Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory

MAYLIS VARVENNE

Institut de Mathématiques de Toulouse, 118 route de Narbonne, 31400 Toulouse, France.
E-mail: maylis.varvenne@math.univ-toulouse.fr

The main objective of the paper is to study the long-time behavior of general discrete dynamics driven by an ergodic stationary Gaussian noise. In our main result, we prove existence and uniqueness of the invariant distribution and exhibit some upper-bounds on the rate of convergence to equilibrium in terms of the asymptotic behavior of the covariance function of the Gaussian noise (or equivalently to its moving average representation). Then, we apply our general results to fractional dynamics (including the Euler Scheme associated to fractional driven Stochastic Differential Equations). When the Hurst parameter H belongs to $(0, 1/2)$ we retrieve, with a slightly more explicit approach due to the discrete-time setting, the rate exhibited by Hairer in a continuous time setting (*Ann. Probab.* **33** (2005) 703–758). In this fractional setting, we also emphasize the significant dependence of the rate of convergence to equilibrium on the local behaviour of the covariance function of the Gaussian noise.

Keywords: discrete stochastic dynamics; Lyapunov function; rate of convergence to equilibrium; stationary Gaussian noise; Toeplitz operator; total variation distance

1. Introduction

Convergence to equilibrium for Stochastic dynamics is one of the most natural and most studied problems in probability theory. Regarding Markov processes, this topic has been deeply undertaken through various approaches: spectral analysis, functional inequalities or coupling methods. However, in many applications (Physics, Biology, Finance, . . .) the future evolution of a quantity may depend on its own history, and thus, noise with independent increments does not accurately reflect reality. A classical way to overcome this problem is to consider dynamical systems driven by processes with stationary increments like fractional Brownian motion (fBm) for instance which is widely used in applications (see e.g. [11,16,18,19]). In a continuous time framework, Stochastic Differential Equations (SDEs) driven by Gaussian processes with stationary increments have been introduced to model random evolution phenomena with long range dependence properties. Consider SDEs of the following form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dZ_t, \tag{1.1}$$

where $(Z_t)_{t \geq 0}$ is a Gaussian process with ergodic stationary increments and $\sigma : \mathbb{R}^d \to \mathcal{M}_d(\mathbb{R})$, $b : \mathbb{R}^d \to \mathbb{R}^d$ are functions defined in a such a way that existence of a solution holds. The ergodic properties of such processes have been a topic of great interest over the last decade. For

general Gaussian processes, existence and approximation of stationary solutions are provided in [5] in the additive case (i.e., when σ is constant). The specific situation where $(Z_t)_{t\geq 0}$ is a fractional Brownian motion has received significant attention since in the seminal paper [12] by Hairer, a definition of invariant distribution is given in the additive case through the embedding of the solution to an infinite-dimensional markovian structure. This point of view led to some probabilistic uniqueness criteria (on this topic, see, e.g., [14,15]) and to some coupling methods in view of the study of the rate of convergence to equilibrium. More precisely, in [12], some coalescent coupling arguments are also developed and lead to the convergence of the process in total variation to the stationary regime with a rate upper-bounded by $C_\varepsilon t^{-(\alpha_H - \varepsilon)}$ for any $\varepsilon > 0$, with

$$\alpha_H = \begin{cases} \dfrac{1}{8} & \text{if } H \in \left(\dfrac{1}{4}, 1\right) \setminus \left\{\dfrac{1}{2}\right\}, \\[2ex] H(1 - 2H) & \text{if } H \in \left(0, \dfrac{1}{4}\right). \end{cases} \tag{1.2}$$

In the multiplicative noise setting (i.e., when σ is not constant), Fontbona and Panloup in [9] extended those results under selected assumptions on σ to the case where $H \in (\frac{1}{2}, 1)$ and finally Deya, Panloup and Tindel obtained in [7] this type of results in the rough setting $H \in (\frac{1}{3}, \frac{1}{2})$.

In this paper, we focus on a general class of recursive discrete dynamics of the following form: we consider $\mathbb{R}^d$-valued sequences $(X_n)_{n\geq 0}$ satisfying

$$X_{n+1} = F(X_n, \Delta_{n+1}), \tag{1.3}$$

where $(\Delta_n)_{n\in\mathbb{Z}}$ is an ergodic stationary Gaussian sequence and $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a deterministic function. As a typical example for F, we can think about Euler discretization of (1.1) (see Section 2.5 for a detailed study) or to autoregressive processes in the particular case where F is linear. Note that such dynamics can be written as (1.3) through the so-called Wold's decomposition theorem which implies that we can see $(\Delta_n)_{n\in\mathbb{Z}}$ as a moving-average of infinite order (see [2] to get more details). To the best of our knowledge, in this linear setting the literature mainly focuses on the statistical analysis of the model (see, for instance, [3,4]) and on mixing properties of such Gaussian processes when it is in addition stationary (on this topic see, e.g., [1,17,22]). Back to the main topic, namely ergodic properties of (1.3) for general functions F, Hairer in [13] has provided technical criteria using ergodic theory to get existence and uniqueness of invariant distribution for this kind of a priori non-Markovian processes. Here, the objective is to investigate the problem of the long-time behavior of (1.3). To this end, we first explain how it is possible to define invariant distributions in this non-Markovian setting and to obtain existence results. More precisely, with the help of the moving average representation of the noise process $(\Delta_n)_{n\in\mathbb{Z}}$, we prove that $(X_n, (\Delta_{n+k})_{k\leq 0})_{n\geq 0}$ can be realized through a Feller transformation $\mathcal{Q}$. In particular, an initial distribution of the dynamical system (X, Δ) is a distribution μ on $\mathbb{R}^d \times (\mathbb{R}^d)^{\mathbb{Z}^-}$. Rephrased in more probabilistic terms, an initial distribution is the distribution of the couple $(X_0, (\Delta_k)_{k\leq 0})_{n\geq 0}$. Then, such an initial distribution is called an invariant distribution if it is invariant by the transformation $\mathcal{Q}$. The first part of the main theorem establishes the existence of such an invariant distribution.

Then, our main contribution is to state a general result about the rate of convergence to equilibrium in terms of the covariance structure of the Gaussian process. To this end, we use a coalescent coupling strategy.

Let us briefly explain how this coupling method is organized in this discrete-time framework. First, one considers two processes $(X_n^1, (\Delta_{n+k}^1)_{k\leq 0})_{n\geq 0}$ and $(X_n^2, (\Delta_{n+k}^2)_{k\leq 0})_{n\geq 0}$ following (1.3) starting respectively, from μ_0 and $\mu_\star$ (an invariant distribution). As a preliminary step, one waits that the two paths get close. Then, at each trial, the coupling attempt is divided in two steps. First, one tries in Step 1 to stick the positions together at a given time. Then, in Step 2, one attempts to ensure that the paths stay clustered until $+\infty$. Actually, oppositely to the Markovian setting where the paths remain naturally fastened together (by putting the same innovation on each marginal), the main difficulty here is that, staying together has a cost. In other words, this property can be ensured only with a non trivial coupling of the noises. Finally, if one of the two previous steps fails, one begins Step 3 by putting the same noise on each coordinate until the "cost" to attempt Step 1 is not too big. More precisely, during this step one waits for the memory of the coupling cost to decrease sufficiently and for the two trajectories to be in a compact set with lower-bounded probability.

In the main theorem previously announced, as a result of this strategy, we are able to prove that the law of the process $(X_{n+k})_{k\geq 0}$ following (1.3) converges in total variation to the stationary regime with a rate upper-bounded by $Cn^{-\upsilon}$. The quantity υ is directly linked to the assumed exponential or polynomial asymptotic decay of the sequence involved in the moving-average representation of $(\Delta_n)_{n\in\mathbb{Z}}$, see (2.2) (or equivalently in its covariance function, see Remarks 2.1 and 2.3).

Then, we apply our main theorem to fractional memory (including the Euler Scheme associated to fractional Stochastic Differential Equations). We first emphasize that with covariance structures with the same order of memory but different local behavior, we can get distinct rates of convergence to equilibrium. Second, we highlight that the computation of the asymptotic decay of the sequence involved in the inversion of the Toeplitz operator (related to the moving-average representation) can be a very technical task (see proof of Proposition 2.3).

