Fast Langevin based algorithm for MCMC in high dimensions

Alain Durmus[1], Gareth O. Roberts[2]*, Gilles Vilmart[3]†, and Konstantinos C. Zygalakis[4]‡

November 25, 2016

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: weak convergence, Markov Chain Monte Carlo, diffusion limit, exponential ergodicity.

AMS subject classification (2010): 60F05, 65C05

1 Introduction

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

$$\mathrm{d}x_t = \frac{1}{2}\Sigma\nabla\log\pi(x_t)\mathrm{d}t + \Sigma^{1/2}\mathrm{d}W_t \,, \tag{1}$$

where $\{W_t,\ t \geq 0\}$ is a standard d-dimensional Brownian motion, and Σ is a given positive definite symmetric matrix. Under appropriate assumptions [10] on π, it can be shown that the dynamic generated by (1) is ergodic with unique invariant distribution π. This is a key property of (1) and taking advantage of it permits to sample from the invariant distribution π. In particular, if one could solve (1) analytically and then take time t to infinity then it

would be possible to generate samples from π. However, there exists a limited number of cases [13] where such an analytical formula exists. A standard approach is to discretise (1) using a one step integrator. The drawback of this approach is that it introduces a bias, because in general π is not invariant with respect to the Markov chain defined by the discretization, [26, 15, 1]. In addition, the discretization might fail to be ergodic [24], even though (1) is geometrically ergodic.

An alternative way of sampling from π, which does not face the bias issue introduced by discretizing (1), is by using the Metropolis-Hastings algorithm [11]. 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 candidate y_{j+1} is generated from a proposal density $q(x_j, \cdot)$. This candidate is then accepted $(x_{j+1} = y_{j+1})$ with probability $\alpha(x_j, y_{j+1})$ given by

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

and rejected $(x_{j+1} = x_j)$ otherwise. The resulting Markov chain $\{x_j, \ j \in \mathbb{N}\}$ is reversible with respect to π and under mild assumptions is ergodic [14, 19].

The simplest proposals are random walks for which q is the transition kernel associated with the proposal

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

where ξ 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 does not use information about π to construct appropriate candidate moves.

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

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

where ξ 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 better convergence properties than the RWM algorithm in general. This method directs the proposed moves towards areas of high probability for the distribution π, using the gradient of $\log\pi$. There is now a growing literature on gradient-based MCMC algorithms, as exemplified through the two papers [8, 5] and the references therein. We also mention here function space MCMC methods [5]. Assuming that the target measure has a density w.r.t. a Gaussian measure on a Hilbert space, these algorithms are defined in infinite dimension and avoid completely the dependence on the dimension d faced by standard MCMC algorithms.

A natural question is if one can improve on the behaviour of MALA by incorporating more information about the properties of π 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 [8]. Although this turns out to not be sufficient, we shall show that, by slightly modifying this approach and not focusing on the weak order itself, we are able to construct a new proposal with better convergence properties than MALA. We mention that an analogous proposal is presented independently in [7] in a different context to improve the strong order of convergence of MALA.

Thus our main contribution in this paper is the introduction and theoretical analysis of the fMALA algorithm (*fast* MALA), and its cousins which will be introduced in Section 3.

These algorithms provide for the first time, implementable gradient-based MCMC algorithms which can achieve convergence in $\mathcal{O}(d^{1/5})$ iterations, thus improving on the $\mathcal{O}(d^{1/3})$ of MALA and many related methods. These results are demonstrated as a result of high-dimensional diffusion approximation results. As well as giving these order of magnitude results for high-dimensional problems, we shall also give stochastic stability results, specifically results about the geometric ergodicity of the algorithms we introduce under appropriate regularity conditions.

Whilst the algorithms we describe have clear practical relevance for MCMC use, it is important to recognise the limitations of this initial study of these methodologies, and we shall note and comment on two which are particularly important. In order to obtain the diffusion limit results we give, it is necessary to make strong assumptions about the structure of the sequence of target distributions as d increases. In our analysis we assume that the target distribution consists of d i.i.d. components as in the initial studies of both high-dimensional RWM and MALA algorithms [20, 21]. Those analyses were subsequently extended (see for example [22]) and supported by considerable empirical evidence from applied MCMC use. We also expect that in the context of this paper, our conclusions should provide practical guidance for MCMC practitioners well beyond the cases where rigorous results can be demonstrated, and we provide an example to illustrate this in Section 5.

Secondly, our diffusion limit results depend on the initial distribution of the Markov chain being the target distribution π, clearly impractical in real MCMC contexts. The works [4, 12] study the case of MCMC algorithms (specifically RWM and MALA algorithms) started away from stationarity. On the one hand, it turns out that MALA algorithms are less robust than RWM when starting at under-dispersed values in that scaling strategies. Indeed, optimising mixing in stationarity can be highly suboptimal in the transient phase, often with initial moves having exponentially small acceptance probabilities (in d). On the other hand, a slightly more conservative strategy for MALA still achieves $\mathcal{O}(d^{1/2})$ compared to $\mathcal{O}(d)$ for RWM. It is natural to expect the story for fMALA to be at least as involved as that for MALA, and we give some empirical evidence to support this in the simulations study of Section 5. Future work will underpin these investigations with theoretical results analogous to those of [4, 12]. From a practical MCMC perspective however, it should be noted that strategies which mix MALA-transient optimal scaling with fMALA-stationary optimal scaling will perform in a robust manner, both in the transient and stationary phases. Two of these effective strategies are illustrated in Section 5.

The paper is organised as follows. In Section 2 we provide a heuristic for the choice of the parameter 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 fMALA and its variants, and prove our main results for the introduced methods. Section 4 investigates the ergodic properties of the different proposals for a wide variety of target densities π. 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 convergence of MCMC algorithms. In particular, in Section 2.1 we discuss some issues related to the computational complexity of MCMC methods in high dimensions, while in Section 2.2 we present a useful heuristic for understanding the optimal scaling 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 parameter h in all proposals mentioned above as the dimension of the space d goes to infinity. In particular, we choose h proportional to an inverse power of the dimension d such that

$$h \propto d^{-\gamma} . \tag{5}$$

This implies that the proposal y is now a function of: (i) the current state x; (ii) the parameter γ through the scaling above; and (iii) the random variable ξ which appears in all the considered proposals. Thus $y = y(x, \xi; \gamma)$. Ideally γ should be as small as possible so the chain makes large steps and samples are correlated as little as possible. At the same time, the acceptance probability should not degenerate to 0 as $d \to \infty$, also to prevent high correlation amongst samples. This naturally leads to the definition of a critical exponent γ_0 given by

$$\gamma_0 = \inf_{\gamma_c \geq 0} \left\{ \gamma_c : \liminf_{d \to \infty} \mathbb{E}\left[\alpha(x, y)\right] > 0 , \quad \forall \gamma \in [\gamma_c, \infty) \right\} . \tag{6}$$

The expectation here is with respect to x distributed according to π and y chosen from the proposal distribution. In other words, we take the largest possible value for h, as function 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.

If h is of the form (5), with $\gamma \geq \gamma_0$, the acceptance probability does not degenerate, and the Markov chain arising from the Metropolis-Hastings method can be thought of as an approximation of the Langevin SDE (1). This Markov chain travels with time-steps h on the paths of this SDE, and therefore requires a minimal number of steps to reach timescales of $\mathcal{O}(1)$ given by

$$M(d) = d^{\gamma_0} . \tag{7}$$

If it takes $\mathcal{O}(1)$ for the limiting SDE to reach stationarity, then we obtain that $M(d)$ gives the computational complexity of the algorithm.[1]

If we now consider the case of a product measure where

$$\pi(x) = \pi_d(x) = Z_d \prod_{i=1}^{d} e^{g(x_i)} , \tag{8}$$

and Z_d is the normalizing constant, then it is well known [20] that for the RWM it holds $\gamma_0 = 1$, while for MALA it holds $\gamma_0 = 1/3$ [21]. 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 a better convergence scaling.

---

[1]In this definition of the cost one does not take into account the cost of generating a proposal. This is discussed in Remark 2.3.

2.2 Formal derivation

Here we explain the main idea that is used for proving the scaling of a Gaussian[2] proposal in high dimensions. In particular, the proposal y is now of the form

$$y = \mu(x,h) + S(x,h)\xi \,, \tag{9}$$

where $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$ is a standard d dimensional Gaussian random variable. Note that in the case of the 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 acceptance probability 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 (9). Now using the fact that y is related to x according to (9), $R_d(x,x) = 0$, together with appropriate smoothness properties on the function $g(x)$, one can expand R_d in powers of $\sqrt{h}$ using a Taylor expansion

$$R_d(x,y) = \sum_{i=1}^{k}\sum_{j=1}^{d} h^{i/2}C_{ij}(x,\xi) + h^{(k+1)/2}L_{k+1}(x,h^*,\xi) \,. \tag{10}$$

It turns out [2] that the scaling associated with each proposal relates directly with how many of the C_{ij} terms are zero in (10). This simplifies if we further assume that $\Sigma = \mathrm{I}_d$ in (1) and that π satisfies (8), because we get for all $i \in \{1, \cdots, k\}$, $j \in \{1, \cdots, j\}$, $C_{ij}(x,\xi) = C_i(x_j, \xi_j)$ and (10) can be written as

$$R_d(x,y) = \sum_{i=1}^{k}\sum_{j=1}^{d} \frac{\sqrt{h^i d}}{\sqrt{d}}C_i(x_j,\xi_j) + h^{(k+1)/2}L_{k+1}(x,h^*,\xi) \,. \tag{11}$$

We then see that if $C_i = 0$, for $i = 1, \cdots, m$, then this implies that $\gamma_0 = 1/(m+1)$. Indeed, this value of γ_0 yields $h^{m+1}d = 1$ and the leading order term in (10) 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 for known proposals, and the above expression is frequently approximated by a central limit theorem. The second dominant term in (10) 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]We point out that Gaussianity here is not necessary but it greatly simplifies the calculations.

