$

To isolate the first component of the process  $\Gamma^{d,\text{fM}}$ , we consider the modified generators defined for  $\psi \in C_c^2(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$  by

$$\tilde{A}_d^{\text{fM}} \psi(x) = d^{1/5} \mathbb{E} \left[ (\psi(y^{\text{fM}}) - \psi(x)) \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right]$$

where for all  $x, y \in \mathbb{R}^d$ ,

$$\alpha_{-1,d}^{\text{fM}}(x, y) = \prod_{i=2}^d \frac{\pi_1(y_i) q_{1,\text{fM}}(y_i, x_i)}{\pi_1(x_i) q_{1,\text{fM}}(x_i, y_i)} .$$

The next lemma shows that we can approximate  $A_d^{\text{fM}}$  by  $\tilde{A}_d^{\text{fM}}$ , and thus, in essence, the first component becomes ‘‘asymptotically independent’’ from the others.

**Theorem A.2.** *There exists a sequence of sets  $\{F_d^2 \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$  such that  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_d^2)^c) = 0$  and for all  $\psi \in C_c^\infty(\mathbb{R})$  (seen as function of  $\mathbb{R}^d$  for all  $d$  which only depends on the first component):*

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^2} \left| A_d^{\text{fM}} \psi(x) - \tilde{A}_d^{\text{fM}} \psi(x) \right| = 0 .$$

In addition,

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^2} d^{1/5} \mathbb{E} \left[ \left| \alpha_d^{\text{fM}}(x, y^{\text{fM}}) - \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right| \right] = 0 . \quad (55)$$

*Proof.* Using that  $\psi$  is bounded and the Jensen inequality, there exists a constant  $M_1$  such that

$$\left| A_d^{\text{fM}}\psi(x) - \tilde{A}_d^{\text{fM}}\psi(x) \right| \leq M_1 d^{1/5} \mathbb{E} \left[ \left| \alpha_d^{\text{fM}}(x, y^{\text{fM}}) - \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right| \right].$$

Thus it suffices to show (55). Set  $\sigma_d = \sqrt{h_d}$ . Since  $t \mapsto 1 \wedge \exp(t)$  is 1-Lipschitz on  $\mathbb{R}$  and, by definition we have

$$d^{1/5} \mathbb{E} \left[ \left| \alpha_d^{\text{fM}}(x, y^{\text{fM}}) - \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right| \right] \leq d^{1/5} \mathbb{E} [|\Theta(x_1, \xi_1, \sigma_d)|], \quad (56)$$

where  $\Theta(x_1, \xi_1, \sigma_d) = \log(\pi_1(y_1^{\text{fM}})q_1^{\text{fM}}(y_1^{\text{fM}}, x_1)) - \log(\pi_1(x_1)q_1^{\text{fM}}(x_1, y_1^{\text{fM}}))$ . By a fifth order Taylor expansion of  $\Theta$  in  $\sigma_d$ , and since by (43)  $\partial^j \Theta(x_1, \xi_1, 0)/(\partial \sigma_d^j) = 0$  for  $j = 0 \dots 4$ , we have

$$\Theta(x_1, \xi_1, \sigma_d) = \frac{\partial^5 \Theta}{\partial \sigma_d^5}(x_1, \xi_1, u_d)(\sigma_d^5/5!),$$

for some  $u_d \in [0, \sigma_d]$ . Using Assumption 2-(1)-(2), and an explicit expression of  $\partial^j \Theta(x_1, \xi_1, u_d)/(\partial \sigma_d^j)$ , there exists two positive polynomials  $P_1$  and  $P_2$  such that

$$|\Theta(x_1, \xi_1, \sigma_d)| \leq (\sigma_d^5/5!)P_1(x_1)P_2(\xi_1).$$

Plugging this result in (56) and since  $\sigma_d^5 = \ell^{5/2}d^{-1/2}$ , we get

$$d^{1/5} \mathbb{E} \left[ \left| \alpha_d^{\text{fM}}(x, y^{\text{fM}}) - \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right| \right] \leq \ell^{5/2}d^{-3/10}P_1(x_1).$$