Now, let us discuss about specific contributions of this discrete-time approach. Above all, our result is quite general since, for instance, it includes discretization of (1.1) for a large class of Gaussian noise processes. Then, in several ways, we get a further understanding of arguments used in the coupling procedure. We better target the impact of the memory through the sequence both appearing in the moving-average representation and the covariance function of the Gaussian noise. Regarding Step 1 of the coupling strategy, the "admissibility condition" (which means that we are able to attempt Step 1 with a controlled cost) is rather more explicit than in the continuous-time setting. Finally, this paper, by deconstructing the coupling method through this explicit discrete-time framework, may weigh in favour of the sharpness of Hairer's approach.

The following section gives more details on the studied dynamics, describes the assumptions required to get the main result, namely Theorem 2.1 and discuss about the application of our main result to the case of fractional memory in Section 2.7. Then, the proof of Theorem 2.1 is achieved in Sections 3, 4, 5, 6 and 7, which are outlined at the end of Section 2.

2. Setting and main results

2.1. Notations

The usual scalar product on $\mathbb{R}^d$ is denoted by $\langle \cdot, \cdot \rangle$ and $| \cdot |$ stands either for the Euclidean norm on $\mathbb{R}^d$ or the absolute value on $\mathbb{R}$. For a given $K > 0$ we denote by $B(0, K)$ the $\mathbb{R}^d$ closed ball centered in 0 with radius K. Then, the *state space* of the process X and the *noise space* associated to $((\Delta_{n+k})_{k \leq 0})_{n \geq 0}$ are respectively, denoted by $\mathcal{X} := \mathbb{R}^d$ and $\mathcal{W} := (\mathbb{R}^d)^{\mathbb{Z}^-}$. For a given measurable space $(\mathcal{X}_1, \mathcal{A}_1)$, $\mathcal{M}_1(\mathcal{X}_1)$ will denote the set of all probability measures on $\mathcal{X}_1$. Let $(\mathcal{X}_2, \mathcal{A}_2)$ be an other measurable space and $f : \mathcal{X}_1 \to \mathcal{X}_2$ be a measurable mapping. Let $\mu \in \mathcal{M}_1(\mathcal{X}_1)$, we denote by $f^* \mu$ the pushforward measure given by:

$$\forall B \in \mathcal{A}_2, \qquad f^* \mu(B) := \mu(f^{-1}(B)).$$

$\Pi_{\mathcal{X}} : \mathcal{X} \times \mathcal{W} \to \mathcal{X}$ and $\Pi_{\mathcal{W}} : \mathcal{X} \times \mathcal{W} \to \mathcal{W}$ stand respectively, for the projection on the marginal $\mathcal{X}$ and $\mathcal{W}$. For a given differentiable function $f : \mathbb{R}^d \to \mathbb{R}^d$ and for all $x \in \mathbb{R}^d$ we will denote by $J_f(x)$ the Jacobian matrix of f valued at point x. Finally, we denote by $\| \cdot \|_{\mathrm{TV}}$ the classical total variation norm: let $\nu, \mu \in \mathcal{M}_1(\mathcal{X}_1)$,

$$\| \mu - \nu \|_{\mathrm{TV}} := \sup_{A \subset \mathcal{X}_1} \big| \mu(A) - \nu(A) \big|.$$

2.2. Dynamical system and Markovian structure

Let $X := (X_n)_{n \geq 0}$ denote an $\mathbb{R}^d$-valued sequence defined by: X_0 is a random variable with a given distribution and

$$\forall n \geq 0, \qquad X_{n+1} = F(X_n, \Delta_{n+1}), \tag{2.1}$$

where $F : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is continuous and $(\Delta_n)_{n \in \mathbb{Z}}$ is a stationary and purely non-deterministic Gaussian sequence with d independent components. Hence, by Wold's decomposition theorem [2] it has a moving average representation

$$\forall n \in \mathbb{Z}, \qquad \Delta_n = \sum_{k=0}^{+\infty} a_k \xi_{n-k} \tag{2.2}$$

with

$$\begin{cases} (a_k)_{k \geq 0} \in \mathbb{R}^{\mathbb{N}} \text{ such that } a_0 \neq 0 \text{ and } \sum_{k=0}^{+\infty} a_k^2 < +\infty, \\ (\xi_k)_{k \in \mathbb{Z}} \text{ an i.i.d. sequence such that } \xi_1 \sim \mathcal{N}(0, I_d). \end{cases} \tag{2.3}$$

Without loss of generality, we assume that $a_0 = 1$. Actually, if $a_0 \neq 1$, we can come back to this case by setting

$$\tilde{\Delta}_n = \sum_{k=0}^{+\infty} \tilde{a}_k \xi_{n-k}$$

with $\tilde{a}_k := \frac{a_k}{a_0}$.

Remark 2.1. The asymptotic behavior of the sequence $(a_k)_{k \geq 0}$ certainly plays a key role to compute the rate of convergence to equilibrium of the process $(X_n)_{n \geq 0}$. Actually, the memory induced by the noise process is quantified by the sequence $(a_k)_{k \geq 0}$ through the identity

$$\forall n \in \mathbb{Z}, \forall k \geq 0, \qquad c(k) := \mathbb{E}[\Delta_n \Delta_{n+k}] = \sum_{i=0}^{+\infty} a_i a_{k+i}.$$

The stochastic dynamics described in (2.1) is clearly non-Markovian. Let us see how it is possible to introduce a Markovian structure and how to define invariant distribution. This method is inspired by [13]. The first idea is to look at $(X_n, (\Delta_{n+k})_{k \leq 0})_{n \geq 0}$ instead of $(X_n)_{n \geq 0}$. Let us introduce the following concatenation operator

$$\sqcup: \quad \mathcal{W} \times \mathbb{R}^d \to \mathcal{W}, \tag{2.4}$$

$$(w, w') \mapsto w \sqcup w',$$

where $(w \sqcup w')_0 = w'$ and $\forall k < 0, (w \sqcup w')_k = w_{k+1}$. Then (2.1) is equivalent to the system

$$\left(X_{n+1}, (\Delta_{n+1+k})_{k \leq 0} \right) = \varphi \left((X_n, (\Delta_{n+k})_{k \leq 0}), \Delta_{n+1} \right), \tag{2.5}$$

where

$$\varphi: \quad (\mathcal{X} \times \mathcal{W}) \times \mathbb{R}^d \to \mathcal{X} \times \mathcal{W},$$

$$((x, w), w') \mapsto \left(F(x, w'), w \sqcup w' \right).$$

Therefore, $(X_n, (\Delta_{n+k})_{k \leq 0})_{n \geq 0}$ can be realized through the Feller Markov transition kernel $\mathcal{Q}$ defined by

$$\int_{\mathcal{X} \times \mathcal{W}} g(x', w') \mathcal{Q}((x, w), (dx', dw')) = \int_{\mathbb{R}^d} g(\varphi((x, w), \delta)) \mathcal{P}(w, d\delta), \tag{2.6}$$

where $\mathcal{P}(w, d\delta) := \mathcal{L}(\Delta_{n+1} \mid (\Delta_{n+k})_{k \leq 0} = w)$ does not depend on n since (Δ_n) is a stationary sequence, and $g : \mathcal{X} \times \mathcal{W} \to \mathbb{R}$ is a measurable function.

Definition 2.1. A measure $\mu \in \mathcal{M}_1(\mathcal{X} \times \mathcal{W})$ is said to be an invariant distribution for (2.5) and then for (2.1) if it is invariant by $\mathcal{Q}$, that is,

$$\mathcal{Q}\mu = \mu.$$

However, the concept of uniqueness will be slightly different from the classical setting. Indeed, denote by $\mathcal{S}\mu$ the distribution of $(X_n^\mu)_{n\geq 0}$ when we realize $(X_n^\mu, (\Delta_{n+k})_{k\leq 0})_{n\geq 0}$ through the transition $\mathcal{Q}$ and with initial distribution μ. Then, we will speak of uniqueness of the invariant distribution up to the equivalence relation: $\mu \sim \nu \Longleftrightarrow \mathcal{S}\mu = \mathcal{S}\nu$.

Moreover, here uniqueness will be deduced from the coupling procedure. There exist some results about uniqueness using ergodic theory, like in [13], but they will be not outlined here.