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

We introduce new Gaussian proposals for which $\gamma_0 = 1/5$ in (7). We start by presenting the simplest method, and then give two variations of it, motivated by the desire to obtain robust and stable ergodic properties (geometric ergodicity). The underlying calculations that show $C_i = 0, i = 1, \ldots, m$ 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 $D^2 f$ 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

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

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

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

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

Remark 2.1. *Since by assumption Σ is positive definite, notice that the Jacobian matrix $Df(x)$ is diagonalizable for all $x \in \mathbb{R}^d$. Indeed, it is similar to the symmetric matrix $\Sigma^{-1/2} Df(x) \Sigma^{1/2} = \Sigma^{1/2} D^2 \log \pi(x) \Sigma^{1/2}$, and we use that a symmetric matrix is always diagonalizable. This will permit us to define analytic functionals of $Df(x)$.*

2.3.1 Fast Metropolis-Adjusted Langevin Algorithm (fMALA)

We first give a natural proposal for which $\gamma_0 = 1/5$ based on the discussion of Section 2.2. We restrict the class of proposal defined by (9) by setting for all $x \in \mathbb{R}^d$ and $h > 0$,

$$\mu(x, h) = x + h\,\mu_1(x) + h^2 \mu_2(x) , \qquad S(x, h) = h^{1/2} S_1(x) + h^{3/2} S_2(x) .$$

By a formal calculation (see the supplementary materials), explicit expressions for the functions μ_1, μ_2, S_1, S_2 have to be imposed for the four first term $C_i(x, \xi)$, $i \in \{1, 2, 3, 4\}$, in (11) to be zero. This result implies the following definition for μ and S:

$$\mu^{\text{fM}}(x, h) = x + \frac{h}{2} f(x) - \frac{h^2}{24}\left(Df(x)f(x) + \{\Sigma : D^2 f(x)\}\right), \tag{12a}$$

$$S^{\text{fM}}(x, h) = \left(h^{1/2} I_d + (h^{3/2}/12) Df(x)\right) \Sigma^{1/2} . \tag{12b}$$

We will refer to (9) when μ, S are given by (12) 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 2.2. *It is interesting to note that 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 [8] about the potential use of higher order integrators for the Langevin equation within the Metropolis-Hastings framework.*

Remark 2.3. *The proposal given by equation* (12) *contains higher order derivatives of the vector field $f(x)$, resulting in higher computational cost than the standard MALA proposal. This additional cost might offset the benefits of the improved scaling, since the corresponding Jacobian and Hessian can be full matrices in general. However, there exist cases of interest[3] where due to the structure of the Jacobian and Hessian the computational cost of the fMALA proposal is of the same order with respect to the dimension d as for the MALA proposal. Furthermore, we note that one possible way to avoid derivatives is by using finite differences or Runge-Kutta type approximations of the proposal* (12). *This, however, is out of the scope of the present paper.*

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 targets π [24]. This issue was addressed in [23] where a modification of MALA based on the Ozaki discretization [18] of (1) was proposed and studied. In the same spirit as in [23] we propose here a modification of fMALA, defined by

$$\mu^{\mathrm{mO}}(x,h) = x + \mathscr{T}_1(Df(x),h,1)f(x) - (h^2/6)Df(x)f(x) \tag{13a}$$
$$- (h^2/24)\{\Sigma : D^2 f(x)\}$$

$$S^{\mathrm{mO}}(x,h) = \left(\mathscr{T}_1(Df(x),2h,1) - (h^2/3)Df(x)\right)^{1/2}\Sigma^{1/2} . \tag{13b}$$

where

$$\mathscr{T}_1(\mathrm{M},h,a) = (a\mathrm{M})^{-1}(e^{(ah/2)\mathrm{M}} - \mathrm{I}_d) \tag{14}$$

for all[4] $\mathrm{M} \in \mathbb{R}^{d\times d}$, $h > 0$, $a \in \mathbb{R}$.

The Markov chain defined by (13) 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 $t \mapsto (e^{ht}-1)/t - (1/3)h^2 t$ is positive on $\mathbb{R}$ for all $h > 0$. It then follows from Remark 2.1 that for all $x \in \mathbb{R}^d$, the matrix $\mathscr{T}_1(Df(x),2h,1) - (h^2/3)Df(x)$ is diagonalizable with non-negative eigenvalues, which permits to define its matrix square-root, and $S^{\mathrm{mO}}(x,h)$ is well defined for all $x \in \mathbb{R}^d$ and $h > 0$.

Remark 2.4. *In regions where $\|\Sigma\nabla\log\pi(x)\|$ is much greater than $\|x\|$, we need in practice to take h very small (of order $\|x\| / \|\Sigma\nabla\log\pi(x)\|$) for MALA to exit these regions. However such a choice of h depends on x and cannot be used directly. Such a value of h can therefore be hard to find theoretically as well as computationally. This issue can be tackled by multiplying $f = \Sigma\nabla\log\pi(x)$ by $\mathscr{T}_1(Df(x),h,a)$ in* (13a). *Indeed under some mild conditions, in that case, we can obtain an algorithm with good mixing properties for all $h > 0$; see [23, 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 2.4 we generalise here the approach in [23] to deal with the complexities arising to the presence of the $Df(x)f(x)$ term.

---

[3]We study one of those in Section 5.

[4]Notice that the matrix functionals in (14),(16),(17) remain valid if matrix aM is not invertible, using the appropriate power series for the matrix exponentials.

In particular we now define

$$
\begin{aligned}
\mu^{\mathrm{gbO}}(x,h) = x &+ \mathscr{T}_1(Df(x),h,a_1)f(x) \\
&- (1/3)\mathscr{T}_3(Df(x),h,a_3)\{\Sigma : D^2 f(x)\} \\
&+ ((a_1/2)+(1/6))\,\mathscr{T}_2(Df(x),h,a_2)f(x)\,, \\
S^{\mathrm{gbO}}(x,h) = (\mathscr{T}_1&(Df(x),2h,a_4) \\
&+ ((a_4/2)-(1/6))\,\mathscr{T}_2(Df(x),2h,a_5))^{1/2}\Sigma^{1/2}\,.
\end{aligned}
$$

(15a)

(15b)

where $a_i,\ i = 1,\cdots,5$ are positive parameters, $\mathscr{T}_1$ is given by (14) and

$$\mathscr{T}_2(\mathrm{M},h,a) = (a\mathrm{M})^{-1}(e^{-(ah^2/4)\mathrm{M}^2} - \mathrm{I}_d) \tag{16}$$

$$\mathscr{T}_3(\mathrm{M},h,a) = (a\mathrm{M})^{-2}(e^{(ah/2)\mathrm{M}} - \mathrm{I}_d - (ah/2)\mathrm{M}) \tag{17}$$

with $\mathrm{M} \in \mathbb{R}^{d\times d}$, $h > 0$, $a \in \mathbb{R}$ and I_d is the identity matrix. The Markov chain defined by (15) 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). Note that S^{gbO} in (15b) is not always well defined in general. However, using Remark 2.1, the following condition is sufficient to define S^{gbO} with the square-root of a diagonalizable matrix with non-negative eigenvalues.

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 $\mathbb{R}$.*

For $a_4 = a_5 = 1$, this assumption is satisfied, and choosing $a_i = 1$ for all $i = 1,\ldots,5$, (15) 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 the optimal scaling results for fMALA and gbOMA introduced in Section 2. We recall from the discussion in Section 2 that the parameter h depends on the dimension and is given as $h_d = \ell^2 d^{-1/5}$, with $\ell > 0$. Finally, we prove our results for the case of target distributions of the product form given by (8), we take $\Sigma = \mathrm{I}_d$, and 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, κ such that*

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

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

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

3.1 Optimal scaling of fMALA

The Markov chain produced by fMALA, with target density π_d and started at stationarity, will be denoted by $\{X_k^{d,\text{fM}}, \ k \in \mathbb{N}\}$. Let q_d^{fM} be the transition density associated with the proposal of fMALA relatively to π_d. In a similar manner, we denote by α_d^{fM} the acceptance probability. Now we introduce the jump process based 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

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

the mean under π_d of the acceptance rate.

Theorem 3.1. *Assume Assumption 2. Then*

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

where $a^{\text{fM}}(\ell) = 2\Phi(-K^{\text{fM}}\ell^5/2)$ with $\Phi(t) = (1/(2\pi)) \int_{-\infty}^{t} e^{-s^2/2} \mathrm{d}s$ and the expression of K^{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, $\{Y^{d,\text{fM}}, \ d \in \mathbb{N}^*\}$ converges weakly (in the Skorokhod topology), as $d \to \infty$, to the solution $\{Y_t^{\text{fM}}, \ t \in \mathbb{R}_+\}$ of the Langevin equation defined by:*

$$\mathrm{d}Y_t^{\text{fM}} = (\mathrm{h}^{\text{fM}}(\ell))^{(1/2)} \mathrm{d}B_t + (1/2)\mathrm{h}^{\text{fM}}(\ell) \nabla \log \pi_1(Y_t^{\text{fM}}) \mathrm{d}t \ , \tag{18}$$

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

Proof. The proof of these two theorems are in Appendix A. $\qquad\square$

Remark 3.3. *The above analysis shows that for fMALA, the optimal exponent defined in (6) is given by $\gamma_0 = 1/5$ as discussed in Section 2.2. Indeed, if h_d has the form $\ell^2 d^{-1/5+\epsilon}$, then an adaptation of the proof of Theorem 3.1 implies that for all $\ell > 0$, if $\epsilon \in (0, 1/5)$, $\lim_{d \to +\infty} a^{\text{fM}}(\ell) = 0$. In contrast, if $\epsilon < 0$ then $\lim_{d \to +\infty} a^{\text{fM}}(\ell) = 1$.*

3.2 Scaling results for gbOMA

As in the case of fMALA, we assume π_d is of the form (8) and we take $\Sigma = \mathrm{I}_d$, $h_d = \ell^2 d^{-1/5}$. The Metropolis-adjusted Markov chain based on gbOMA, with target density π_d and started at stationarity, is denoted by $\{X_k^{d,\text{gbO}}, \ k \in \mathbb{N}\}$. We will denote by q_d^{gbO} the transition density associated with the proposals defined by gbOMA with respect to π_d. In a similar manner, the acceptance probability relatively to π_d and gbOMA will be denoted by α_d^{gbO}. 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 $\Gamma_t^{d,\text{gbO}} = X_{J_t}^{d,\text{gbO}}$. Denote also by

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

the mean under π_d of the acceptance rate of the algorithm.