Setting  $F_d^2 = \{x \in \mathbb{R}^d ; P_1(x_1) \leq d^{1/10}\}$ , we have

$$\sup_{x \in F_d^2} d^{1/5} \mathbb{E} \left[ \left| \alpha_d^{\text{fM}}(x, y^{\text{fM}}) - \alpha_{-1,d}^{\text{fM}}(x, y^{\text{fM}}) \right| \right] \leq \ell^{5/2}d^{-1/5},$$

and (55) follows. Finally,  $F_d^2$  satisfied  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_d^2)^c) = 0$  since by the Markov inequality

$$d^{1/5} \pi_d((F_d^2)^c) \leq d^{-1/10} \mathbb{E} [P_1(X)^3],$$

where  $\mathbb{E} [P_1(X)^3]$  is finite by Assumption 2-(3).  $\square$

**Lemma A.3.** For all  $\psi \in C_c^\infty(\mathbb{R})$ ,

$$\lim_{d \rightarrow +\infty} \sup_{x_1 \in \mathbb{R}} \left| d^{1/5} \mathbb{E} \left[ \psi(y_1^{\text{fM}}) - \psi(x_1) \right] - (\ell^2/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) \right| = 0.$$

*Proof.* Consider  $\sigma_d = \sqrt{h_d}$  and  $W(x_1, \xi_1, \sigma_d) = \psi(y_1^{\text{fM}})$ . Note that  $W(x_1, \xi_1, 0) = \psi(x_1)$ . Then using that  $\psi \in C_c^\infty(\mathbb{R})$ , a third order Taylor expansion of this function in  $\sigma_d$  implies there exists  $u_d \in [0, h_d]$  and  $M_1 \geq 0$  such that

$$\begin{aligned} \mathbb{E} [W(x_1, \xi_1, \sigma_d) - \psi(x_1)] &= (\ell^2 d^{-1/5}/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) + M_1 d^{-3/10} \\ &\quad + \frac{\partial^3 W}{\partial \sigma_d^3}(x_1, \xi_1, u_d)\sigma_d^3. \end{aligned}$$

Moreover since  $\psi \in C_c^\infty(\mathbb{R})$ , the third partial derivative of  $W$  in  $\sigma_d$  are bounded for all  $x_1$ ,  $\xi_1$  and  $\sigma_d$ . Therefore there exists  $M_2 \geq 0$  such that for all  $x_1 \in \mathbb{R}$ ,

$$\left| d^{1/5} \mathbb{E} \left[ \psi(y_1^{\text{fM}}) - \psi(x_1) \right] - (\ell^2/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) \right| \leq M_2 \ell^{3/2} d^{-1/10},$$

which concludes the proof.  $\square$

As proceed in [20], we prove a uniform central limit theorem for the sequence of random variables defined for  $i \geq 2$  and  $x_i \in \mathbb{R}$  by  $C_5^{\text{fM}}(x_i, \xi_i)$ . Define now for  $d \geq 2$  and  $x \in \mathbb{R}^d$ ,

$$\bar{M}_d(x) = n^{-1/2} \sum_{i=2}^d C_5^{\text{fM}}(x_i, \xi_i),$$

and the characteristic function of  $\bar{M}_d$  for  $t \in \mathbb{R}$  by

$$\varphi_d(x, t) = \mathbb{E}[e^{it\bar{M}_d(x)}].$$

Finally define the characteristic function of the zero-mean Gaussian distribution with standard deviation  $K^{\text{fM}}$ , given in Lemma A.1, by: for  $t \in \mathbb{R}$ ,

$$\varphi(t) = e^{-(K^{\text{fM}})^2 t^2 / 2}.$$

**Lemma A.4.** *There exists a sequence of set  $\{F_d^3 \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$ , satisfying  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_d^3)^c) = 0$  and we have the following properties:*

(i) *for all  $t \in \mathbb{R}$ ,  $\lim_{d \rightarrow +\infty} \sup_{x \in F_d^3} |\varphi_d(x, t) - \varphi(t)| = 0$ ,*

(ii) *for all bounded continuous function  $b : \mathbb{R} \rightarrow \mathbb{R}$ ,*

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^3} \left| \mathbb{E}[b(\bar{M}_d(x))] - (2\pi\ell^{10}(K^{\text{fM}})^2)^{-1/2} \int_{\mathbb{R}} b(u) e^{-u^2/(2\ell^{10}(K^{\text{fM}})^2)} du \right| = 0.$$

In particular, we have

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^3} \left| \mathbb{E} \left[ 1 \wedge e^{\bar{M}_d(x) - \ell^{10}(K^{\text{fM}})^2 / 2} \right] - 2\Phi(\ell^5 K^{\text{fM}} / 2) \right| = 0.$$

*Proof.* We first define for all  $d \geq 1$ ,  $F_d^3 = F_{d,1}^3 \cap F_{d,2}^3$  where

$$F_{d,1}^3 = \bigcap_{j=2,4} \left\{ x \in \mathbb{R}^d ; \left| d^{-1} \sum_{i=2}^d \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^j \right] - \mathbb{E} \left[ C_5^{\text{fM}}(X_1, \xi_1)^j \right] \right| \leq d^{-1/4} \right\}, \quad (57)$$

$$F_{d,2}^3 = \left\{ x \in \mathbb{R}^d ; \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^2 \right] \leq d^{3/4} \quad \forall i \in \{2, \dots, d\} \right\}. \quad (58)$$

It follows from (49), and the Chebychev and Markov inequalities that there exists a constant  $M$  such that  $\pi_d((F_{d,1}^3)^c) + \pi_d((F_{d,2}^3)^c) \leq Md^{-1/2}$ . Therefore  $\lim_{d \rightarrow +\infty} d^{1/5} \pi_d((F_d^3)^c) = 0$ .