2.3. Preliminary tool: A Toeplitz type operator

The moving-average representation of the Gaussian sequence $(\Delta_n)_{n\in\mathbb{Z}}$ naturally leads us to define an operator related to the coefficients $(a_k)_{k\geq 0}$. First, set

$$\ell_a\left(\mathbb{Z}^-, \mathbb{R}^d\right) := \left\{ w \in \left(\mathbb{R}^d\right)^{\mathbb{Z}^-} \,\middle|\, \forall k \geq 0, \left|\sum_{l=0}^{+\infty} a_l w_{-k-l}\right| < +\infty \right\}$$

and define $\mathbf{T_a}$ on $\ell_a(\mathbb{Z}^-, \mathbb{R}^d)$ by

$$\mathbf{T_a}(w) = \left(\sum_{l=0}^{+\infty} a_l w_{-k-l}\right)_{k\geq 0}. \tag{2.7}$$

Due to the Cauchy–Schwarz inequality, we can check that for instance $\ell^2(\mathbb{Z}^-, \mathbb{R}^d) := \{w \in (\mathbb{R}^d)^{\mathbb{Z}^-} \mid \sum_{k=0}^{+\infty} |w_{-k}|^2 < +\infty\}$ is included in $\ell_a(\mathbb{Z}^-, \mathbb{R}^d)$ due to the assumption $\sum_{k\geq 0} a_k^2 < +\infty$. This Toeplitz type operator $\mathbf{T_a}$ links $(\Delta_n)_{n\in\mathbb{Z}}$ to $(\xi_n)_{n\in\mathbb{Z}}$. The following proposition spells out the reverse operator.

Proposition 2.1. *Let $\mathbf{T_b}$ be the operator defined on $\ell_b(\mathbb{Z}^-, \mathbb{R}^d)$ in the same way as $\mathbf{T_a}$ but with the following sequence $(b_k)_{k\geq 0}$*

$$b_0 = \frac{1}{a_0} \quad and \quad \forall k \geq 1, \qquad b_k = -\frac{1}{a_0}\sum_{l=1}^{k} a_l b_{k-l}. \tag{2.8}$$

Then,

$$\forall w \in \ell_a\left(\mathbb{Z}^-, \mathbb{R}^d\right), \qquad \mathbf{T_b}\left(\mathbf{T_a}(w)\right) = w \quad and \quad \forall w \in \ell_b\left(\mathbb{Z}^-, \mathbb{R}^d\right), \qquad \mathbf{T_a}\left(\mathbf{T_b}(w)\right) = w$$

that is $\mathbf{T_b} = \mathbf{T_a}^{-1}$ and $\ell_b(\mathbb{Z}^-, \mathbb{R}^d) = \mathbf{T_a}(\ell_a(\mathbb{Z}^-, \mathbb{R}^d))$.

Proof. Let $w \in \ell_a(\mathbb{Z}^-, \mathbb{R}^d)$. Then let $n \geq 0$,

$$\left(\mathbf{T_b}\big(\mathbf{T_a}(w)\big)\right)_{-n} = \sum_{k=0}^{+\infty} b_k \sum_{l=0}^{+\infty} a_l w_{-n-k-l} = \sum_{k=0}^{+\infty} \sum_{i=k}^{+\infty} b_k a_{i-k} w_{-n-i}$$

$$= \sum_{i=0}^{+\infty} \underbrace{\left(\sum_{k=0}^{i} b_k a_{i-k} \right)}_{\substack{=0 \text{ except} \\ \text{for } i=0}} w_{-n-i} = w_{-n}.$$

We show in the same way that for $w \in \ell_b(\mathbb{Z}^-, \mathbb{R}^d)$, we have $\mathbf{T_a}(\mathbf{T_b}(w)) = w$. $\qquad\square$

Remark 2.2. The sequence $(b_k)_{k \geq 0}$ is of first importance in the sequel. The sketch of the proof of Theorem 2.1 will use an important property linked to the sequence $(b_k)_{k \geq 0}$: if two sequences u and v are such that

$$\forall n \geq 1, \qquad u_n = \sum_{k=0}^{n-1} a_k v_{n-k}$$

then,

$$\forall n \geq 1, \qquad v_n = \sum_{k=0}^{n-1} b_k u_{n-k}.$$

This reverse identity and the asymptotic behavior of $(b_k)_{k \geq 0}$ play a significant role in the computation of the rate of convergence.

The following section is devoted to outline assumptions on $(a_k)_{k \geq 0}$ and $(b_k)_{k \geq 0}$ and then on F to get the main result.

2.4. Assumptions and general theorem

First of all, let us introduce assumptions on $(a_k)_{k \geq 0}$ and $(b_k)_{k \geq 0}$. All along the paper, we will switch between two types of assumptions called respectively the *polynomial case* and the *exponential case*.

Hypothesis ($\mathbf{H_{poly}}$). *The following conditions hold,*

- *there exist $\rho, \beta > 1/2$ and $C_\rho, C_\beta > 0$ such that*

$$\forall k \geq 0, \qquad |a_k| \leq C_\rho (k+1)^{-\rho} \quad and \quad \forall k \geq 0, \qquad |b_k| \leq C_\beta (k+1)^{-\beta}$$

- *there exist $\kappa \geq \rho + 1$ and $C_\kappa > 0$ such that*

$$\forall k \geq 0, \qquad |a_k - a_{k+1}| \leq C_\kappa (k+1)^{-\kappa}.$$

Hypothesis ($\mathbf{H_{exp}}$). *There exist $\lambda, \zeta > 0$ and $C_\lambda, C_\zeta > 0$ such that,*

$$\forall k \geq 0, \qquad |a_k| \leq C_\lambda e^{-\lambda k} \quad and \quad \forall k \geq 0, \qquad |b_k| \leq C_\zeta e^{-\zeta k}.$$

Remark 2.3.

$\triangleright$ ($\mathbf{H_{poly}}$) and ($\mathbf{H_{exp}}$) are general parametric hypothesis which apply to a large class of Gaussian driven dynamics. These assumptions implicitly involve the covariance function of the noise process $(c(k))_{k \in \mathbb{N}}$ (see Remark 2.1): there exists $\tilde{C}_\lambda > 0$ and for all $\varepsilon \in [0, \rho]$ such that $\rho + \varepsilon > 1$, there exists $\tilde{C}_{\rho, \varepsilon} > 0$ such that

$$\forall k \geq 0, \qquad |c(k)| \leq \begin{cases} \tilde{C}_{\rho, \varepsilon}(k+1)^{-\rho + \varepsilon} & \text{under } (\mathbf{H_{poly}}), \\ \tilde{C}_\lambda e^{-\lambda k} & \text{under } (\mathbf{H_{exp}}). \end{cases}$$

$\triangleright$ ($\mathbf{H_{poly}}$) and ($\mathbf{H_{exp}}$) also involve the coefficients appearing in the reverse Toeplitz operator $\mathbf{T_a}^{-1}$ (see Proposition 2.1). Even though $(a_k)_{k \geq 0}$ and $(b_k)_{k \geq 0}$ are related by (2.8), there is no general rule which connects ρ and β. This fact will be highlighted in Section 2.7. Moreover, for the sake of clarity, we have chosen to state our main result when (a_k) and (b_k) belong to the same family of asymptotic decay rate.

$\triangleright$ Due to the strategy of the proof (coalescent coupling in a non Markovian setting), we also need a bound on the discrete derivative of $(a_k)_{k \geq 0}$.

Let us now introduce some assumptions on the function F which defines the dynamics (2.1). Throughout this paper $F : \mathcal{X} \times \mathbb{R}^d \to \mathcal{X}$ is a continuous function and the following hypothesis ($\mathbf{H_1}$) and ($\mathbf{H_2}$) are satisfied.

Hypothesis ($\mathbf{H_1}$). *There exists a continuous function $V : \mathcal{X} \to \mathbb{R}_+^*$ satisfying $\lim_{|x| \to +\infty} V(x) = +\infty$ and $\exists \gamma \in (0, 1)$ and $C > 0$ such that for all $(x, w) \in \mathcal{X} \times \mathbb{R}^d$,*

$$V\big(F(x, w)\big) \leq \gamma V(x) + C\big(1 + |w|\big). \tag{2.9}$$

Remark 2.4.

$\triangleright$ As we will see in Section 3, this condition ensures the existence of a Lyapunov function V and then of an invariant distribution. Such a type of assumption also appears in the litterature of discrete Markov Chains (see, e.g., equation (7) in [8]) but in an integrated form. More precisely, in our non-Markovian setting, a pathwise control is required to ensure some control on the moments of the trajectories before the successful coupling procedure (this fact is detailed in Section 6.2).