Theorem 3.4. *Assume Assumptions 1 and 2. Then*

$$\lim_{d \to +\infty} a_d^{\mathrm{gbO}}(\ell) = a^{\mathrm{gbO}}(\ell) \ ,$$

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

Theorem 3.5. *Assume Assumptions 1 and 2. Let $\{G_t^{d,\mathrm{gbO}} = \Gamma_{t,1}^{d,\mathrm{gbO}}, \ t \in \mathbb{R}_+\}$ be the process corresponding to the first component of $\Gamma^{d,\mathrm{gbO}}$. Then, $\{G^{d,\mathrm{gbO}}, \ d \in \mathbb{N}^*\}$ converges weakly (in the Skorokhod topology) to the solution $\{G_t^{\mathrm{gbO}}, \ t \in \mathbb{R}_+\}$ of the Langevin equation defined by:*

$$\mathrm{d}G_t^{\mathrm{gbO}} = (\mathrm{h}^{\mathrm{gbO}}(\ell))^{(1/2)} \mathrm{d}B_t + (1/2)\mathrm{h}^{\mathrm{gbO}}(\ell)\nabla \log \pi_c(G_t^{\mathrm{gbO}})\mathrm{d}t \ ,$$

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

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

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

Using in addition Assumption 1 there exists $m_0 > 0$ such that for all $h \in I$ and for all $x \in \mathbb{R}$,

$$\mathscr{T}_1(g''(x), 2h, a_4) + ((a_4/2) - (1/6)) \, \mathscr{T}_2(g''(x), 2h, a_5) \geq m_0 \ .$$

Using these two results, the proof of both theorems follows the same lines as 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:

$$
\begin{aligned}
y^{\mathrm{bO}} &= \mu^{\mathrm{bO}}(x, h) + S^{\mathrm{bO}}(x, h) \, \xi \ , &&(19) \\
\mu^{\mathrm{bO}}(x, h) &= x + \mathscr{T}_1(Df(x), h, 1)f(x) + (2/3)\mathscr{T}_2(Df(x), h, 1)f(x) \\
&\quad - (1/3)\mathscr{T}_3(Df(x), h, 1)\{\Sigma : D^2 f(x)\} \ , \\
S^{\mathrm{bO}}(x, h) &= (\mathscr{T}_1(Df(x), 2h, 1) + (1/3)\mathscr{T}_2(Df(x), 2h, 1))^{1/2} \ ,
\end{aligned}
$$

where $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{T}_3$ are respectively defined by (14), (16) and (17). First, let us begin with some definitions. For a signed measure ν on $\mathbb{R}^d$, we define the total variation norm of ν by

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

where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra of $\mathbb{R}^d$. Let P be a Markov kernel with invariant measure π. For a given measurable function $V : \mathbb{R}^d \to [1, +\infty)$, 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 ν a signed measure on $\mathbb{R}^d$, the V-norm $\|\cdot\|_V$ is defined by

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

We refer the reader to [17] for the definitions of small sets, φ-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 $V : \mathbb{R}^d \to [1, +\infty)$ 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 [17, Theorem 15.0.1]), namely for some small set $\mathcal{C}$, there exist $\lambda < 1$ and $b < +\infty$ such that for all $x \in \mathbb{R}^d$:

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\mathcal{C}}(x) \,. \tag{20}$$

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:

$$q(x, y) = \frac{1}{(2\pi)^{d/2} |S(x)|} \exp\left(-(1/2)\left\langle S(x)^{-2}(y - \mu(x)), (y - \mu(x))\right\rangle\right) \,, \tag{21}$$

where for $\mathrm{M} \in \mathbb{R}^{d \times d}$, $|\mathrm{M}|$ 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 [24, 9]. But contrary to [9], we assume for simplicity the following assumption on the functions $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and $S : \mathbb{R}^d \to \mathcal{S}_+^*(\mathbb{R}^d)$:

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

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

Assumption 4. *The log-density g of π 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 [9] 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 π satisfying Assumption 4. Then, the produced Markov kernel P is given by

$$P(x, \mathrm{d}y) = \alpha(x, y)q(x, y)\mathrm{d}y + \delta_x(\mathrm{d}y) \int_{\mathbb{R}^d} (1 - \alpha(x, y))q(x, y)\mathrm{d}y \,, \tag{22}$$

where q and α are resp. given by (21) and (2).

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

Note that this condition is necessary to obtain the geometric ergodicity of a Metropolis-Hastings algorithm by [25, 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 [9], 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*

$$\limsup_{\|x\| \to +\infty} \|\mu(x)\| / \|x\| = \tau, \ and \ \limsup_{\|x\| \to +\infty} \|S(x)\| / \|x\| = 0 \,. \tag{23}$$

Then, the Markov kernel P given by (22) are V-geometrically ergodic, where $V(x) = 1 + \|x\|^2$.

Proof. The proof is postponed to Appendix B.1. □

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

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

$$\liminf_{\|x\| \to +\infty} \left\| S(x)^{-1}\mu(x) \right\| \|x\|^{-1} > \epsilon^{-1} \,, \qquad \liminf_{\|x\| \to +\infty} \inf_{\|y\|=1} \|S(x)y\| \geq \epsilon \,, \tag{24}$$

and

$$\lim_{\|x\| \to +\infty} \log \left(|S(x)| \right) / \|x\|^2 = 0 \,. \tag{25}$$

Then, P is not geometrically ergodic.

Proof. The proof is postponed to Appendix B.2. □

4.2 Exponential potentials

We illustrate our results on the following classes of density.

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

Let π be a probability density on $\mathbb{R}$ with respect to the Lebesgue measure. We will say that $\pi \in \mathcal{E}(\beta, \gamma)$ if π 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 $\mathscr{P}_m$

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

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

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

$$\left\| D^2(\nabla r)(x) \right\| \underset{\|x\| \to +\infty}{=} o(\|x\|^{m-3}) , \tag{26}$$

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

$$\mathrm{q}(x) = \mathrm{p}(x) + \mathrm{r}(x) .$$

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

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

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

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 $x \in \mathbb{R}^d$.

$$\mathrm{p}(x) = \|x\|^m \ \mathrm{p}(\vec{n}_x) \tag{27}$$
$$\nabla \mathrm{p}(x) = \|x\|^{m-1} \nabla \mathrm{p}(\vec{n}_x) \tag{28}$$
$$\nabla^2 \mathrm{p}(x) = \|x\|^{m-2} \nabla^2 \mathrm{p}(\vec{n}_x) \tag{29}$$
$$D^2(\nabla \mathrm{p})(x) = \|x\|^{m-3} D^2(\nabla \mathrm{p})(x) \tag{30}$$
$$\langle \nabla \mathrm{p}(x), x \rangle = m \ \mathrm{p}(x) \tag{31}$$
$$\nabla^2 \mathrm{p}(x) x = (m-1) \ \nabla \mathrm{p}(x) \tag{32}$$
$$\langle \nabla^2 \mathrm{p}(x) x, x \rangle = m(m-1) \ \mathrm{p}(x) . \tag{33}$$

From (29), it follows that $\nabla^2 \mathrm{p}(x)$ is definite positive for all $x \in \mathbb{R}^d \setminus 0$ if and only if $\nabla^2 \mathrm{p}(\vec{n})$ is positive definite for all $\vec{n}$, with $\|\vec{n}\| = 1$. Then, p belongs to $\mathscr{P}_m^+$ 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 $\mathscr{E}(\beta, \gamma)$, but also for the polynomial class $\mathscr{P}_m^+$.

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

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 [24] has for the same values of β, 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 [24, Theorem 4.3] we deduce that the two algorithms are not geometrically ergodic for $\beta \in (0, 1)$. Similarly, the proof in [24, 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 [24, Theorem 4.1], and not geometrically ergodic if $h\gamma(1 + h\gamma/6) > 2$ by Theorem 4.2, since μ^{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 μ^{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 [24, Theorem 4.1], and not geometrically ergodic if $h\gamma > 1.23$ by Theorem [25, 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 (19) when x is large, which are implied by the behaviour of $\lambda_{\min}(Df(x)))$ and $\lambda_{\max}(Df(x))$.

If $\liminf_{\|x\| \to +\infty} \lambda_{\min}(Df(x)) = 0$. Then, $g(x) = o(\|x\|^2)$ as $\|x\| \to \infty$, and y^{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 [24] for this one.

Example 4.3. *By [24, Theorem 4.1] bOMA is geometrically ergodic for $\pi \in \mathscr{E}(\gamma, \beta)$ with $\beta \in [1, 2)$.*

Now, we focus on the case where $\limsup_{\|x\| \to +\infty} \lambda_{\max}(Df(x)) < 0$. For instance, this condition holds for $\pi \in \mathscr{E}(\gamma, \beta)$ when $\beta \geq 2$. We give conditions similar to the one for geometric convergence of the Ozaki discretization, given in [9], 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. $\limsup_{\|x\| \to +\infty} \lambda_{\max}(Df(x)) < 0$;

2. $\lim_{\|x\| \to +\infty} Df(x)^{-2}\{\mathrm{I}_d : D^2 f(x)\} = 0$;

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

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

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

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

Proof. Since $S^{\mathrm{bO}}(x, h)$ 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 $x, \|x\| \geq R$, $\mathrm{Sp}(Df(x)) \subset \mathbb{R}_-$. 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*

$$0 < \inf_{n \in \mathbb{S}^d} c(n) \leq \sup_{n \in \mathbb{S}^d} c(n) < 6/5 \, , \tag{34}$$

then bOMA is geometrically ergodic.

Proof. We check that the conditions of Lemma 4.1 hold. By Assumption 4 and (19), Assumption 3 holds, thus it remains to check (23). First, Lemma 4.4 implies that the second equality of (23) is satisfied, and we just need to prove the first equality. By [9, Lemma 3.4], it suffices to prove that

$$\limsup_{\|x\| \to +\infty} \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle < 0 \, , \tag{35}$$

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

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

and $x \mapsto \mathscr{B}(x)$ is bounded on $\mathbb{R}^d$. Since $\mathscr{B}$ is bounded on $\mathbb{R}^d$, by Assumption 6-(2)-(3) and (34),

$$\lim_{\|x\| \to +\infty} \left| \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle - \|\mathscr{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2\langle \mathscr{B}(x)\vec{n}_x, \vec{n}_x \rangle c(\vec{n}_x) \right| = 0 \, . \tag{36}$$

15

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

$$\|\mathscr{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2\langle \mathscr{B}(x)\vec{n}_x, \vec{n}_x\rangle c(\vec{n}_x)$$
$$= \sum_{i=1}^{d} c(\vec{n}_x)\lambda_i(x)\langle e_i(x),\vec{n}_x\rangle^2 (c(\vec{n}_x)\lambda_i(x) + 2) \quad (37)$$

Since $\limsup_{\|x\|\to+\infty} Df(x) < 0$, for all i and $\|x\|$ large enough, $\lambda_i(x) \in [-5/3, 0)$. Therefore using (34) we get from (37):

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

The proof is concluded using this result in (36). $\qquad\square$

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

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

Lemma 4.6 ([9, Proof of Theorem 4.10]). *Let $\pi \in \mathscr{P}_m^+$ for $m \geq 2$, then π 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 $\pi \in \mathscr{P}_m^+$ for $m \geq 2$, then bOMA is V-geometrically ergodic, with $V(x) = \|x\|^2 + 1$.*