(i). Let  $t \in \mathbb{R}$  and  $x \in F_d^3$  and denote

$$V(x_i) = \text{Var}[C_5^{\text{fM}}(x_i, \xi_i)] = \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^2 \right],$$

where the second equality follows from Lemma A.1. By the triangle inequality

$$\begin{aligned} |\varphi_d(x, t) - \varphi(t)| &\leq \left| \varphi_d(x, t) - \prod_{i=2}^d \left( 1 - \frac{\ell^{10} V(x_i) t^2}{2d} \right) \right| \\ &\quad + \left| \prod_{i=2}^d \left( 1 - \frac{\ell^{10} V(x_i) t^2}{2d} \right) - e^{-\ell^{10}(K^{\text{fM}})^2 t^2 / 2} \right|. \quad (59) \end{aligned}$$

We bound the two terms of the right hand side separately. Note that by independence for all  $d$ ,  $\varphi_d(x, t) = \prod_{i=2}^d \varphi_1(x_i, t/\sqrt{d})$ . Since  $x \in F_d^3$ , by (58), for  $d$  large enough  $\ell^{10}V(x_i)t^2/(2d) \leq 1$  for all  $i \in \{2, \dots, d\}$ . Thus, by [3, Eq. 26.5], we have for such large  $d$ , all  $i \in \{2, \dots, d\}$  and all  $\delta > 0$ :

$$\begin{aligned} \left| \varphi_1(x_i, t/\sqrt{d}) - \left(1 - \frac{\ell^{10}V(x_i)t^2}{2d}\right) \right| &\leq \mathbb{E} \left[ \left( \frac{|t|^3 \ell^{15}}{6d^{3/2}} |C_5^{\text{fM}}(x_i, \xi_i)|^3 \right) \wedge \left( \frac{t^2 \ell^{10}}{d} C_5^{\text{fM}}(x_i, \xi_i)^2 \right) \right] \\ &\leq \mathbb{E} \left[ \frac{|t|^3 \ell^{15}}{6d^{3/2}} |C_5^{\text{fM}}(x_i, \xi_i)|^3 \mathbf{1}_{\{|C_5^{\text{fM}}(x_i, \xi_i)| \leq \delta d^{1/2}\}} \right] \\ &\quad + \mathbb{E} \left[ \frac{t^2 \ell^{10}}{d} C_5^{\text{fM}}(x_i, \xi_i)^2 \mathbf{1}_{\{|C_5^{\text{fM}}(x_i, \xi_i)| > \delta d^{1/2}\}} \right] \\ &\leq \frac{\delta |t|^3 \ell^{15}}{6d} \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^2 \right] + \frac{\ell^{10}t^2}{\delta^2 d^2} \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^4 \right], \end{aligned}$$

In addition, by [3, Lemma 1, Section 27] and using this result we get:

$$\begin{aligned} \left| \varphi_d(x, t) - \prod_{i=2}^d \left(1 - \frac{\ell^{10}V(x_i)t^2}{2d}\right) \right| &\leq \sum_{i=2}^d \frac{\delta |t|^3 \ell^{15}}{6d} \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^2 \right] \\ &\quad + \frac{\ell^{10}t^2}{\delta^2 d^2} \mathbb{E} \left[ C_5^{\text{fM}}(x_i, \xi_i)^4 \right] \\ &\leq \left( \mathbb{E} \left[ C_5^{\text{fM}}(X_1, \xi_1)^2 \right] + d^{-1/4} \right) \ell^{15} \delta |t|^3 / 6 \\ &\quad + \left( \mathbb{E} \left[ C_5^{\text{fM}}(X_1, \xi_1)^4 \right] + d^{-1/4} \right) \ell^{10} t^2 / (\delta^2 d), \end{aligned}$$

where the last inequality follows from  $x \in F_d^3$  and (57). Let now  $\epsilon > 0$ , and choose  $\delta$  small enough such that the first term is smaller than  $\epsilon/2$ . Then there exists  $d_0 \in \mathbb{N}^*$  such that for all  $d \geq d_0$ , the second term is smaller than  $\epsilon/2$  as well. Therefore, for  $d \geq d_0$  we get

$$\sup_{x \in F_d^3} \left| \varphi_d(x, t) - \prod_{i=2}^d \left(1 - \frac{\ell^{10}V(x_i)t^2}{2d}\right) \right| \leq \epsilon.$$