$\triangleright$ This assumption is also fulfilled if we have a function V with $V(F(x, w)) \leq \gamma V(x) + C(1 + |w|^p)$ for a given $p > 1$ instead of (2.9) since the function $V^{1/p}$ satisfies ($\mathbf{H_1}$) (with the help of the elementary inequality $|u + v|^{1/p} \leq |u|^{1/p} + |v|^{1/p}$ when $p > 1$).

We define $\tilde{F} : \mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathcal{X}$ by $\tilde{F}(x, u, y) = F(x, u + y)$. We assume that $\tilde{F}$ satisfies the following conditions:

Hypothesis (H₂). *Let $K > 0$. We assume that there exists $\tilde{K} > 0$ such that for every $\mathbf{x} :=$ (x, x', y, y') in $B(0, K)^4$, there exist $\Lambda_{\mathbf{x}} : \mathbb{R}^d \to \mathbb{R}^d$, $M_K > 0$ and $C_{\tilde{K}}$ such that the following holds*

- $\Lambda_{\mathbf{x}}$ *is a bijection from $\mathbb{R}^d$ to $\mathbb{R}^d$. Moreover, it is a $\mathcal{C}^1$-diffeomorphism between two open sets U and D such that $\mathbb{R}^d \backslash U$ and $\mathbb{R}^d \backslash D$ are negligible sets.*
- *for all $u \in B(0, \tilde{K})$,*

$$\tilde{F}(x, u, y) = \tilde{F}(x', \Lambda_{\mathbf{x}}(u), y') \quad \text{and} \tag{2.10}$$

$$\left| \det\big(J_{\Lambda_{\mathbf{x}}}(u)\big) \right| \geq C_{\tilde{K}}. \tag{2.11}$$

- *for all $u \in \mathbb{R}^d$,*

$$\left| \Lambda_{\mathbf{x}}(u) - u \right| \leq M_K. \tag{2.12}$$

Remark 2.5.

▷ Let us make a few precisions on the arguments of $\tilde{F}$: x is the position of the process, u the increment of the innovation process and y is related to the past of the process (see next item for more details). The boundary $C_{\tilde{K}}$ and M_K are independent from x, x', y and y'. This assumption can be viewed as a kind of controlability assumption in the following sense: the existence of $\Lambda_{\mathbf{x}}$ leads to the coalescence of the positions by (2.10). Rephrased in terms of our coalescent coupling strategy, this *ad hoc* assumption is required to achieve the first step. More precisely, as announced in the introduction, we take two trajectories (X^1, Δ^1) and (X^2, Δ^2) following (2.1) and we want to stick X^1 and X^2 at a given time $n + 1$. Through the function $\Lambda_{\mathbf{x}}$ in (H₂), we can build a couple of Gaussian innovations $(\xi_{n+1}^1, \xi_{n+1}^2)$ with marginal distribution $\mathcal{N}(0, I_d)$ to achieve this goal (with lower-bounded probability), so that: $X_{n+1}^1 = X_{n+1}^2$ which is equivalent to $\tilde{F}(x, \xi_{n+1}^1, y) = \tilde{F}(x', \xi_{n+1}^2, y')$ with $(x, x', y, y') = (X_n^1, X_n^2, \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^1, \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^2)$ (see Section 4.2.1).
▷ Assumption (H₂) can be applied to a large class of functions F, as for example: $F(x, w) = f(b(x) + \sigma(x)w)$ where σ is continuously invertible and σ^{-1} and b are continuous functions (we do not need any assumption on f as we will see in the companion paper [23], Remark 1.2). Actually, Condition (2.10) can be obtained through an application of the implicit function theorem: if we assume that there exists a point $(0, u'_{\mathbf{x}})$ such that $\tilde{F}(x, 0, y) = \tilde{F}(x', u'_{\mathbf{x}}, y')$ and denote by $G_{\mathbf{x}} : (u, u') \mapsto \tilde{F}(x, u, y) - \tilde{F}(x', u' + u'_{\mathbf{x}}, y')$, then if $(\partial_{u'} G_{\mathbf{x}})(0, 0) \neq 0$, the implicit function theorem yields (2.10). As we will see in Section 2.5, Condition (2.12) can be also easily fulfilled (see proof of Theorem 2.2).

We are now in position to state our main result.

Theorem 2.1. *Assume (H₁) and (H₂). Then,*

(i) *There exists an invariant distribution $\mu_\star$ associated to (2.1).*
(ii) *Assume that (H_poly) is true with $\rho, \beta > 1/2$ and $\rho + \beta > 3/2$. Then, uniqueness holds for the invariant distribution $\mu_\star$. Furthermore, for every initial distribution μ_0 for which*

$\int_{\mathcal{X}} V(x) \Pi^*_{\mathcal{X}} \mu_0(dx) < +\infty$ and for all $\varepsilon > 0$, there exists $C_\varepsilon > 0$ such that

$$\left\| \mathcal{L}\left((X^{\mu_0}_{n+k})_{k \geq 0}\right) - \mathcal{S}\mu_\star \right\|_{\mathrm{TV}} \leq C_\varepsilon n^{-(v(\beta, \rho) - \varepsilon)}.$$

where the function v is defined by

$$v(\beta, \rho) = \sup_{\alpha \in (\frac{1}{2} \vee (\frac{3}{2} - \beta), \rho)} \min\{1, 2(\rho - \alpha)\}\big(\min\{\alpha, \beta, \alpha + \beta - 1\} - 1/2\big).$$

(iii) *Assume that* ($\mathbf{H_{exp}}$) *is true, then uniqueness holds for the invariant distribution* $\mu_\star$. *Furthermore, for every initial distribution* μ_0 *for which* $\int_{\mathcal{X}} V(x) \Pi^*_{\mathcal{X}} \mu_0(dx) < +\infty$ *and for all* $p > 0$, *there exists* $C_p > 0$ *such that*

$$\left\| \mathcal{L}\left((X^{\mu_0}_{n+k})_{k \geq 0}\right) - \mathcal{S}\mu_\star \right\|_{\mathrm{TV}} \leq C_p n^{-p}.$$

Remark 2.6.

▷ In view of Theorem 2.1(iii), one can wonder if we could obtain exponential or subexponential rates of convergence in this case. We focus on this question in Remark 6.1.

▷ The rates obtained in Theorem 2.1 hold for a large class of dynamics. This generality implies that the rates are not optimal in all situations. In particular, when F have "nice" properties an adapted method could lead to better rates. For example, let us mention the particular case where the dynamical system (2.1) is reduced to: $X_{n+1} = A X_n + \sigma \Delta_{n+1}$ where A and σ are some given matrices. As for (fractional) Ornstein–Uhlenbeck processes in a continuous setting, the study of linear dynamics can be achieved with specific methods. Here, the sequence X benefits of a Gaussian structure so that the convergence in distribution could be studied through the covariance of the process. One can also remark that since for two paths X and $\tilde{X}$ built with the same noise, we have: $\tilde{X}_{n+1} - X_{n+1} = A(\tilde{X}_n - X_n)$, a simple induction leads to $\mathbb{E}[|\tilde{X}_n - X_n|^2] \leq |A^n|^2 \mathbb{E}[|\tilde{X}_0 - X_0|^2]$. So without going into the details, if $\rho(A) := \lim_{n \to +\infty} |A^n|^{1/n} < 1$, such bounds lead to geometric rates of convergence in Wasserstein distance and also in total variation distance (on this topic, see, e.g., [21]).

In the following subsection, we test the assumptions of our main result Theorem 2.1 (especially ($\mathbf{H_1}$) and ($\mathbf{H_2}$)) on the Euler scheme of SDEs like (1.1).

2.5. The Euler scheme

Recall that $\mathcal{X} = \mathbb{R}^d$. In this subsection, set

$$\begin{aligned} F_h: \quad & \mathcal{X} \times \mathbb{R}^d \to \mathcal{X}, \\ & (x, w) \mapsto x + h b(x) + \sigma(x) w. \end{aligned} \tag{2.13}$$

where $h > 0$, $b : \mathcal{X} \to \mathcal{X}$ is continuous and $\sigma : \mathcal{X} \to \mathcal{M}_d(\mathbb{R})$ is a continuous and bounded function on $\mathcal{X}$. For all $x \in \mathcal{X}$, we suppose that $\sigma(x)$ is invertible and we denote by $\sigma^{-1}(x)$ the

inverse. Moreover, we assume that σ^{-1} is a continuous function and that b satisfies a Lyapunov type assumption that is:

(L1) $\exists C > 0$ such that

$$\forall x \in \mathcal{X}, \qquad |b(x)| \leq C(1 + |x|). \tag{2.14}$$

(L2) $\exists \tilde{\beta} \in \mathbb{R}$ and $\tilde{\alpha} > 0$ such that

$$\forall x \in \mathcal{X}, \qquad \langle x, b(x) \rangle \leq \tilde{\beta} - \tilde{\alpha}|x|^2. \tag{2.15}$$

Remark 2.7. This function F_h corresponds to the Euler scheme associated to SDEs like (1.1). The conditions on the function b are classical to get existence of invariant distribution.