Proof. Let us denote $\pi \propto \exp(-\mathrm{p}(x) - \mathrm{r}(x))$, with p and r satisfying the conditions from the definition in Section 4.2.2. We prove that if $\pi \in \mathscr{P}_m^+$, Theorem 4.5 can be applied. First, by definition of $\mathscr{P}_m^+$, Assumption 4 is satisfied. Furthermore, Assumption 6-(1)-(2) follows from (26), (29), (30) and the condition that $\nabla^2\mathrm{p}(\vec{n})$ 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 μ^{bO} and S^{bO} by μ and S, and do not mention the dependence in the parameter h of μ and S when it does not play any role. Note that

$$\int_{\mathbb{R}^d} \alpha(x,y) q(x,y)\mathrm{d}y = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \{1 \wedge \exp \widetilde{\alpha}(x,\xi)\} \exp(-\|\xi\|^2/2)\mathrm{d}\xi , \quad (38)$$

where

$$\widetilde{\alpha}(x,\xi) = -\mathrm{p}(\mu(x) + S(x)\xi) + \mathrm{p}(x) - \mathrm{r}(\mu(x)$$
$$+ S(x)\xi) + \mathrm{r}(x) - \log(|S(\mu(x) + S(x)\xi)|) + \log(|S(x)|) + (1/2)\|\xi\|^2$$
$$- (1/2)\left\langle (\widetilde{S}(x,\xi))^{-1}\{x - \mu(\mu(x) + S(x)\xi)\}, x - \mu(\mu(x) + S(x)\xi)\right\rangle , \quad (39)$$

and $\widetilde{S}(x,\xi) = S(\mu(x) + S(x)\xi)S(\mu(x) + S(x)\xi)^T$. First, we consider $m \geq 3$, then we have the following estimate of the terms in (39) by (26)-(30) and Lemma 4.6:

$$\mu(w) \underset{\|w\|\to+\infty}{=} \{1 - 5/(3(m-1))\}w + o(\|w\|) \quad (40)$$

$$(S(w)S(w)^T)^{-1} \underset{\|w\|\to+\infty}{=} \frac{3}{4}m(m-1)\|w\|^{m-2}\nabla^2\mathrm{p}(\vec{n}_w) + o(\|w\|^{m-2}) \quad (41)$$

$$\log(|S(w)|) \underset{\|w\|\to+\infty}{=} o(\|w\|) \quad (42)$$

16

Then by (40)-(42), if we define $\Psi : [3, +\infty) \to \mathbb{R}$ 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

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

Since Ψ is positive on $[3, +\infty)$, for all $\xi \in \mathbb{R}^d \lim_{\|x\| \to +\infty} \widetilde{\alpha}(x, \xi) = +\infty$. This result, (38) and Fatou's Lemma imply that Assumption 5 is satisfied.

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

$$\varrho(\mathrm{M}) = (e^{-\mathrm{M}} - \mathrm{I}_p) + (2/3)(e^{-\mathrm{M}^2} - \mathrm{I}_p)$$

$$\varsigma(\mathrm{M}) = (e^{-2\mathrm{M}} - \mathrm{I}_p) + (1/3)(e^{-4\mathrm{M}^2} - \mathrm{I}_p) .$$

Then we have the following estimates:

$$\widetilde{\alpha}(x, \xi) \underset{\|x\| \to +\infty}{=} \left\langle \mathrm{A}(\varsigma(h\mathrm{A}))^{-1} \left\{ \left(2\varrho(h\mathrm{A}) + \varrho(h\mathrm{A})^2 \right) x \right\}, \left(2\varrho(h\mathrm{A}) + \varrho(h\mathrm{A})^2 \right) x \right\rangle$$

$$+ \langle \mathrm{A}x, x \rangle - \langle \mathrm{A} \left\{ (\mathrm{I}_d + \varrho(h\mathrm{A}))x \right\}, (\mathrm{I}_d + \varrho(h\mathrm{A}))x \rangle + o(\|x\|^2) \tag{43}$$

If we denote the eigenvalues of A by $\{\lambda_i, i = 1 \ldots d\}$ and $\{x_i, i = 1, \ldots, d\}$ the coordinates of x in an orthonormal basis of eigenvectors for A, (43) becomes

$$\widetilde{\alpha}(x, \xi) \underset{\|x\| \to +\infty}{=} \sum_{i=1}^d \Xi(h, \lambda_i) x_i^2 + o(\|x\|^2) . \tag{44}$$

where for $h, \lambda > 0$,

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

Using that for any $h, \lambda > 0$, $\Xi(h, \lambda) > 0$ and (44), we have for all $\xi \in \mathbb{R}^d$, $\lim_{\|x\| \to +\infty} \widetilde{\alpha}(x, \xi) = +\infty$, 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 $\pi \in \mathscr{E}(\beta, \gamma)$ with $\beta \geq 2$.*

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

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 [24] for the ULA to Gaussian Markov kernels on $\mathbb{R}$. Under Assumption 3, it is straightforward to see that Q is Leb^d-irreducible, where Leb^d is the Lebesgue measure, aperiodic and all compact set of $\mathbb{R}^d$ are small; see [9, Theorem 3.1]. We now state our main theorems, which essentially complete [24, Theorem 3.1-3.2]. Since their proof are very similar, they are omitted.

Method	$\beta \in [1, 2)$	$\beta = 2$	$\beta > 2$
fMALA (12)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
mOMA (13)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
bOMA (19)	geometrically ergodic	geometrically ergodic	geometrically ergodic

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

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

$$\limsup_{|x| \to +\infty} S(x) \le s_\wedge \ ,$$

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

(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 [24, 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:*

$$\liminf_{|x| \to +\infty} S(x) \ge s_\vee \ ,$$

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

(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 [24, Theorem 3.2]. □

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

We now apply Theorems 4.9 and 4.10 in order to study the ergodicity of the different unadjusted proposals applied to $\pi \in \mathscr{E}(\beta, \gamma)$. 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 [24] has for the same values of β, 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 $\beta \ge 2$ we have the following cases.

18

(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 transcient if $h\gamma(1 + h\gamma/6) > 2$ by Theorem 4.10-(2), since μ^{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 transcient if $1 + 2(h\gamma)^2/3 - e^{-\gamma h} > 2$ by Theorem 4.10-(2), since μ^{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 μ^{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 transcient 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 (12)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
mUOA (13)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
bUOA (19)	ergodic	geometrically ergodic	geometrically ergodic	geometrically ergodic

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

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 [21], 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 (8) using the double well potential with $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$ in dimensions $d = 10, 100, 500, 1000$, respectively. We consider many time stepsizes $h = \ell^2 d^{-1/5}$, plotting the first order efficiency (multiplied by $d^{1/5}$ because this is the scale which is asymptotically constant for fMALA as $d \to \infty$) as a function of the acceptance rate for the standard MALA (white bullets) and the acceptance rate $a_d^{\text{fM}}(\ell)$ of the improved version fMALA (black bullets), respectively. For simplicity, each
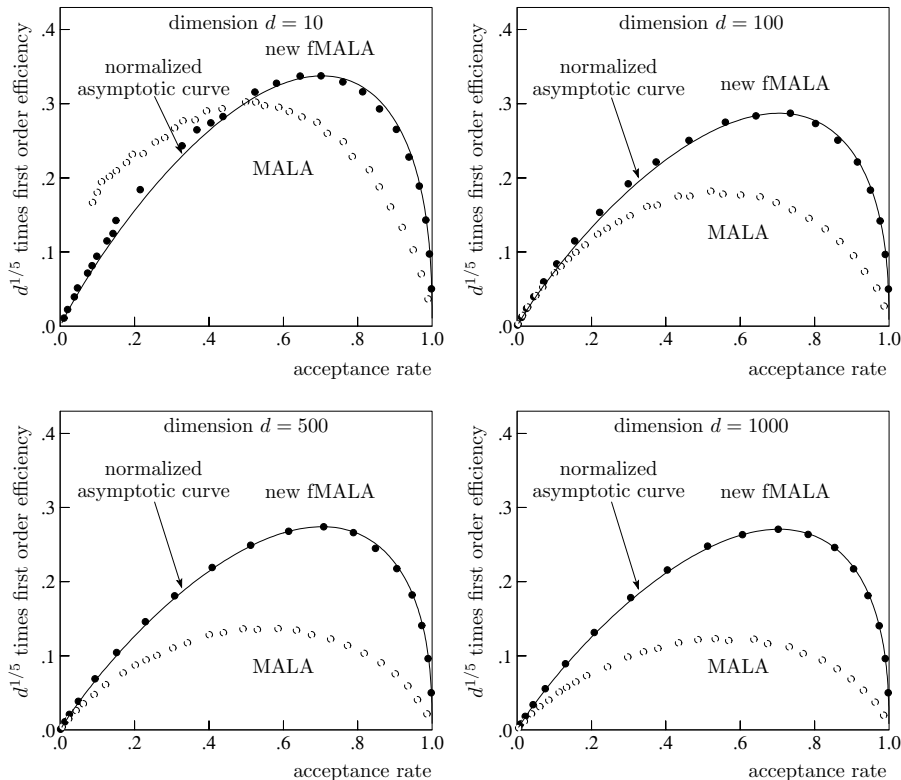
19

Figure 1: First-order efficiency of the new fMALA and the standard MALA for the double well potential $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$, as a function of the overall acceptance rates in dimensions $d = 10, 100, 500, 1000$. 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.

chain is started from the origin. The expectations are approximated as the average over 2×10^5 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 goes to infinity, normalised to have the same maximum as fMALA in finite dimension d. This corresponds to the (rescaled) limiting diffusion speed $h^{fM}(\ell)$ as a function of $a^{fM}(\ell)$ (quantities given respectively in Theorems 3.1 and 3.2). We observe excellent agreement of the numerical first order efficiency compared to the asymptotic one, especially as d increases, which corroborates the scaling results of fMALA. In addition, 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.