Consider now the second term of (59), by the triangle inequality,

$$\begin{aligned} \left| \prod_{i=2}^d \left(1 - \frac{\ell^{10}V(x_i)t^2}{2d}\right) - e^{-\ell^{10}(K^{\text{fM}})^2 t^2 / 2} \right| &\leq \left| \prod_{i=2}^d \left(1 - \frac{\ell^{10}V(x_i)t^2}{2d}\right) - \prod_{i=2}^d e^{-\ell^{10}V(x_i)t^2 / (2d)} \right| \\ &\quad + \left| \prod_{i=2}^d e^{-\ell^{10}V(x_i)t^2 / (2d)} - e^{-\ell^{10}(K^{\text{fM}})^2 t^2 / 2} \right|. \quad (60) \end{aligned}$$

We deal with the two terms separately. First since for all  $x_i$ ,  $V(x_i) \geq 0$ , we have

$$\left| 1 - V(x_i) \ell^{10} t^2 / (2d) - e^{-V(x_i) \ell^{10} t^2 / (2d)} \right| \leq V(x_i)^2 \ell^{20} t^4 / (8d^2).$$

Using this result, [3, Lemma 1, Section 27] and the Cauchy-Schwarz inequality, it follows:

$$\begin{aligned}
& \left| \prod_{i=2}^d \left( 1 - \frac{\ell^{10} \mathbf{V}(x_i) t^2}{2d} \right) - \prod_{i=2}^d e^{-\ell^{10} \mathbf{V}(x_i) t^2 / (2d)} \right| \\
& \leq \sum_{i=2}^d \left| 1 - \mathbf{V}(x_i) \ell^{10} t^2 / (2d) - e^{-\mathbf{V}(x_i) \ell^{10} t^2 / (2d)} \right| \\
& \leq \sum_{i=2}^d \mathbf{V}(x_i)^2 \ell^{20} t^4 / (8d^2) \leq \left( \mathbb{E} \left[ C_5^{\text{fM}}(\mathbf{X}_1, \xi_1)^4 \right] + d^{-1/4} \right) \ell^{20} t^4 / (8d), \tag{61}
\end{aligned}$$

where the last inequality is implied by (57). Finally since on  $\mathbb{R}_-$ ,  $u \mapsto e^u$  is 1-Lipschitz and using (57), we get

$$\begin{aligned}
\left| \prod_{i=2}^d e^{-\ell^{10} \mathbf{V}(x_i) t^2 / (2d)} - e^{-\ell^{10} (K^{\text{fM}})^2 t^2 / 2} \right| & \leq (t^2 \ell^{10} / 2) \left| \sum_{i=2}^d d^{-1} \mathbf{V}(x_i) - (K^{\text{fM}})^2 \right| \\
& \leq t^2 \ell^{10} d^{-1/4} / 2. \tag{62}
\end{aligned}$$

Therefore, combining (61) and (62) in (60), we get:

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d^3} \left| \prod_{i=2}^d \left( 1 - \frac{\ell^{10} \mathbf{V}(x_i) t^2}{2d} \right) - e^{-\ell^{10} (K^{\text{fM}})^2 t^2 / 2} \right| = 0,$$

which concludes the proof of (i).

(ii) follows now from (i) by the continuity theorem applied to an appropriate sequence  $\{x^d, d \in \mathbb{N}^*\}$ .  $\square$

*proof of Theorem 3.1.* The theorem follows from Lemma A.1, (55) in Theorem A.2 and the last statement in Lemma A.4.  $\square$

*proof of Theorem 3.2.* Consider  $F_d = \bigcap_{j=1,2,3} F_d^j$ , where the sets  $F_d^j$  are given resp. in Lemma A.1 Theorem A.2 and Lemma A.4. We then obtain  $\lim_{d \rightarrow +\infty} d^{-1/5} \pi_d((F_d)^c) = 0$  and by the union bound, for all  $T \geq 0$ ,

$$\lim_{d \rightarrow +\infty} \mathbb{P} \left[ \Gamma_s^{d, \text{fM}} \in F_d, \forall s \in [0, T] \right] = 1.$$