In this setting the function $\tilde{F}_h$ (introduced in Hypothesis **(H₂)**) is given by

$$\tilde{F}_h: \quad \mathcal{X} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathcal{X},$$

$$(x, u, y) \mapsto x + hb(x) + \sigma(x)(u + y).$$

Theorem 2.2. *Let $h > 0$. Let F_h be the function defined above. Assume that $b : \mathcal{X} \to \mathcal{X}$ is a continuous function satisfying **(L1)** and **(L2)** and $\sigma : \mathcal{X} \to \mathcal{M}_d(\mathbb{R})$ is a continuous and bounded function such that for all $x \in \mathcal{X}$, $\sigma(x)$ is invertible and $x \mapsto \sigma^{-1}(x)$ is a continuous function. Then, **(H₁)** and **(H₂)** hold for F_h as soon as $0 < h < \min\{\sqrt{1 + \frac{\tilde{\alpha}}{2C^2}} - 1, \frac{1}{\tilde{\alpha}}\}$ where $\tilde{\alpha}$ and C are given by **(L1)** and **(L2)**.*

Proof. For the sake of conciseness, the proof is detailed in the supplementary material [23], Section 1. Regarding **(H₁)**, it makes use of ideas developed in [5,6]. For **(H₂)**, the construction of $\Lambda_{\mathbf{x}}$ is explicit. The idea is to take $\Lambda_{\mathbf{x}}(u) := \sigma^{-1}(x')\sigma(x)u + \sigma^{-1}(x')(x - x' + h(b(x) - b(x'))) + \sigma^{-1}(x')\sigma(x)y - y'$ inside $B(0, K)$ (which ensures (2.10) with $\tilde{K} = K$), to set $\Lambda_{\mathbf{x}}(u) := u$ outside an other ball $B(0, K_1)$ with a well-chosen K_1 (which almost gives (2.12)) and to extend $\Lambda_{\mathbf{x}}$ into $\mathbb{R}^d$ by taking into account the various hypothesis on $\Lambda_{\mathbf{x}}$. □

Remark 2.8. The coefficient σ is assumed to be bounded but we can relax a bit this assumption: if $|\sigma(x)| \leq C(1 + |x|^\kappa)$ for some $\kappa \in (0, 1)$, then Theorem 2.2 is true again. We provide in [23], Remark 1.2, the necessary adjustments of the proof to get this result.

The two following subsections are devoted to outline examples of sequences which satisfy hypothesis **(H_{exp})** or **(H_{poly})**. In particular, Section 2.7 includes the case where the process $(\Delta_n)_{n \in \mathbb{Z}}$ corresponds to fractional Brownian motion increments.

2.6. Two explicit cases which satisfy (H_{exp})

First, let us mention the explicit exponential case with the following definition for the sequence $(a_k)_{k \geq 0}$

$$a_0 = 1 \quad \text{and} \quad \forall k \in \mathbb{N}^*, \qquad a_k = C_a e^{-\lambda k} \tag{2.16}$$

with $C_a \in \mathbb{R}$. Let us recall that $b_0 = 1$ (since $a_0 = 1$) and for all $k \geq 1$, we can get the following general expression of b_k (see [23], Section 2, for more details):

$$b_k = \sum_{p=1}^{k} \frac{(-1)^p}{a_0^{p+1}} \left(\sum_{\substack{k_1,\ldots,k_p \geq 1 \\ k_1+\cdots+k_p=k}} \prod_{i=1}^{p} a_{k_i} \right). \tag{2.17}$$

A classical combinatorial argument shows that $\sharp\{(k_1,\ldots,k_p) \in \mathbb{N}^* \mid k_1 + \cdots + k_p = k\} = \binom{k-1}{p-1}$. As a consequence, when the sequence $(a_k)_{k\geq0}$ is defined by (2.16), we can easily prove that for $k \geq 1$,

$$b_k = -C_a(1 - C_a)^{k-1}e^{-\lambda k}.$$

Hence, to satisfy ($\mathbf{H_{exp}}$), we only need C_a to be such that $\zeta := \lambda - \ln|1 - C_a| > 0$ and then for all $k \in \mathbb{N}^*$, we get

$$|b_k| \leq C_b e^{-\zeta k} \tag{2.18}$$

with $C_b > 0$ a constant depending on C_a.

Remark 2.9.

$\triangleright$ In this setting where everything is computable, it's interesting to see that the asymptotic decrease of the sequence $(|b_k|)$ is not only related to the one of the sequence $(|a_k|)$. For instance, if we take $C_a < 0$, the simple fact that $a_0 > 0$ and $a_k < 0$ for all $k > 0$ makes (b_k) diverge to $+\infty$ and nevertheless, $(|a_k|)$ decreases to 0 at an exponential rate.

$\triangleright$ If we take $C_a = 1$, we can reduce $(\Delta_n)_{n\in\mathbb{Z}}$ to the following induction: $\forall n \in \mathbb{Z}$, $\Delta_{n+1} = \xi_{n+1} + e^{-\lambda}\Delta_n$.

Let us finally consider finite moving averages, that is, when $a_k = 0$ for all $k > m$ (for some $m \geq 1$). In this setting, one can expect ($\mathbf{H_{exp}}$) to be satisfied since the memory is finite. This is actually the case when the finite moving average is invertible, namely: $P(\lambda) := 1 + \sum_{k=1}^{m} a_k \lambda^k$ has all its roots outside the unit circle (see [2] Theorem 3.1.2). In that case, there exists $\lambda \in \mathbb{C}$ such that $|\lambda| > 1$ and $\frac{1}{P(\lambda)} = \sum_{k=0}^{+\infty} b_k \lambda^k < +\infty$ (to get more details on this equality, see [23], Section 2). Then, there exists $C > 0$ such that for all $k \geq 0$,

$$|b_k| \leq C \left(\frac{1}{|\lambda|} \right)^k$$

and finally ($\mathbf{H_{exp}}$) holds true.

When the invertibility is not fulfilled, the situation is more involved but ($\mathbf{H_{exp}}$) is still true up to another Wold decomposition. More precisely, one can find another white noise ξ and another set of coefficients a_k such that the invertibility holds true, on this topic see, for example, [2] Proposition 4.4.2.

2.7. Polynomial case: From a general class of examples to the fractional case

A natural example of Gaussian sequence $(\Delta_n)_{n\in\mathbb{Z}}$ which leads to polynomial memory is to choose $a_k = (k+1)^{-\rho}$ for a given $\rho > 1/2$. In that case, we have the following result.

Proposition 2.2. *Assume* (**H$_1$**) *and* (**H$_2$**). *Let $\rho > 1/2$. If for all $k \geq 0$, $a_k = (k+1)^{-\rho}$, then we have $|b_k| \leq (k+1)^{-\rho}$. Moreover, if $\rho > 3/4$ Theorem* 2.1(ii) *holds with the rate*

$$v(\rho, \rho) = \begin{cases} 2(\rho - 3/4)^2 & \text{if } \rho \in (3/4, 1], \\ \dfrac{1}{2}(\rho - 1/2)^2 & \text{if } \rho > 1. \end{cases}$$

Remark 2.10. The main novelty here comes from the proof of the inequality $|b_k| \leq (k+1)^{-\rho}$ for all $k \geq 0$ which is outlined in the supplementary material [23], Section 3, and which is based on results of [10]. The key argument in this proof is the log-convexity property of the sequence $(a_k)_{k\in\mathbb{N}}$, which means that for all $k \in \mathbb{N}$, $a_k \geq 0$ and for $k \geq 1$, $a_k^2 - a_{k-1}a_{k+1} \leq 0$.