For our next experiments, we consider the d-dimensional zero-mean Gaussian distribution with covariance matrix I_d for $d = 1000$, as target distribution. We aim to numerically study the transient behaviour of fMALA and propose some solutions to overcome this issue. In Figure 2, we plot the squared norm of 10^4 samples generated by the RWM, MALA, fMALA and some hybrid strategies for MALA and fMALA, all started from the origin. We also include a zoom on the first 100 steps. In Figure 2a, we use standard implementations of the
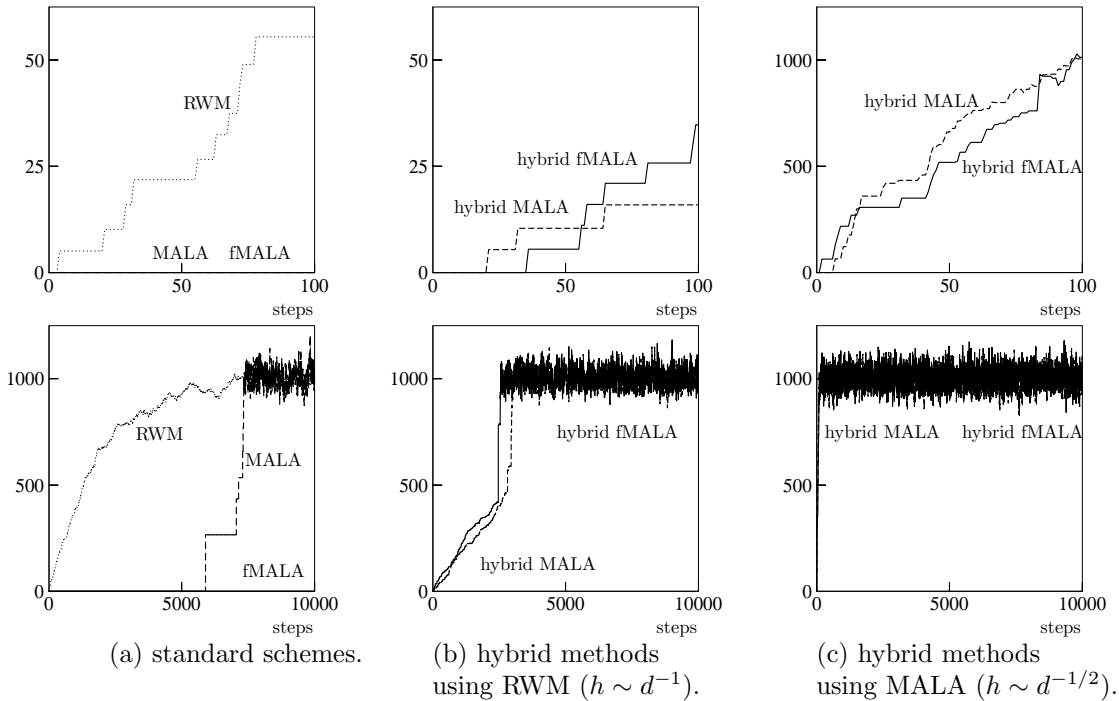
Figure 2: Trace plots of $\|X\|^2$ for the Gaussian target density in dimension $d = 1000$ when starting at the origin. Comparison of fMALA with $h \sim d^{-1/5}$ (solid lines), MALA with $h \sim d^{-1/3}$ (dashed lines), RWM with $h \sim d^{-1}$ (dotted lines).
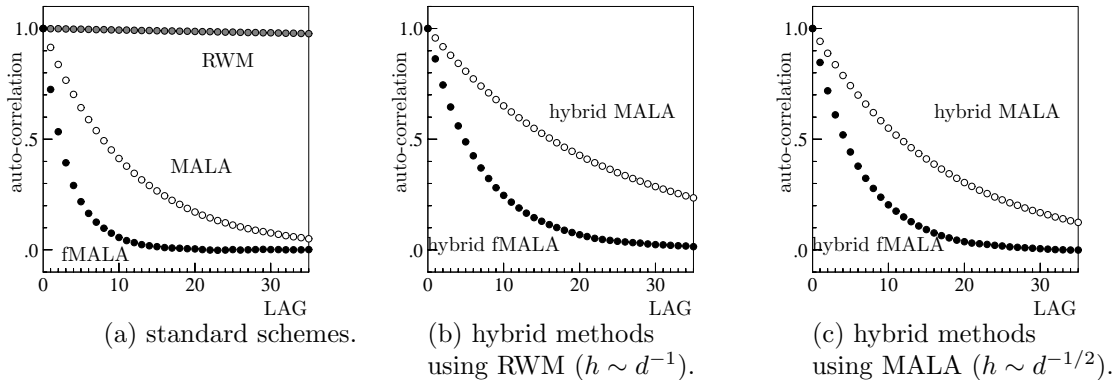

Figure 3: Auto-correlation versus LAG for the Gaussian target density in dimension $d = 1000$. Comparison of fMALA with $h \sim d^{-1/5}$ (black), MALA with $h \sim d^{-1/3}$ (white), RWM with $h \sim d^{-1}$ (gray).

schemes. The time step h for each algorithm is chosen as the optimal parameter based on the optimal scaling results of all the algorithms at stationarity: for the RWM $h = 2.38^2 d^{-1}$, for MALA $h = 1.65^2 d^{-1/3}$ and for fMALA $h = 1.79^2 d^{-1/5}$. It can observed that MALA exhibits many rejected steps in contrast to RWM. This is a known issue of MALA in the transient phase [4, 12] due to a tiny acceptance probability at first steps, and the same behaviour can be observed for fMALA, with zero accepted step in the present simulation. To circumvent this issue, the following hybrid MALA scheme was presented in [4]. The idea is to combine MALA with RWM at each step: with probability 1/2, we apply the MALA proposal (4) with

21

step size $h = 1.65^2 d^{-1/3}$, the optimal parameter for MALA at stationarity. Otherwise, the RWM proposal (3) is used with step size $h = 2.38^2 d^{-1}$, the optimal parameter for the RWM at stationarity. Indeed, [4] and [12] have shown that the optimal scaling in the transient phase and at stationarity is the same and scales as d^{-1}. In Figure 2b, the plots for this hybrid MALA are presented, the same methodology is also applied for the hybrid fMALA scheme, showing a behaviour similar to hybrid MALA. In Figure 2c, the RWM proposal is replaced by the MALA proposal (4) with a different step size $h = 2d^{-1/2}$, which is the optimal parameter for MALA in the transient phase according to [4]. Again, hybrid fMALA exhibits a behaviour similar to hybrid MALA.

In Figure 3, we consider again the same schemes and hybrid versions as in Figure 2, with the same step sizes, and we compare their autocorrelation function. We consider for each algorithms $2 \cdot 10^5$ iterations started at stationarity, where the first 10^3 iterations were discarded as burn-in. In Figure 3a, it can be observed that the autocorrelation associated with fMALA goes to 0 quicker than the RWM and MALA. In Figure 3b, and Figure 3c, we observe that by using hybrid strategies which are designed to robustify convergence from the transient phase, fMALA still comfortably outperforms MALA in terms of expected square efficiency (which is a stationary quantity).

Although our analysis applies only to product measure densities of the form (8), we next consider the following non-product density in $\mathbb{R}^d$, defined using a normalization constant Z_d and for $X_0 = 0$ as

$$\pi(X_1, \ldots, X_d) = Z_d \prod_{i=1}^{d} \frac{1}{1 + (X_i - \alpha(X_{i-1}))^2}, \tag{45}$$

where we consider the scalar functions $\alpha(x) = x/2$ and $\alpha(x) = \sin(x)$, respectively. Notice that the density (45) is associated with the AR(1) process $X_i = \alpha(X_{i-1}) + Z_n$ with non Gaussian (Cauchy) increments Z_n. Furthermore, we observe that in this case the Jacobian in (12) is a symmetric tridiagonal matrix, which implies that the computational cost of the fMALA proposal is of the same order $\mathcal{O}(d)$ as the standard MALA proposal.

In Figure 4, we compare for many timesteps the standard MALA (left pictures) and the new fMALA (right pictures), and plot the (scaled) first order efficiency $\mathbb{E}[\|X_{k+1} - X_k\|^2/d]$ as a function of the overall acceptance rates, using the averages over 2×10^4 iterations of the algorithms. The initial condition for both algorithms is the same and is obtained after running 10^4 steps of the RWM algorithm to get close to the target probability measure. Analogously to the product case studied in Figure 1, we observe in both cases $\alpha(x) = x/2$ and $\alpha(x) = \sin(x)$ that the first-order efficiency of fMALA converges to a non-zero limiting curve with maximum close to the value 0.704. In contrast, the efficiency of the standard MALA drops to zero in this scaling where the first-order efficiency is multiplied with $d^{1/5}$. This numerical experiment suggests that our analysis in the product measure setting persists in the non product measure case.

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.
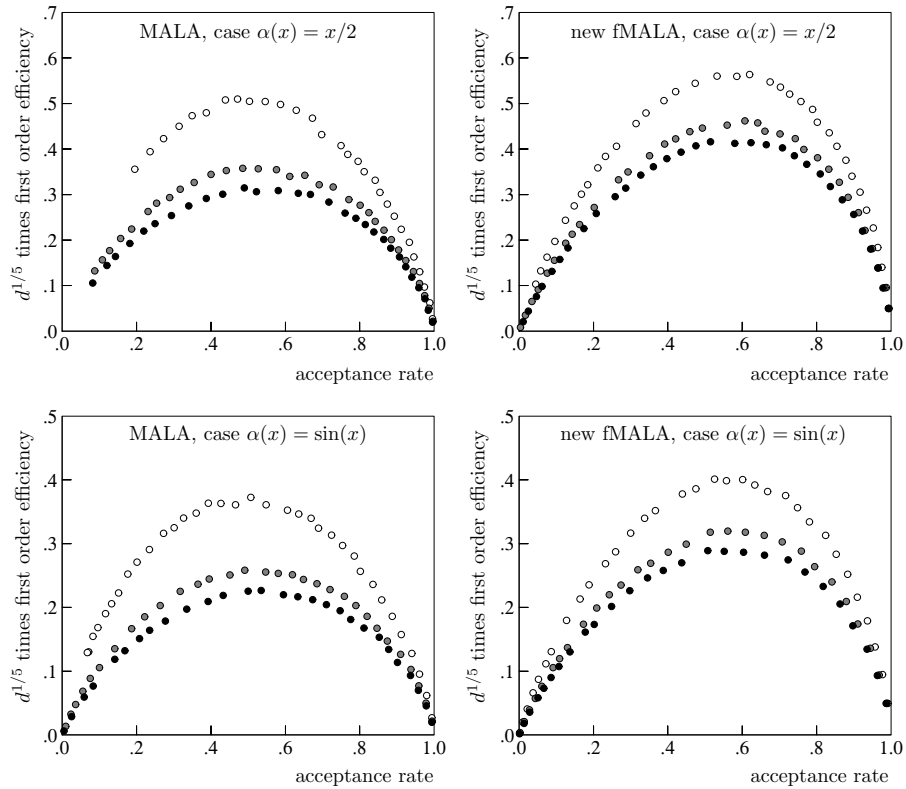
Figure 4: First-order efficiency of the new fMALA and the standard MALA as a function of the overall acceptance rates for the dimensions $d = 100$ (white points), $d = 500$ (gray points), $d = 1000$ (dark points), respectively, for the non product density (45) with $\alpha(x) = x/2$ (top pictures) and $\alpha(x) = \sin(x)$ (bottom pictures).

[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] O. F. Christensen, G. O. Roberts, and J. S. Rosenthal. Scaling limits for the transient phase of local Metropolis–Hastings algorithms. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):253–268, 2005.

[5] 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, 08 2013.

[6] 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.

[7] M. Fathi and G. Stoltz. Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics. *preprint*, 2015.

[8] 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.

[9] N. R. Hansen. Geometric ergodicity of discrete-time approximations to multivariate diffusions. *Bernoulli*, 9(4):725–743, 2003.

[10] 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.

[11] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[12] B. Jourdain, T. Lelièvre, and B. Miasojedow. Optimal scaling for the transient phase of metropolis hastings algorithms: the longtime behavior. *Bernoulli*, 20(4):1930–1978, 2014.

[13] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[14] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008.

[15] 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.

[16] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[17] 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.

[18] 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.

[19] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[20] 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.

[21] 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.

[22] G. O. Roberts and J. S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, 16(4):351–367, 2001.

[23] 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).

[24] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[25] 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.

[26] 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.

[27] 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 [20] and [21]. 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 (18), defined for $\psi \in C_c^2(\mathbb{R})$, and $x \in \mathbb{R}^d$ by

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

We check that the assumptions of [6, 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 $T \geq 0$:

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

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

for all functions ψ in a core of A^{fM}, which strongly separates points. Since A^{fM} is an operator on the set of functions 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 [6, Theorem 2.1, Chapter 8], this set of functions is a core for A^{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 π_1. Also, for all $x \in \mathbb{R}^d$, denote by y^{fM} the proposal of fMALA, defined by (9), (12a) and (12b), started at $x \in \mathbb{R}^d$, with parameter h_d and associated with the d-dimensional Gaussian random variable $\{\xi_i, i = 1, \cdots, d\}$.