Furthermore, combining the former results with Lemma A.3, we have for all  $\psi \in C_c^\infty(\mathbb{R})$  (seen as a function of the first component):

$$\lim_{d \rightarrow +\infty} \sup_{x \in F_d} \left| A_d^{\text{fM}} \psi(x) - A^{\text{fM}} \psi(x) \right| = 0.$$

Then, the weak convergence follows from [5, Corollary 8.7, Chapter 4].  $\square$

## B Postponed proofs

### B.1 Proof of Lemma 4.1

By Assumption 3-4,  $\pi$  and  $q$  are positive and continuous. It follows from [14, Lemma 1.2] that  $P$  is  $\text{Leb}^d$ -irreducible aperiodic, where  $\text{Leb}^d$  is the Lebesgue measure on  $\mathbb{R}^d$ . In addition, all compact set  $\mathcal{C}$  such that  $\text{Leb}^d(\mathcal{C}) > 0$  are small for  $P$ . Now by [15, Theorem 15.0.1], we just need to check the drift condition (18). But by a simple calculation, using  $\alpha(x, y) \leq 1$  for all  $x, y \in \mathbb{R}^d$ , and the Cauchy-Schwarz inequality, we get

$$\begin{aligned} PV(x) &\leq 1 + \|x\|^2 + (\|\mu(x)\|^2 - \|x\|^2) \int_{\mathbb{R}^d} \alpha(x, y)q(x, y)dy \\ &\quad + (2\pi)^{-d/2} (2\|\mu(x)\|\|S(x)\| + \|S(x)\|^2) \int_{\mathbb{R}^d} \max(\|\xi\|^2, 1) e^{-\|\xi\|^2/2} d\xi. \end{aligned}$$

By (21),  $\limsup_{\|x\| \rightarrow +\infty} (2\|\mu(x)\|\|S(x)\| + \|S(x)\|^2) \|x\|^{-2} = 0$ . Therefore, using again the first inequality of (21) and Assumption 5:

$$\limsup_{\|x\| \rightarrow +\infty} PV(x)/V(x) \leq 1 - (1 - \tau^2) \liminf_{\|x\| \rightarrow +\infty} \int_{\mathbb{R}^d} \alpha(x, y)q(x, y)dy < 1.$$

This concludes the proof of Lemma 4.1.  $\square$

### B.2 Proof of Theorem 4.2

We prove this result by contradiction. The strategy of the proof is the following: first, under our assumptions, most of the proposed moves by the algorithm has a norm which is greater than the current point. However, if  $P$  is geometrically ergodic, then it implies a upper bound on the rejection probability of the algorithm by some constant strictly smaller than 1. But combining these facts, we can exhibit a sequence of point  $\{x_n, n \in \mathbb{N}\}$ , such that  $\lim_{n \rightarrow +\infty} \pi(x_n) = +\infty$ . Since we assume that  $\pi$  is bounded, we have our contradiction.

If  $P$  is geometrically ergodic, then by [23, Theorem 5.1], there exists  $\eta > 0$  such that for almost every  $x \in \mathbb{R}^d$ ,

$$\int_{\mathbb{R}^d} \alpha(x, y)q(x, y)dy \geq \eta, \quad (63)$$

and let  $M \geq 0$  such that

$$\mathbb{P}[\|\xi\| \geq M] \leq \eta/2, \quad (64)$$

where  $\xi$  is a standard  $d$ -dimensional Gaussian random variable. By (22), there exists  $R_\epsilon, \delta > 0$  such that

$$\inf_{\{\|x\| \geq R_\epsilon\}} \|S(x)^{-1}\mu(x)\| \|x\|^{-1} \geq \epsilon^{-1} + \delta \quad \text{and} \quad \inf_{\{\|x\| \geq R_\epsilon\}} \inf_{\|y\|=1} \|S(x)y\| \geq \epsilon. \quad (65)$$

Note that we can assume  $R_\epsilon$  is large enough so that

$$\epsilon\delta R_\epsilon \geq M. \quad (66)$$

Now define for  $x \in \mathbb{R}^d, \|x\| \geq R_\epsilon$

$$B(x) = \left\{ y \in \mathbb{R}^d \mid \|S(x)^{-1}(y - \mu(x))\| \leq M \right\}. \quad (67)$$

Note if  $y \in B(x)$ , we have by definition and the triangle inequality  $\|S(x)^{-1}y\| \geq \|S(x)^{-1}\mu(x)\| - M$ . Therefore by (65) and (66)

$$\|y\| = \|S(x)S(x)^{-1}y\| \geq \epsilon \|S(x)^{-1}y\| \geq \epsilon(\epsilon^{-1} + \delta) \|x\| - M \geq \|x\|. \quad (68)$$