With Proposition 2.2 in hand, the purpose of the remainder of this section is to focus on Gaussian sequences of *fractional type*, that is, when the sequence $(a_k)_{k\geq 0}$ satisfies:

$$\forall k \geq 0, \qquad |a_k| \leq C_\rho (k+1)^{-\rho} \quad \text{and} \quad |a_k - a_{k+1}| \leq \tilde{C}_\rho (k+1)^{-(\rho+1)} \tag{2.19}$$

with $\rho := 3/2 - H$ and $H \in (0, 1)$ is the so-called Hurst parameter. In particular, through this class of sequences, we provide an explicit example which shows that computing the rate of convergence of the sequence $(b_k)_{k\geq 0}$ is a hard task and strongly depends on the variations of $(a_k)_{k\geq 0}$. Condition (2.19) includes both cases of Proposition 2.2 with $\rho := 3/2 - H$ and when $(\Delta_n)_{n\in\mathbb{Z}}$ corresponds to the fractional Brownian motion (fBm) increments (as we will see below), we therefore decided to use the terminology "*fractional type*". Recall that a d-dimensional fBm with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process $(B_t)_{t\geq 0}$ with stationary increments satisfying

$$\forall t, s \geq 0, \forall i, j \in \{1, \ldots, d\}, \qquad \mathbb{E}\big[\big(B_t^i - B_s^i\big)\big(B_t^j - B_s^j\big)\big] = \delta_{ij}|t - s|^{2H}.$$

In our discrete-time setting, we are thus concerned by the long time behavior of (2.1) if we take for $h > 0$

$$(\Delta_n)_{n\in\mathbb{Z}} = (B_{nh} - B_{(n-1)h})_{n\in\mathbb{Z}} \tag{2.20}$$

which is a stationary Gaussian sequence. It can be realized through a moving average representation with coefficients $(a_k^H)_{k\geq 0}$ defined by (see [20]):

$$a_0^H = h^H \kappa(H) 2^{1/2-H} \quad \text{and}$$

$$\text{for } k \geq 1, \qquad a_k^H = h^H \kappa(H)\left(\left(k + \frac{1}{2}\right)^{H-1/2} - \left(k - \frac{1}{2}\right)^{H-1/2}\right), \tag{2.21}$$

where

$$\kappa(H) = \frac{\sqrt{\sin(\pi H)\Gamma(2H+1)}}{\Gamma(H+1/2)}.$$

One can easily check that $a_k^H \underset{k \to +\infty}{\sim} C_{h,H}(k+1)^{-(3/2-H)}$ and $|a_k^H - a_{k+1}^H| \le C'_{h,H}(k+1)^{-(5/2-H)}$.

Hence $(a_k^H)_{k \ge 0}$ is of *fractional type* in the sense of (2.19). Now, the question is: how does the corresponding (b_k^H) behave? When H belongs to $(0, 1/2)$, only a_0^H is positive and then (a_k^H) is not log-convex. Therefore, we cannot use this property to get the asymptotic behavior of (b_k^H) as we did in Proposition 2.2. However, thanks to simulations (see Figure 1(a) and 1(b)), we conjectured and we proved Proposition 2.3.
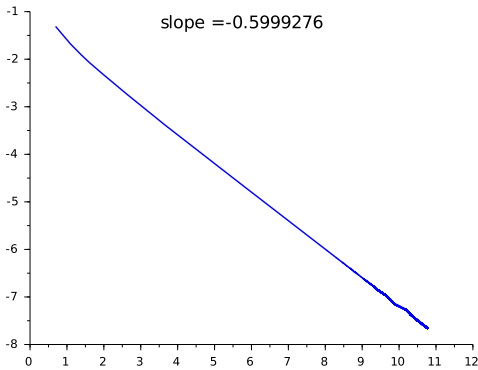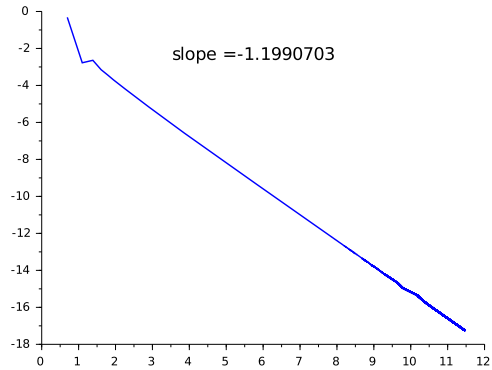

(a) parameter: $H = 0.1$
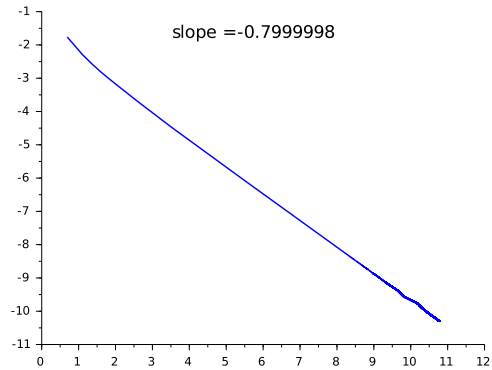
(b) parameter: $H = 0.3$
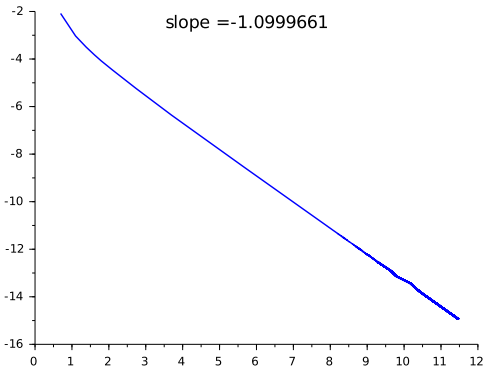
(c) parameter: $H = 0.6$

(d) parameter: $H = 0.9$

Figure 1. $(\log |b_k^H|)$ according to $(\log(k+1))$ with different Hurst parameters H.

Proposition 2.3. *There exists $C''_{h,H} > 0$ such that for all $H \in (0, 1/2)$*

$$\forall k \geq 0, \qquad \left| b_k^H \right| \leq C''_{h,H}(k+1)^{-(H+1/2)}. \tag{2.22}$$

Then, if we assume (**H$_1$**) *and* (**H$_2$**), *Theorem* 2.1(ii) *holds with the rate*

$$v(\rho, 2 - \rho) = v(3/2 - H, H + 1/2) = \begin{cases} H(1 - 2H) & \text{if } H \in (0, 1/4], \\ \dfrac{1}{8} & \text{if } H \in (1/4, 1/2). \end{cases}$$

Remark 2.11. For the sake of conciseness, we provided the details of the proof in the companion paper [23], Section 4. The proof of (2.22) is based on a technical induction which involves very sharp inequalities.

In Proposition 2.2 (with $\rho = 3/2 - H$ and $H \in (0, 1/2)$) and in the above proposition, dealing with the same order of memory, we get really different orders of rate of convergence: one easily checks that for all $H \in (0, 1/2)$, $v(3/2 - H, H + 1/2) < v(3/2 - H, 3/2 - H)$. Finally, we have seen that managing the asymptotic behavior of (b_k) is both essential and a difficult task.

To end this section, let us briefly discuss on the specific statements on fBm increments and compare with the continuous time setting (see [7,9,12]). For this purpose, we introduce the following conjecture (based on simulations, see Figure 1(c) and 1(d)) when H belongs to $(1/2, 1)$:

Conjecture. *There exists $C''_{h,H} > 0$ and $\beta_H > 1$ such that*

$$\forall k \geq 0, \qquad \left| b_k^H \right| \leq C''_{h,H}(k+1)^{-\beta_H}. \tag{2.23}$$

Remark 2.12. We do not have a precise idea of the expression of β_H with respect to H. But, we can note that if $\rho < 1$ and $\beta > 1$ in (**H$_{\text{poly}}$**), then the rate of convergence in Theorem 2.1 is $v(\rho, \beta) = \frac{(2\rho - 1)^2}{8}$ and does not depend on β. Hence, if $H \in (1/2, 1)$, $\rho = 3/2 - H$ and $\beta_H > 1$, we fall into this case and then the dependence of β_H in terms of H does not matter.