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

$$\log\left(\frac{\pi(y_i^{\text{fM}})q^{\text{fM}}(y_i^{\text{fM}}, x_i)}{\pi(x_i)q^{\text{fM}}(x_i, y_i^{\text{fM}})}\right) = \sum_{j=5}^{10} C_j^{\text{fM}}(x_i, \xi_i)d^{-j/10} + C_{11}^{\text{fM}}(x_i, \xi_i, h_d) , \qquad (46)$$

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

$$\mathbb{E}\left[C_j^{\text{fM}}(X, \xi_1)\right] = 0 \ for \ j = 5, \cdots, 9 , \tag{47}$$

$$\mathbb{E}\left[\left(\mathbb{E}\left[C_5^{\text{fM}}(X, \xi_1) \,|X\right]\right)^2\right] = \ell^{10}(K^{\text{fM}})^2 = -2\mathbb{E}\left[C_{10}^{\text{fM}}(X, \xi_1)\right] . \tag{48}$$

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

$$\lim_{d \to +\infty} d^{-3/5} \sup_{x \in F_d^1} \mathbb{E}\left[\left|\sum_{i=2}^{d} C_j(x_i^d, \xi_i) - \mathbb{E}\left[C_j^{\text{fM}}(X, \xi_i)\right]\right|\right] = 0 , \tag{49}$$

and

$$\lim_{d \to +\infty} \sup_{x \in F_d^1} \mathbb{E}\left[\left|\sum_{i=2}^{d} C_{11}(x_i^d, \xi_i, h_d)\right|\right] = 0 . \tag{50}$$

Finally,

$$\lim_{d \to +\infty} \sup_{x \in F_d^1} \mathbb{E}\left[\left|\zeta^d\right|\right] = 0 , \tag{51}$$

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 [27]. Then, since just odd powers of ξ_i occur in C_5, C_7 and C_9, we deduce (47) 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 (47) for $j = 6, 8$ and (48). We now build the sequence of sets F_d^1, which satisfies the claimed properties. Denote for $j = 6, \cdots, 10$ 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 ξ_i and the derivatives of g at x_i. Therefore for all $k \in \mathbb{N}^*$,

$$\mathbb{E}\left[\left|\tilde{C}_j^{\text{fM}}(X)\right|^k\right] + \mathbb{E}\left[\left|V_j^{\text{fM}}(X)\right|^k\right] < +\infty . \tag{52}$$

Consider for all $j = 6, \cdots, 10$, 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

$$F_{d,j,1}^1 = \left\{x \in \mathbb{R}^d \ ; \ \left|\sum_{i=2}^{d} \tilde{C}_j^{\text{fM}}(x_i) - \mathbb{E}\left[\tilde{C}_j^{\text{fM}}(X)\right]\right| \leq d^{23/40}\right\} \tag{53}$$

$$F_{d,j,2}^1 = \left\{x \in \mathbb{R}^d \ ; \ \left|\sum_{i=2}^{d} V_j^{\text{fM}}(X) - \mathbb{E}\left[V_j^{\text{fM}}(X)\right]\right| \leq d^{23/20}\right\} . \tag{54}$$

Note that $\lim_{d \to +\infty} d^{1/5} \pi_d((F^1_{d,j})^c) = 0$ for all $j = 6 \cdots 10$, is implied by $\lim_{d \to +\infty} d^{1/5} \pi_d((F^1_{d,j,1})^c) = 0$ and $\lim_{d \to +\infty} d^{1/5} \pi_d((F^1_{d,j,2})^c) = 0$. Let $\{X_i, i \geq 2\}$ be a sequence of i.i.d. random variables with distribution π_1. By definition of $F^1_{d,j,1}$, the Markov inequality and independence, we get

$$
\begin{aligned}
d^{1/5} \pi_d((F^1_{d,j,1})^c) &\leq d^{-21/10} \mathbb{E}\left[\left(\sum_{i=2}^d \tilde{C}^{\mathrm{fM}}_j(X_i) - \mathbb{E}\left[\tilde{C}^{\mathrm{fM}}_j(X) \right] \right)^4 \right] \\
&\leq \sum_{i_1,i_2=2}^d \mathbb{E}\left[\left(\tilde{C}^{\mathrm{fM}}_j(X_{i_1}) - \mathbb{E}\left[\tilde{C}^{\mathrm{fM}}_j(X) \right] \right)^2 \left(\tilde{C}^{\mathrm{fM}}_j(X_{i_2}) - \mathbb{E}\left[\tilde{C}^{\mathrm{fM}}_j(X) \right] \right)^2 \right] \\
&\leq d^{-1/10} \mathbb{E}\left[\left(\tilde{C}^{\mathrm{fM}}_j(X) - \mathbb{E}\left[\tilde{C}^{\mathrm{fM}}_j(X) \right] \right)^4 \right] ,
\end{aligned}
\tag{55}
$$

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

$$
\begin{aligned}
d^{1/5} \pi_d((F^1_{d,j,2})^c) &\leq d^{-21/10} \mathbb{E}\left[\left(\sum_{i=2}^d V^{\mathrm{fM}}_j(X_i) - \mathbb{E}\left[V^{\mathrm{fM}}_j(X) \right] \right)^2 \right] \\
&\leq d^{-1/10} \mathbb{E}\left[\left(V^{\mathrm{fM}}_j(X) - \mathbb{E}\left[V^{\mathrm{fM}}_j(X) \right] \right)^2 \right] .
\end{aligned}
\tag{56}
$$

Therefore (52), (55) and (56) imply that $\lim_{d \to +\infty} d^{1/5} \pi_d((F^1_{d,j})^c) = 0$ for all $j = 6, \cdots, 10$. In addition, for all $x \in F^1_{d,j}$, by the triangle inequality and the Cauchy-Schwarz inequality we have for all $j = 6, \cdots, 10$

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

Therefore by this inequality, (53) and (54), there exists a constant M_1 such that

$$
d^{3/5} \sup_{x \in F^1_{d,j}} \mathbb{E}\left[\left| \sum_{i=2}^d C^{\mathrm{fM}}_j(x_i, \xi_i) - \mathbb{E}\left[C^{\mathrm{fM}}_j(X, \xi) \right] \right| \right] \leq d^{-1/40} M_1 ,
$$

and (49) follows. It remains to show (50). By definition, C_{11} is the remainder in the eleventh order expansion in $\sigma_d := \sqrt{h_d}$ given by (46) of the function Θ defined by $\Theta(x_i, \xi_i, \sigma_d) = \log(\pi_1(y^{\mathrm{fM}}_i) q^{\mathrm{fM}}_1(y^{\mathrm{fM}}_i, x_i)) - \log(\pi_1(x_i) q^{\mathrm{fM}}_1(x_i, y^{\mathrm{fM}}_i))$. 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 Θ with respect to σ_d, taken in (x_i, ξ_i, u_d), can be bounded

by a positive polynomial in (x_i, ξ_i) on the form $P_2(x_i)P_3(\xi_i)$. Hence, there exists a constant M_2 such that

$$\mathbb{E}\left[|C_{11}(x_i, \xi_i, h_d)|\right] \leq M_2 \, d^{-11/10} \, P_2(x_i) \,. \tag{57}$$

And if we define

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

then we have by the Chebychev inequality, this definition and (57)

$$d^{1/5}\pi_d((F_{d,11}^1)^c) \leq \mathrm{Var}\left[P_2(X)\right] d^{-4/5}$$

$$\sup_{x \in F_{d,11}^1} \sum_{i=2}^d \mathbb{E}\left[|C_{11}(x_i, \xi_i, h_d)|\right] \leq M_2(\mathbb{E}\left[P_2(x)\right] + 1)d^{-1/10} \,.$$

These results, combined with Assumption 2-(3), imply $\lim_{d\to+\infty} d^{1/5}\pi_d((F_{d,11}^1)^c) = 0$ and (50). Finally, $F_d^1 = \bigcap_{j=6}^{11} F_{d,j}^1$ satisfies the claimed properties of the Lemma, and (51) directly follows from all the previous results. $\qquad\square$

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

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

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

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

The next lemma shows that we can approximate A_d^{fM} by $\tilde{A}_d^{\mathrm{fM}}$, 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 $\lim_{d\to+\infty} d^{1/5}\pi_d((F_d^2)^c) = 0$ 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):*

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

In addition,

$$\lim_{d\to+\infty} \sup_{x \in F_d^2} d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] = 0 \,. \tag{58}$$

Proof. Using that ψ is bounded and the Jensen inequality, there exists a constant M_1 such that

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

Thus it suffices to show (58). Set $\sigma_d = \sqrt{h_d}$. Since $t \mapsto 1 \wedge \exp(t)$ is 1-Lipschtz on $\mathbb{R}$ and, by definition we have

$$d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] \leq d^{1/5}\mathbb{E}\left[|\Theta(x_1, \xi_1, \sigma_d)|\right] \,, \tag{59}$$

28

where $\Theta(x_1, \xi_1, \sigma_d) = \log(\pi_1(y_1^{\mathrm{fM}})q_1^{\mathrm{fM}}(y_1^{\mathrm{fM}}, x_1)) - \log(\pi_1(x_1)q_1^{\mathrm{fM}}(x_1, y_1^{\mathrm{fM}}))$. By a fifth order Taylor expansion of Θ in σ_d, and since by (46) $\partial^j \Theta(x_1, \xi_1, 0)/(\partial\sigma_d^j) = 0$ for $j = 0 \cdots 4$, 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

$$|\Theta(x_1, \xi_1, \sigma_d)| \le (\sigma_d^5/5!)\mathrm{P}_1(x_1)\mathrm{P}_2(\xi_1) \, .$$

Plugging this result in (59) and since $\sigma_d^5 = \ell^{5/2}d^{-1/2}$, we get

$$d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] \le \ell^{5/2}d^{-3/10}\mathrm{P}_1(x_1) \, .$$

Setting $F_d^2 = \{x \in \mathbb{R}^d \, ; \, \mathrm{P}_1(x_1) \le d^{1/10}\}$, we have

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

and (58) follows. Finally, F_d^2 satisfied $\lim_{d \to +\infty} d^{1/5}\pi_d((F_d^2)^c) = 0$ since by the Markov inequality

$$d^{1/5}\pi_d((F_d^2)^c) \le d^{-1/10}\mathbb{E}\left[\mathrm{P}_1(\mathrm{X})^3\right] \, ,$$

where $\mathbb{E}\left[\mathrm{P}_1(\mathrm{X})^3\right]$ is finite by Assumption 2-(3). $\qquad\square$

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

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

Proof. Consider $\sigma_d = \sqrt{h_d}$ and $W(x_1, \xi_1, \sigma_d) = \psi(y_1^{\mathrm{fM}})$. 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 σ_d implies there exists $u_d \in [0, h_d]$ and $M_1 \ge 0$ such that