We then show that this inequality implies

$$\liminf_{\|x\| \rightarrow +\infty} \inf_{y \in B(x)} \frac{q(y, x)}{q(x, y)} = 0. \quad (69)$$

First, it is straightforward by (67), that the denominator is uniformly bounded from 0. As for the numerator, first we can notice by (65)-(68), we have  $\|y\| \geq R_\epsilon$  and for all  $z \in \mathbb{R}^d$ ,  $\|S(y)z\| \geq \epsilon \|z\|$ , which implies for all  $z \in \mathbb{R}^d$ ,  $\epsilon^{-1} \|z\| \geq \|S(y)^{-1}z\|$ . Using this result, the triangle inequality and (65)-(68) again, we get

$$\begin{aligned} q(y, x) &= (2\pi)^{-d/2} \exp \left( -\|S(y)^{-1}(x - \mu(y))\|^2 - \log(|S(y)(S(y))^T|) \right) \\ &\leq (2\pi)^{-d/2} \exp \left( -(\|S(y)^{-1}\mu(y)\| - \|S(y)^{-1}x\|)^2 - \log(|S(y)(S(y))^T|) \right) \\ &\leq (2\pi)^{-d/2} \exp \left( -((\epsilon^{-1} + \delta) \|y\| - \epsilon^{-1} \|y\|)^2 - \log(|S(y)(S(y))^T|) \right) \\ &\leq (2\pi)^{-d/2} \exp \left( -\delta^2 \|y\|^2 - \log(|S(y)(S(y))^T|) \right). \end{aligned}$$

Using this inequality and (23) imply  $\lim_{\|x\| \rightarrow +\infty} \inf_{y \in B(x)} q(y, x) = 0$  and then (69). Therefore there exists  $R_q \geq 0$  such that for all  $x \in \mathbb{R}^d$ ,  $\|x\| \geq R_q$

$$\inf_{y \in B(x)} \frac{q(y, x)}{q(x, y)} \leq \eta/4. \quad (70)$$

Now we are able to build the sequence  $\{x_n, n \in \mathbb{N}\}$  such that for all  $n \in \mathbb{N}$ ,  $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$  and  $\lim_{n \rightarrow +\infty} \pi(x_n) = +\infty$ . Indeed let  $x_0 \in \mathbb{R}^d$  such that  $\|x_0\| \geq \max(R_\epsilon, R_q)$ . Assume, we have built the sequence up to the  $n$ th term and such that for all  $k = 0, \dots, n-1$ ,  $\|x_{k+1}\| \geq \max(R_\epsilon, R_q)$  and  $\pi(x_{k+1}) \geq (3/2)\pi(x_k)$ . Now we choose  $x_{n+1}$  depending on  $x_n$ , satisfying  $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$  and  $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$ . Since  $\|x_n\| \geq \max(R_\epsilon, R_q)$ , by (63)-(64) and (70)

$$\begin{aligned} \eta &\leq \int_{\mathbb{R}^d} \alpha(x_n, y) q(x_n, y) dy \leq \eta/2 + \int_{B(x_n)} \min \left( 1, \frac{\pi(y)q(y, x_n)}{\pi(x_n)q(x_n, y)} \right) q(x_n, y) dy \\ &\leq \eta/2 + (\eta/4) \int_{B(x_n)} \frac{\pi(y)}{\pi(x_n)} q(x_n, y) dy. \end{aligned}$$

This inequality implies that  $\int_{B(x_n)} \frac{\pi(y)}{\pi(x_n)} q(x_n, y) dy \geq 2$  and therefore there exists  $x_{n+1} \in B(x_n)$  such that  $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$ , and since  $x_{n+1} \in B(x_n)$  by (68),  $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$ . Therefore, we have a sequence  $\{x_n, n \in \mathbb{N}\}$  such that for all  $n \in \mathbb{N}$ ,  $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$ . Since by assumption  $\pi(x_0) > 0$ , we get  $\lim_{n \rightarrow +\infty} \pi(x_n) = +\infty$ , which contradicts the assumption that  $\pi$  is bounded. This concludes the proof of Theorem 4.2.  $\square$