If the previous conjecture is true, we get the following rate of convergence for $H \in (1/2, 1)$ in Theorem 2.1:

$$v(\rho, \beta_H) = v(3/2 - H, \beta_H) = \frac{(1 - H)^2}{2}.$$

Then, when H belongs to $(0, 1/2)$ Proposition 2.3 gives exactly the same rate of convergence obtained in [7,12]. However, when $H > 1/2$ it seems that we will get a smaller rate than in a continuous time setting. The reason for this may be that Theorem 2.1 is a result with quite general hypothesis on the Gaussian noise process $(\Delta_n)_{n \in \mathbb{Z}}$. In the case of fBm increments, the moving average representation is explicit. Hence, we may use a more specific approach and significantly closer to Hairer's, especially with regard to Step 2 in the coupling method (see Section 5.2.2) by not exploiting the technical lemma 5.3 for instance. This seems to be a right track in order to improve our results on this precise example.

We are now ready to begin the proof of Theorem 2.1. In Section 3, we establish the first part of the theorem, that is, (i). Then, in Section 4 we explain the scheme of coupling before implementing this strategy in Sections 5 and 6. Finally, in Section 7, we achieve the proof of (ii) and (iii) of Theorem 2.1.

3. Existence of invariant distribution

Denote by $\mathbb{P}_w$ the law of $(\Delta_k)_{k \leq 0}$. Since $(\Delta_n)_{n \in \mathbb{Z}}$ is stationary we immediately get the following property:

Property 3.1. If a measure $\mu \in \mathcal{M}_1(\mathcal{X} \times \mathcal{W})$ is such that $\Pi^*_{\mathcal{W}}\mu = \mathbb{P}_w$ then $\Pi^*_{\mathcal{W}}\mathcal{Q}\mu = \mathbb{P}_w$.

We can now define the notion of Lyapunov function.

Definition 3.1. A function $\psi : \mathcal{X} \to [0, +\infty)$ is called a Lyapunov function for $\mathcal{Q}$ if ψ is continuous and if the following holds:

(i) $\psi^{-1}([0, a])$ is compact for all $a \in [0, +\infty)$.
(ii) $\exists \beta > 0$ and $\alpha \in (0, 1)$ such that:

$$\int_{\mathcal{X} \times \mathcal{W}} \psi(x) \mathcal{Q}\mu(\mathrm{d}x, \mathrm{d}w) \leq \beta + \alpha \int_{\mathcal{X}} \psi(x) \big(\Pi^*_{\mathcal{X}}\mu\big)(\mathrm{d}x)$$

for all $\mu \in \mathcal{M}_1(\mathcal{X} \times \mathcal{W})$ such that $\Pi^*_{\mathcal{W}}\mu = \mathbb{P}_w$ and $\int_{\mathcal{X}} \psi(x)(\Pi^*_{\mathcal{X}}\mu)(\mathrm{d}x) < +\infty$.

The following result ensures the existence of invariant distribution for $\mathcal{Q}$.

Theorem 3.1. *If there exists a Lyapunov function ψ for $\mathcal{Q}$, then $\mathcal{Q}$ has at least one invariant distribution $\mu_\star$, in other words $\mathcal{Q}\mu_\star = \mu_\star$.*

A detailed proof of this result is given in the supplementary material [23], Section 5. Finally, we get the first part (i) of Theorem 2.1 about the existence of an invariant distribution by setting $\psi := V$ (with V the function appearing in $(\mathbf{H_1})$) and by saying that ψ is a Lyapunov function for $\mathcal{Q}$.

4. General coupling procedure

We now turn to the proof of the main result of the paper, Theorem 2.1(ii) and (iii) about the convergence in total variation. This result is based on a coupling method first introduced in [12], but also used in [9] and [7], in a continuous time framework. The coupling strategy is slightly different in our discrete context, the following part is devoted to explain this procedure.

4.1. Scheme of coupling

Let $(\Delta_n^1)_{n\in\mathbb{Z}}$ and $(\Delta_n^2)_{n\in\mathbb{Z}}$ be two stationary and purely non-deterministic Gaussian sequences with the following moving average representations

$$
\begin{cases}
\Delta_n^1 = \displaystyle\sum_{k=0}^{+\infty} a_k \xi_{n-k}^1, \\
\Delta_n^2 = \displaystyle\sum_{k=0}^{+\infty} a_k \xi_{n-k}^2
\end{cases}
$$

with

$$
\begin{cases}
(a_k)_{k\geq 0} \in \mathbb{R}^{\mathbb{N}} \text{ such that } a_0 = 1 \text{ and } \displaystyle\sum_{k=0}^{+\infty} a_k^2 < +\infty, \\
\xi^i := \left(\xi_k^i\right)_{k\in\mathbb{Z}} \text{ an i.i.d sequence such that } \xi_1^i \sim \mathcal{N}(0, I_d) \text{ for } i = 1, 2.
\end{cases}
\tag{4.1}
$$

We denote by (X^1, X^2) the solution of the system:

$$
\begin{cases}
X_{n+1}^1 = F\left(X_n^1, \Delta_{n+1}^1\right), \\
X_{n+1}^2 = F\left(X_n^2, \Delta_{n+1}^2\right)
\end{cases}
\tag{4.2}
$$

with initial conditions $(X_0^1, (\Delta_k^1)_{k\leq 0})$ and $(X_0^2, (\Delta_k^2)_{k\leq 0})$. We assume that $(X_0^2, (\Delta_k^2)_{k\leq 0}) \sim \mu_\star$ where $\mu_\star$ denotes a fixed invariant distribution associated to (2.1). The previous section ensures that such a measure exists. We define the natural filtration associated to (4.2) by

$$
(\mathcal{F}_n)_{n\in\mathbb{N}} = \left(\sigma\left(\left(\xi_k^1\right)_{k\leq n}, \left(\xi_k^2\right)_{k\leq n}, X_0^1, X_0^2\right)\right)_{n\geq 0}.
$$

To lower the "weight of the past" at the beginning of the coupling procedure, we assume that a.s,

$$
\left(\Delta_k^1\right)_{k\leq 0} = \left(\Delta_k^2\right)_{k\leq 0}
$$

which is actually equivalent to assume that a.s $(\xi_k^1)_{k\leq 0} = (\xi_k^2)_{k\leq 0}$ since the invertible Toeplitz operator defined in Section 2.3 links $(\Delta_k^i)_{k\leq 0}$ to $(\xi_k^i)_{k\leq 0}$ for $i = 1, 2$. Lastly, we denote by $(g_n)_{n\in\mathbb{Z}}$ and $(f_n)_{n\in\mathbb{Z}}$ the random variable sequences defined by

$$
\xi_{n+1}^1 = \xi_{n+1}^2 + g_n \quad \text{and} \quad \Delta_{n+1}^1 = \Delta_{n+1}^2 + f_n.
\tag{4.3}
$$

They respectively represent the "drift" between the underlying noises (ξ_k^i) and the real noises (Δ_k^i). By assumption, we have $g_n = f_n = 0$ for $n < 0$.

Remark 4.1. From the moving average representations, we deduce immediately the following relation for all $n \geq 0$,

$$
f_n = \sum_{k=0}^{+\infty} a_k g_{n-k} = \sum_{k=0}^{n} a_k g_{n-k}.
\tag{4.4}
$$

The aim is now to build $(g_n)_{n \geq 0}$ and $(f_n)_{n \geq 0}$ in order to stick X^1 and X^2. We set

$$\tau_\infty = \inf\{n \geq 0 \mid X_k^1 = X_k^2, \forall k \geq n\}.$$

In a purely Markovian setting, when the paths coincide at time n then they remain stuck for all $k \geq n$ by putting the same innovation into both processes. Due to the memory this phenomenon cannot happen here. Hence, this involves a new step in the coupling scheme: try to keep the paths fathened together (see below).

Recall that $\mathcal{L}((X_k^2)_{k \geq n}) = \mathcal{S}\mu_\star$. The purpose of the coupling procedure is to bound the quantity $\mathbb{P}(\tau_\infty > n)$ since by a classical result we have

$$\left\| \mathcal{L}((X_k^1)_{k \geq n}) - \mathcal{S}\mu_\star \right\|_{\mathrm{TV}} \leq \mathbb{P}(\tau_\infty > n). \tag{4.5}$$

Hence, we realize the coupling after a series of trials which follows three steps:

* *Step 1*: Try to stick the positions at a given time with a "controlled cost".
* *Step 2*: (specific to non-Markov processes) Try to keep the paths fastened together.
* *Step 3*: If Step 2 fails, we wait long enough so as to allow Step 1 to be realized with a "controlled cost" and with a positive probability. During this step, we assume that $g_n = 0$.