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

Moreover since $\psi \in C_c^\infty(\mathbb{R})$, the third partial derivative of W in σ_d are bounded for all x_1, ξ_1 and σ_d. Therefore there exists $M_2 \ge 0$ such that for all $x_1 \in \mathbb{R}$,

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

which concludes the proof. $\qquad\square$

As in [21], 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 $C_5^{\mathrm{fM}}(x_i, \xi_i)$. Define now for $d \geq 2$ and $x \in \mathbb{R}^d$,

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

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

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

Finally define the characteristic function of the zero-mean Gaussian distribution with standard deviation $\ell^5 K^{\mathrm{fM}}$, given in Lemma A.1, by: for $t \in \mathbb{R}$,

$$\varphi(t) = e^{-(\ell^5 K^{\mathrm{fM}} t)^2/2} \,.$$

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

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

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

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

In particular, we have

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

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

$$F_{d,1}^3 = \bigcap_{j=2,4} \left\{ x \in \mathbb{R}^d \; ; \; \left| d^{-1} \sum_{i=2}^{d} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^j \right] - \mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1, \xi_1)^j \right] \right| \leq d^{-1/4} \right\} \,, \tag{60}$$

$$F_{d,2}^3 = \left\{ x \in \mathbb{R}^d \; ; \; \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^2 \right] \leq d^{3/4} \; \forall i \in \{2, \cdots, d\} \right\} \,. \tag{61}$$

It follows from (52), and the Chebychev and Markov inequalities that there exists a constant M such that $\pi_d((F_{d,1}^3)^c) + \pi_d((F_{d,2}^3)^c) \leq M d^{-1/2}$. Therefore $\lim_{d \to +\infty} d^{1/5} \pi_d((F_d^3)^c) = 0$. (i). Let $t \in \mathbb{R}$ and $x \in F_d^3$ and denote

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

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

$$|\varphi_d(x, t) - \varphi(t)| \leq \left| \varphi_d(x, t) - \prod_{i=2}^{d} \left(1 - \frac{\ell^{10} \mathrm{V}(x_i) t^2}{2d} \right) \right|$$

$$+ \left| \prod_{i=2}^{d} \left(1 - \frac{\ell^{10} \mathrm{V}(x_i) t^2}{2d} \right) - e^{-\ell^{10}(K^{\mathrm{fM}})^2 t^2/2} \right| \,. \tag{62}$$

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 (61), for d large enough $\ell^{10}\mathrm{V}(x_i)t^2/(2d) \leq 1$ for all $i \in \{2, \cdots, d\}$. Thus, by [3, Eq. 26.5], we have for such large d, all $i \in \{2, \cdots, d\}$ and all $\delta > 0$:

$$\left| \varphi_1(x_i, t/\sqrt{d}) - \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d} \right) \right|$$

$$\leq \mathbb{E}\left[\left(\frac{|t|^3 \ell^{15}}{6d^{3/2}} \left| C_5^{\mathrm{fM}}(x_i, \xi_i) \right|^3 \right) \wedge \left(\frac{t^2 \ell^{10}}{d} C_5^{\mathrm{fM}}(x_i, \xi_i)^2 \right) \right]$$

$$\leq \mathbb{E}\left[\frac{|t|^3 \ell^{15}}{6d^{3/2}} \left| C_5^{\mathrm{fM}}(x_i, \xi_i) \right|^3 \mathbb{1}_{\{|C_5^{\mathrm{fM}}(x_i, \xi_i)| \leq \delta d^{1/2}\}} \right]$$

$$+ \mathbb{E}\left[\frac{t^2 \ell^{10}}{d} C_5^{\mathrm{fM}}(x_i, \xi_i)^2 \mathbb{1}_{\{|C_5^{\mathrm{fM}}(x_i, \xi_i)| > \delta d^{1/2}\}} \right]$$

$$\leq \frac{\delta |t|^3 \ell^{15}}{6d} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^2 \right] + \frac{\ell^{10}t^2}{\delta^2 d^2} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^4 \right] \,,$$

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

$$\left| \varphi_d(x,t) - \prod_{i=2}^d \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d} \right) \right| \leq \sum_{i=2}^d \frac{\delta |t|^3 \ell^{15}}{6d} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^2 \right]$$

$$+ \frac{\ell^{10}t^2}{\delta^2 d^2} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^4 \right]$$

$$\leq \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1, \xi_1)^2 \right] + d^{-1/4} \right) \ell^{15}\delta |t|^3 /6$$

$$+ \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1, \xi_1)^4 \right] + d^{-1/4} \right) \ell^{10}t^2/(\delta^2 d) \,,$$

where the last inequality follows from $x \in F_d^3$ and (60) Let now $\epsilon > 0$, and choose δ small enough such that the fist 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

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

Consider now the second term of (62), by the triangle inequality,

$$\left| \prod_{i=2}^d \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d} \right) - e^{-\ell^{10}(K^{\mathrm{fM}})^2 t^2/2} \right|$$

$$\leq \left| \prod_{i=2}^d \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d} \right) - \prod_{i=2}^d e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)} \right|$$

$$+ \left| \prod_{i=2}^d e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)} - e^{-\ell^{10}(K^{\mathrm{fM}})^2 t^2/2} \right| \,. \quad (63)$$

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

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

31

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

$$\left| \prod_{i=2}^{d} \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d} \right) - \prod_{i=2}^{d} e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)} \right|$$

$$\leq \sum_{i=2}^{d} \left| 1 - \mathrm{V}(x_i)\ell^{10}t^2/(2d) - e^{-\mathrm{V}(x_i)\ell^{10}t^2/(2d)} \right|$$

$$\leq \sum_{i=2}^{d} \mathrm{V}(x_i)^2 \ell^{20} t^4/(8d^2) \leq \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1, \xi_1)^4 \right] + d^{-1/4} \right) \ell^{20} t^4/(8d) , \qquad (64)$$

where the last inequality is implied by (60). Finally since on $\mathbb{R}_-$, $u \mapsto e^u$ is 1-Lipschitz and using (60), we get

$$\left| \prod_{i=2}^{d} e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)} - e^{-\ell^{10}(K^{\mathrm{fM}})^2 t^2/2} \right| \leq (t^2\ell^{10}/2) \left| \sum_{i=2}^{d} d^{-1}\mathrm{V}(x_i) - (K^{\mathrm{fM}})^2 \right|$$

$$\leq t^2 \ell^{10} d^{-1/4}/2 . \qquad (65)$$

Therefore, combining (64) and (65) in (63), we get:

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

which concludes the proof of (i).
(ii) Let $\mathrm{b} : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. Consider the sequence $\{x^d , d \in \mathbb{N}^*\}$ of elements of F_d^3 which satisfies for all $d \in \mathbb{N}^*$,

$$\sup_{y \in F_d^3} \left| \mathbb{E}\left[\mathrm{b}\left(\bar{\mathsf{M}}_d(y) \right) \right] - (2\pi\ell^{10}(K^{\mathrm{fM}})^2)^{-1/2} \int_{\mathbb{R}} \mathrm{b}(u) e^{-u^2/(2\ell^{10}(K^{\mathrm{fM}})^2)} \mathrm{d}u \right|$$

$$\leq \left| \mathbb{E}\left[\mathrm{b}\left(\bar{\mathsf{M}}_d(x^d) \right) \right] - (2\pi\ell^{10}(K^{\mathrm{fM}})^2)^{-1/2} \int_{\mathbb{R}} \mathrm{b}(u) e^{-u^2/(2\ell^{10}(K^{\mathrm{fM}})^2)} \mathrm{d}u \right| + d^{-1} . \quad (66)$$

Then using (i) and Levy's continuity theorem, we get

$$\lim_{d \to +\infty} \left| \mathbb{E}\left[\mathrm{b}\left(\bar{\mathsf{M}}_d(x^d) \right) \right] - (2\pi\ell^{10}(K^{\mathrm{fM}})^2)^{-1/2} \int_{\mathbb{R}} \mathrm{b}(u) e^{-u^2/(2\ell^{10}(K^{\mathrm{fM}})^2)} \mathrm{d}u \right| = 0 .$$

This limit and (66) conclude the proof. $\qquad \square$

proof of Theorem 3.1. The theorem follows from Lemma A.1, (58) in Theorem A.2 and the last statement in Lemma A.4. $\qquad \square$

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 $\lim_{d \to +\infty} d^{-1/5} \pi_d((F_d)^c) = 0$ and by the union bound, for all $T \geq 0$,

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

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):

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

Then, the weak convergence follows from [6, Corollary 8.7, Chapter 4]. □

B Postponed proofs

B.1 Proof of Lemma 4.1

By Assumption 3-4, π and q are positive and continuous. It follows from [16, Lemma 1.2] that P is Leb^d-irreducible aperiodic, where 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 [17, Theorem 15.0.1], we just need to check the drift condition (20). 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

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

By (23), $\limsup_{\|x\| \to +\infty} (2 \|\mu(x)\| \|S(x)\| + \|S(x)\|^2) \|x\|^{-2} = 0$. Therefore, using again the first inequality of (23) and Assumption 5:

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

This concludes the proof of Lemma 4.1. □

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 $\lim_{n \to +\infty} \pi(x_n) = +\infty$. Since we assume that π is bounded, we have our contradiction.

If P is geometrically ergodic, then by [25, Theorem 5.1], there exists $\eta > 0$ such that for almost every $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \alpha(x,y) q(x,y) \mathrm{d}y \geq \eta \ , \tag{67}$$

and let $M \geq 0$ such that

$$\mathbb{P}\left[\|\xi\| \geq M\right] \leq \eta/2 \ , \tag{68}$$

where ξ is a standard d-dimensional Gaussian random variable. By (24), there exist $R_\epsilon, \delta > 0$ such that

$$\inf_{\{\|x\| \geq R_\epsilon\}} \left\| S(x)^{-1} \mu(x) \right\| \|x\|^{-1} \geq \epsilon^{-1} + \delta \tag{69}$$

$$\inf_{\{\|x\| \geq R_\epsilon\}} \inf_{\|z\|=1} \|S(x)z\| \geq \epsilon(1 + \delta\epsilon/2)^{-1} . \tag{70}$$

Note that we can assume R_ϵ is large enough so that

$$\epsilon \delta R_\epsilon / 2 \geq M . \tag{71}$$

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

$$B(x) = \left\{ y \in \mathbb{R}^d \mid \left\| S(x)^{-1}(y - \mu(x)) \right\| \leq M \right\} . \tag{72}$$

Note if $y \in B(x)$, we have by definition and the triangle inequality $\left\| S(x)^{-1}y \right\| \geq \left\| S(x)^{-1}\mu(x) \right\| - M$. Therefore by (69)-(70) and (71)

$$\|y\| = \left\| S(x)S(x)^{-1}y \right\| \geq \epsilon(1 + \delta\epsilon/2)^{-1} \left\| S(x)^{-1}y \right\|$$
$$\geq \epsilon(1 + \delta\epsilon/2)^{-1} \left\{ (\epsilon^{-1} + \delta) \|x\| - M \right\} \geq \|x\| . \tag{73}$$