## C Expressions of $C_5^\bullet(x_1, \xi_1)$

$$C_5^{\text{fM}}(x_1, \xi_1) = \frac{\ell^5}{720} \left( \xi_1^5 g^{(5)}(x_1) + 5\xi_1^3 g^{(5)}(x_1) + 15\xi_1^3 g^{(4)}(x_1)g'(x_1) \right. \\ \left. + 15\xi_1 g^{(4)}(x_1)g'(x_1) + 30\xi_1^3 g^{(3)}(x_1)g''(x_1) \right. \\ \left. + 10\xi_1 g^{(3)}(x_1)g''(x_1) + 30\xi_1 g^{(3)}(x_1)g'(x_1)^2 + 35\xi_1 g'(x_1)g''(x_1)^2 \right)$$

$$C_5^{\text{mO}}(x_1, \xi_1) = \ell^5 \left( \frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1) \right. \\ \left. + \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{29}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) - \frac{7}{48}\xi_1 g^{(3)}(x_1)g''(x_1) \right. \\ \left. + \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 + \frac{1}{6}\xi_1 g'(x_1)g''(x_1)^2 \right) .$$

$$C_5^{\text{bO}}(x_1, \xi_1) = \ell^5 \left( \frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1) \right. \\ \left. + \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{29}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) - \frac{19}{144}\xi_1 g^{(3)}(x_1)g''(x_1) \right. \\ \left. + \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 + \frac{1}{6}\xi_1 g'(x_1)g''(x_1)^2 \right) .$$

$$C_5^{\text{gbO}}(x_1, \xi_1) = \ell^5 \left( \frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1) \right. \\ \left. + \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{1}{72}a_3 \xi_1 g^{(3)}(x_1)g''(x_1) + \frac{1}{6}a_4^2 \xi_1^3 g^{(3)}(x_1)g''(x_1) \right. \\ \left. - \frac{1}{6}a_4^2 \xi_1 g^{(3)}(x_1)g''(x_1) + \frac{5}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) + \frac{1}{48}\xi_1 g^{(3)}(x_1)g''(x_1) \right. \\ \left. + \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 - \frac{1}{24}a_1^2 \xi_1 g'(x_1)g''(x_1)^2 + \frac{1}{6}a_4^2 \xi_1 g'(x_1)g''(x_1)^2 \right. \\ \left. + \frac{1}{24}\xi_1 g'(x_1)g''(x_1)^2 \right) .$$

## D Expressions of $K^\bullet$

We provide here the expressions of the quantities  $K^\bullet$  involved in Theorems 3.1, 3.2, 3.3, 3.4. Let  $X$  be a random variable distributed according to  $\pi_1$ .

$$K^{\text{fM}} = \mathbb{E} \left[ \frac{79g^{(5)}(X)^2}{17280} + \frac{11g^{(4)}(X)^2 g'(X)^2}{1152} + \frac{77g^{(3)}(X)^2 g''(X)^2}{2592} + \frac{1}{576}g^{(3)}(X)^2 g'(X)^4 \right. \\ \left. + \frac{49g'(X)^2 g''(X)^4}{20736} + \frac{7}{576}g^{(4)}(X)g^{(5)}(X)g'(X) + \frac{19}{864}g^{(3)}(X)g^{(5)}(X)g''(X) \right]$$

$$\begin{aligned}
& + \frac{1}{288}g^{(3)}(X)g^{(5)}(X)g'(X)^2 + \frac{7g^{(5)}(X)g'(X)g''(X)^2}{1728} \\
& + \frac{1}{144}g^{(3)}(X)g^{(4)}(X)g'(X)^3 + \frac{7}{864}g^{(4)}(X)g'(X)^2g''(X)^2 + \frac{7g^{(3)}(X)g'(X)^3g''(X)^2}{1728} \\
& + \frac{5}{432}g^{(3)}(X)^2g'(X)^2g''(X) + \frac{35g^{(3)}(X)g'(X)g''(X)^3}{2592} + \frac{29}{864}g^{(3)}(X)g^{(4)}(X)g'(X)g''(X) \Big].
\end{aligned}$$

$$\begin{aligned}
K^{\text{mO}} = \mathbb{E} & \left[ \frac{79g^{(5)}(X)^2}{17280} + \frac{11g^{(4)}(X)^2g'(X)^2}{1152} + \frac{1567g^{(3)}(X)^2g''(X)^2}{3456} \right. \\
& + \frac{1}{576}g^{(3)}(X)^2g'(X)^4 + \frac{1}{36}g'(X)^2g''(X)^4 + \frac{7}{576}g^{(4)}(X)g^{(5)}(X)g'(X) \\
& + \frac{17}{192}g^{(3)}(X)g^{(5)}(X)g''(X) + \frac{1}{288}g^{(3)}(X)g^{(5)}(X)g'(X)^2 \\
& + \frac{1}{72}g^{(5)}(X)g'(X)g''(X)^2 + \frac{1}{144}g^{(3)}(X)g^{(4)}(X)g'(X)^3 + \\
& \frac{1}{36}g^{(4)}(X)g'(X)^2g''(X)^2 + \frac{1}{72}g^{(3)}(X)g'(X)^3g''(X)^2 \\
& + \frac{11}{288}g^{(3)}(X)^2g'(X)^2g''(X) + \frac{11}{72}g^{(3)}(X)g'(X)g''(X)^3 + \\
& \left. \frac{73}{576}g^{(3)}(X)g^{(4)}(X)g'(X)g''(X) \right].
\end{aligned}$$