We then show that this inequality implies

$$\liminf_{\|x\| \to +\infty} \inf_{y \in B(x)} \frac{q(y,x)}{q(x,y)} = 0 . \tag{74}$$

Let $x \in \mathbb{R}^d$, $\|x\| \geq R_\epsilon$, $y \in B(x)$. First, it is straightforward by (72), that $|S(x)| q(x,y)$ is uniformly bounded away from 0, and it suffices to consider $|S(x)| q(y,x)$. By (70)-(73), we have $\|y\| \geq R_\epsilon$ and for all $z \in \mathbb{R}^d$, $\|S(y)z\| \geq \epsilon(1 + \delta\epsilon/2)^{-1} \|z\|$, which implies for all $z \in \mathbb{R}^d$, $\epsilon^{-1}(1 + \delta\epsilon/2) \|z\| \geq \left\| S(y)^{-1}z \right\|$. By this inequality and (69), we have

$$\left| \left\| S(y)^{-1}\mu(y) \right\| - \left\| S(y)^{-1}x \right\| \right| \geq \left\| S(y)^{-1}\mu(y) \right\| - \left\| S(y)^{-1}x \right\|$$
$$\geq (\epsilon^{-1} + \delta) \|y\| - \epsilon^{-1}(1 + \delta\epsilon/2) \|x\| \geq (\delta/2) \|y\| , \tag{75}$$

where the last inequality follows from (73). Using this result, the triangle inequality, (75)-(70) and (73), we get

$$q(y,x) = (2\pi)^{-d/2} \exp \left\{ -(1/2) \left\| S(y)^{-1}(x - \mu(y)) \right\|^2 - \log(|S(y)|) \right\}$$
$$\leq (2\pi)^{-d/2} \exp \left\{ -(1/2) \left(\left\| S(y)^{-1}\mu(y) \right\| - \left\| S(y)^{-1}x \right\| \right)^2 - \log(|S(y)|) \right\}$$
$$\leq (2\pi)^{-d/2} \exp \left\{ -(\delta^2/8) \|y\|^2 - \log(|S(y)|) \right\}$$
$$\leq (2\pi)^{-d/2} \exp \left\{ -(\delta^2/8) \|x\|^2 - d\log(\epsilon(1 + \delta\epsilon/2)^{-1}) \right\} .$$

Using this inequality and (25) imply $\lim_{\|x\| \to +\infty} \inf_{y \in B(x)} |S(x)| q(y,x) = 0$ and then (74). Therefore there exists $R_q \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq R_q$

$$\inf_{y \in B(x)} \frac{q(y,x)}{q(x,y)} \leq \eta/4 . \tag{76}$$

34

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 $\lim_{n \to +\infty} \pi(x_n) = +\infty$. 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 nth term and such that for all $k = 0, \ldots, n-1$, $\|x_{k+1}\| \geq \max(R_\epsilon, R_q)$ 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 (67)-(68) and (76)

$$\eta \leq \int_{\mathbb{R}^d} \alpha(x_n, y)q(x_n, y)\mathrm{d}y \leq \eta/2 + \int_{B(x_n)} \min\left(1, \frac{\pi(y)q(y, x_n)}{\pi(x_n)q(x_n, y)}\right)q(x_n, y)\mathrm{d}y$$

$$\leq \eta/2 + (\eta/4)\int_{B(x_n)} \frac{\pi(y)}{\pi(x_n)}q(x_n, y)\mathrm{d}y .$$

This inequality implies that $\int_{B(x_n)} \frac{\pi(y)}{\pi(x_n)}q(x_n, y)\mathrm{d}y \geq 2$ 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 (73), $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$. 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 $\lim_{n \to +\infty} \pi(x_n) = +\infty$, which contradicts the assumption that π is bounded. This concludes the proof of Theorem 4.2. □

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

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

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

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

$$C_5^{\text{gbO}}(x_1, \xi_1) = \ell^5 \left(\frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1) \right.$$

$$+ \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{1}{72}a_3\xi_1 g^{(3)}(x_1)g''(x_1) + \frac{1}{6}a_4^2\xi_1^3 g^{(3)}(x_1)g''(x_1)$$

$$- \frac{1}{6}a_4^2\xi_1 g^{(3)}(x_1)g''(x_1) + \frac{5}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) + \frac{1}{48}\xi_1 g^{(3)}(x_1)g''(x_1)$$

$$+ \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 - \frac{1}{24}a_1^2\xi_1 g'(x_1)g''(x_1)^2 + \frac{1}{6}a_4^2\xi_1 g'(x_1)g''(x_1)^2$$

$$\left. + \frac{1}{24}\xi_1 g'(x_1)g''(x_1)^2 \right) .$$

D Expressions of $K^\bullet$

We provide here the expressions of the quantities $K^\bullet$ involved in Theorems 3.1, 3.2, 3.4, 3.5. Let X be a random variable distributed according to π_1.

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

$$+ \frac{49g'(X)^2g''(X)^4}{20736} + \frac{7}{576}g^{(4)}(X)g^{(5)}(X)g'(X) + \frac{19}{864}g^{(3)}(X)g^{(5)}(X)g''(X)$$

$$+ \frac{1}{288}g^{(3)}(X)g^{(5)}(X)g'(X)^2 + \frac{7g^{(5)}(X)g'(X)g''(X)^2}{1728}$$

$$+ \frac{1}{144}g^{(3)}(X)g^{(4)}(X)g'(X)^3 + \frac{7}{864}g^{(4)}(X)g'(X)^2g''(X)^2 + \frac{7g^{(3)}(X)g'(X)^3g''(X)^2}{1728}$$

$$\left. + \frac{5}{432}g^{(3)}(X)^2g'(X)^2g''(X) + \frac{35g^{(3)}(X)g'(X)g''(X)^3}{2592} + \frac{29}{864}g^{(3)}(X)g^{(4)}(X)g'(X)g''(X) \right] .$$

$$K^{\text{mO}} = \mathbb{E}\left[\frac{79g^{(5)}(X)^2}{17280} + \frac{11g^{(4)}(X)^2g'(X)^2}{1152} + \frac{1567g^{(3)}(X)^2g''(X)^2}{3456} \right.$$

$$+ \frac{1}{576}g^{(3)}(X)^2g'(X)^4 + \frac{1}{36}g'(X)^2g''(X)^4 + \frac{7}{576}g^{(4)}(X)g^{(5)}(X)g'(X)$$

$$+ \frac{17}{192}g^{(3)}(X)g^{(5)}(X)g''(X) + \frac{1}{288}g^{(3)}(X)g^{(5)}(X)g'(X)^2$$

$$+ \frac{1}{72}g^{(5)}(X)g'(X)g''(X)^2 + \frac{1}{144}g^{(3)}(X)g^{(4)}(X)g'(X)^3 +$$

$$\frac{1}{36}g^{(4)}(X)g'(X)^2g''(X)^2 + \frac{1}{72}g^{(3)}(X)g'(X)^3g''(X)^2$$

$$+ \frac{11}{288}g^{(3)}(X)^2g'(X)^2g''(X) + \frac{11}{72}g^{(3)}(X)g'(X)g''(X)^3 +$$

$$\left. \frac{73}{576}g^{(3)}(X)g^{(4)}(X)g'(X)g''(X) \right] .$$

$$K^{\text{gbO}} = \mathbb{E}\left[\frac{1}{36}g'(X)^2g''(X)^4a_4^4 + \frac{5}{18}g''(X)^2g^{(3)}(X)^2a_4^4 \right.$$

$$+ \frac{1}{9}g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})a_4^4 - \frac{1}{72}a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^4 a_4^2 + \frac{1}{72}g'(\mathrm{X})^2 g''(\mathrm{X})^4 a_4^2$$

$$+ \frac{11}{72}g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 a_4^2 + \frac{1}{108}a_3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 a_4^2$$

$$+ \frac{1}{36}g'(\mathrm{X})^2 g''(\mathrm{X})g^{(3)}(\mathrm{X})^2 a_4^2 - \frac{1}{36}a_1^2 g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})a_4^2$$

$$+ \frac{5}{72}g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})a_4^2 + \frac{1}{216}a_3 g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})a_4^2$$

$$+ \frac{1}{72}g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})a_4^2 + \frac{1}{36}g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X})a_4^2$$

$$+ \frac{7}{72}g'(\mathrm{X})g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})a_4^2 + \frac{1}{72}g'(\mathrm{X})g''(\mathrm{X})^2 g^{(5)}(\mathrm{X})a_4^2$$

$$+ \frac{5}{72}g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})a_4^2 + \frac{1}{576}a_1^4 g'(\mathrm{X})^2 g''(\mathrm{X})^4$$

$$- \frac{1}{288}a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^4 + \frac{1}{576}g'(\mathrm{X})^2 g''(\mathrm{X})^4 + \frac{1}{576}g'(\mathrm{X})^4 g^{(3)}(\mathrm{X})^2$$

$$+ \frac{a_3^2 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2}{5184} + \frac{1}{288}a_3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2$$

$$+ \frac{79 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2}{3456} + \frac{1}{96}g'(\mathrm{X})^2 g''(\mathrm{X})g^{(3)}(\mathrm{X})^2$$

$$+ \frac{1}{864}a_3 g'(\mathrm{X})^2 g''(\mathrm{X})g^{(3)}(\mathrm{X})^2 + \frac{11 g'(\mathrm{X})^2 g^{(4)}(\mathrm{X})^2}{1152}$$

$$+ \frac{79 g^{(5)}(\mathrm{X})^2}{17280} - \frac{1}{96}a_1^2 g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})$$

$$+ \frac{1}{96}g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) - \frac{1}{864}a_1^2 a_3 g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X})$$

$$+ \frac{1}{864}a_3 g'(\mathrm{X})g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) - \frac{1}{288}a_1^2 g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})$$

$$+ \frac{1}{288}g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X}) - \frac{1}{144}a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X})$$

$$+ \frac{1}{144}g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X}) + \frac{1}{144}g'(\mathrm{X})^3 g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})$$

$$+ \frac{17}{576}g'(\mathrm{X})g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X}) + \frac{1}{432}a_3 g'(\mathrm{X})g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})$$

$$- \frac{1}{288}a_1^2 g'(\mathrm{X})g''(\mathrm{X})^2 g^{(5)}(\mathrm{X}) + \frac{1}{288}g'(\mathrm{X})g''(\mathrm{X})^2 g^{(5)}(\mathrm{X})+$$

$$\frac{1}{288}g'(\mathrm{X})^2 g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X}) + \frac{11}{576}g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})$$

$$+ \frac{1}{864}a_3 g''(\mathrm{X})g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X}) + \frac{7}{576}g'(\mathrm{X})g^{(4)}(\mathrm{X})g^{(5)}(\mathrm{X}) \Bigg] \ .$$