$$\begin{aligned}
K^{\text{gbO}} = \mathbb{E} & \left[ \frac{1}{36}g'(X)^2g''(X)^4a_4^4 + \frac{5}{18}g''(X)^2g^{(3)}(X)^2a_4^4 \right. \\
& + \frac{1}{9}g'(X)g''(X)^3g^{(3)}(X)a_4^4 - \frac{1}{72}a_1^2g'(X)^2g''(X)^4a_4^2 + \frac{1}{72}g'(X)^2g''(X)^4a_4^2 \\
& + \frac{11}{72}g''(X)^2g^{(3)}(X)^2a_4^2 + \frac{1}{108}a_3g''(X)^2g^{(3)}(X)^2a_4^2 \\
& + \frac{1}{36}g'(X)^2g''(X)g^{(3)}(X)^2a_4^2 - \frac{1}{36}a_1^2g'(X)g''(X)^3g^{(3)}(X)a_4^2 \\
& + \frac{5}{72}g'(X)g''(X)^3g^{(3)}(X)a_4^2 + \frac{1}{216}a_3g'(X)g''(X)^3g^{(3)}(X)a_4^2 \\
& + \frac{1}{72}g'(X)^3g''(X)^2g^{(3)}(X)a_4^2 + \frac{1}{36}g'(X)^2g''(X)^2g^{(4)}(X)a_4^2 \\
& + \frac{7}{72}g'(X)g''(X)g^{(3)}(X)g^{(4)}(X)a_4^2 + \frac{1}{72}g'(X)g''(X)^2g^{(5)}(X)a_4^2 \\
& + \frac{5}{72}g''(X)g^{(3)}(X)g^{(5)}(X)a_4^2 + \frac{1}{576}a_1^4g'(X)^2g''(X)^4 \\
& - \frac{1}{288}a_1^2g'(X)^2g''(X)^4 + \frac{1}{576}g'(X)^2g''(X)^4 + \frac{1}{576}g'(X)^4g^{(3)}(X)^2 \\
& + \frac{a_3^2g''(X)^2g^{(3)}(X)^2}{5184} + \frac{1}{288}a_3g''(X)^2g^{(3)}(X)^2 \\
& + \frac{79g''(X)^2g^{(3)}(X)^2}{3456} + \frac{1}{96}g'(X)^2g''(X)g^{(3)}(X)^2 \\
& \left. + \frac{1}{864}a_3g'(X)^2g''(X)g^{(3)}(X)^2 + \frac{11g'(X)^2g^{(4)}(X)^2}{1152} \right]
\end{aligned}$$

$$\begin{aligned}
& + \frac{79g^{(5)}(X)^2}{17280} - \frac{1}{96}a_1^2g'(X)g''(X)^3g^{(3)}(X) \\
& + \frac{1}{96}g'(X)g''(X)^3g^{(3)}(X) - \frac{1}{864}a_1^2a_3g'(X)g''(X)^3g^{(3)}(X) \\
& + \frac{1}{864}a_3g'(X)g''(X)^3g^{(3)}(X) - \frac{1}{288}a_1^2g'(X)^3g''(X)^2g^{(3)}(X) \\
& + \frac{1}{288}g'(X)^3g''(X)^2g^{(3)}(X) - \frac{1}{144}a_1^2g'(X)^2g''(X)^2g^{(4)}(X) \\
& + \frac{1}{144}g'(X)^2g''(X)^2g^{(4)}(X) + \frac{1}{144}g'(X)^3g^{(3)}(X)g^{(4)}(X) \\
& + \frac{17}{576}g'(X)g''(X)g^{(3)}(X)g^{(4)}(X) + \frac{1}{432}a_3g'(X)g''(X)g^{(3)}(X)g^{(4)}(X) \\
& - \frac{1}{288}a_1^2g'(X)g''(X)^2g^{(5)}(X) + \frac{1}{288}g'(X)g''(X)^2g^{(5)}(X) + \\
& \frac{1}{288}g'(X)^2g^{(3)}(X)g^{(5)}(X) + \frac{11}{576}g''(X)g^{(3)}(X)g^{(5)}(X) \\
& + \frac{1}{864}a_3g''(X)g^{(3)}(X)g^{(5)}(X) + \frac{7}{576}g'(X)g^{(4)}(X)g^{(5)}(X) \Big].
\end{aligned}$$