More precisely, let us introduce some notations,

* Let $\tau_0 \geq 0$. We begin the first trial at time $\tau_0 + 1$, in other words we try to stick $X_{\tau_0+1}^1$ and $X_{\tau_0+1}^2$. Hence, we assume that

$$\forall n < \tau_0, \qquad g_n = f_n = 0. \tag{4.6}$$

* For $j \geq 1$, let τ_j denote the end of trial j. More specifically,
  ▷ If $\tau_j = +\infty$ for some $j \geq 1$, it means that the coupling tentative has been successful.
  ▷ Else, τ_j corresponds to the end of Step 3, that is $\tau_j + 1$ is the beginning of Step 1 of trial $j + 1$.

The real meaning of "controlled cost" will be clarified on Section 5.1. But the main idea is that at Step 1 of trial j, the "cost" is represented by the quantity $g_{\tau_{j-1}}$ that we need to build to get $X_{\tau_{j-1}+1}^1 = X_{\tau_{j-1}+1}^2$ with positive probability. Here the cost does not only depend on the positions at time τ_{j-1} but also on all the past of the underlying noises ξ^1 and ξ^2. Hence, we must have a control on $g_{\tau_{j-1}}$ in case of failure and to this end we have to wait enough during Step 3 before beginning a new attempt of coupling.

4.2. Coupling lemmas to achieve Steps 1 and 2

This section is devoted to establish coupling lemmas in order to build (ξ^1, ξ^2) during Step 1 and Step 2.

4.2.1. *Hitting step*

If we want to stick X^1 and X^2 at time $n + 1$, we need to build $(\xi_{n+1}^1, \xi_{n+1}^2)$ in order to get $F(X_n^1, \Delta_{n+1}^1) = F(X_n^2, \Delta_{n+1}^2)$ with positive probability, that is to get

$$F\left(X_n^1, \xi_{n+1}^1 + \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^1\right) = F\left(X_n^2, \xi_{n+1}^2 + \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^2\right)$$

$$\Longleftrightarrow \quad \tilde{F}\left(X_n^1, \xi_{n+1}^1, \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^1\right) = \tilde{F}\left(X_n^2, \xi_{n+1}^2, \sum_{k=1}^{+\infty} a_k \xi_{n+1-k}^2\right). \tag{4.7}$$

The following lemma will be the main tool to achieve this goal.

Lemma 4.1. *Let $K > 0$ and $\mu := \mathcal{N}(0, I_d)$. Under the controlability assumption* (**H₂**), *there exists $\tilde{K} > 0$ (given by* (**H₂**)), *such that for every $\mathbf{x} := (x, x', y, y')$ in $B(0, K)^4$, we can build a random variable (Z_1, Z_2) with values in $(\mathbb{R}^d)^2$ such that*

(i) $\mathcal{L}(Z_1) = \mathcal{L}(Z_2) = \mu$,
(ii) *there exists $\delta_{\tilde{K}} > 0$ depending only on $\tilde{K}$ such that*

$$\mathbb{P}\big(\tilde{F}(x, Z_1, y) = \tilde{F}(x', Z_2, y')\big) \geq \mathbb{P}\big(Z_2 = \Lambda_{\mathbf{x}}(Z_1), |Z_1| \leq \tilde{K}\big) \geq \delta_{\tilde{K}} > 0, \tag{4.8}$$

where $\Lambda_{\mathbf{x}}$ is the function given by hypothesis (**H₂**),
(iii) *there exists $M_K > 0$ given by* (**H₂**) *depending only on K such that*

$$\mathbb{P}\big(|Z_2 - Z_1| \leq M_K\big) = 1. \tag{4.9}$$

Proof. Let $\mathbf{x} := (x, x', y, y') \in B(0, K)^4$. First, let us denote by π_1 (resp. π_2) the projection from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ of the first (resp. the second) coordinate. Introduce the two following functions defined on $\mathbb{R}^d$

$$\Lambda_1 : \quad u_1 \mapsto \big(u_1, \Lambda_{\mathbf{x}}(u_1)\big),$$

$$\Lambda_2 : \quad u_2 \mapsto \big(\Lambda_{\mathbf{x}}^{-1}(u_2), u_2\big),$$

where $\Lambda_{\mathbf{x}}$ is the function given by (**H₂**). Now, we set

$$\mathbf{P}_1 = \frac{1}{2}\big(\Lambda_1^* \mu \wedge \Lambda_2^* \mu\big).$$

Let us find a simplest expression for $\mathbf{P}_1$. For every measurable function $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$, we have

$$\Lambda_1^* \mu(f) = \int_{\mathbb{R}^d} f\big(u_1, \Lambda_{\mathbf{x}}(u_1)\big) \mu(du_1) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(u_1, u_2) \delta_{\Lambda_{\mathbf{x}}(u_1)}(du_2) \mu(du_1)$$

and

$$\Lambda_2^* \mu(f) = \int_{\mathbb{R}^d} f\left(\Lambda_{\mathbf{x}}^{-1}(u_2), u_2\right) \mu(du_2)$$

$$= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f\left(\Lambda_{\mathbf{x}}^{-1}(u_2), u_2\right) \exp\left(-\frac{|u_2|^2}{2}\right) du_2$$

$$= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f\left(u_1, \Lambda_{\mathbf{x}}(u_1)\right) \exp\left(-\frac{|\Lambda_{\mathbf{x}}(u_1)|^2}{2}\right) \left|\det\left(J_{\Lambda_{\mathbf{x}}}(u_1)\right)\right| du_1$$

$$\left(\text{by setting } u_1 = \Lambda_{\mathbf{x}}^{-1}(u_2)\right)$$

$$= \int_{\mathbb{R}^d} f\left(u_1, \Lambda_{\mathbf{x}}(u_1)\right) \underbrace{\exp\left(\frac{|u_1|^2}{2} - \frac{|\Lambda_{\mathbf{x}}(u_1)|^2}{2}\right) \left|\det\left(J_{\Lambda_{\mathbf{x}}}(u_1)\right)\right|}_{=:D_{\Lambda_{\mathbf{x}}}(u_1)} \mu(du_1)$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} f(u_1, u_2) \delta_{\Lambda_{\mathbf{x}}(u_1)}(du_2) D_{\Lambda_{\mathbf{x}}}(u_1) \mu(du_1).$$

By construction, we then have

$$\mathbf{P}_1(du_1, du_2) = \frac{1}{2} \delta_{\Lambda_{\mathbf{x}}(u_1)}(du_2) \left(D_{\Lambda_{\mathbf{x}}}(u_1) \wedge 1\right) \mu(du_1). \tag{4.10}$$

Write $S(u_1, u_2) = (u_2, u_1)$ and denote by $\tilde{\mathbf{P}}_1$ the "symmetrized" non-negative measure induced by $\mathbf{P}_1$,

$$\tilde{\mathbf{P}}_1 = \mathbf{P}_1 + S^* \mathbf{P}_1. \tag{4.11}$$

We then define (Z_1, Z_2) as follows:

$$\mathcal{L}(Z_1, Z_2) = \tilde{\mathbf{P}}_1 + \Delta^*\left(\mu - \pi_1^* \tilde{\mathbf{P}}_1\right) = \mathbf{P}_1 + \mathbf{P}_2 \tag{4.12}$$

with $\Delta(u) = (u, u)$ and $\mathbf{P}_2 = S^* \mathbf{P}_1 + \Delta^*(\mu - \pi_1^* \tilde{\mathbf{P}}_1)$. It remains to prove that $\mathcal{L}(Z_1, Z_2)$ is well defined and satisfies all the properties required by the lemma.

First step: Prove that $\mathbf{P}_2$ is the sum of two non-negative measures.

Using (4.10), we can check that for all non-negative function f,

$$\pi_1^* \mathbf{P}_1(f) \le \frac{1}{2} \mu(f)$$

and

$$\pi_2^* \mathbf{P}_1(f) = \pi_1^*\left(S^* \mathbf{P}_1\right)(f) \le \frac{1}{2} \mu(f).$$

By adding the two previous inequalities, we deduce that the measure $\mu - \pi_1^* \tilde{\mathbf{P}}_1$ is non-negative. This concludes the first step.

Second step: Prove that $\pi_1^*(\mathbf{P}_1 + \mathbf{P}_2) = \pi_2^*(\mathbf{P}_1 + \mathbf{P}_2) = \mu$.
This fact is almost obvious. We just need to use the fact that

$$\pi_1 \circ \Delta = \pi_2 \circ \Delta = \mathrm{Id}$$

and the symmetry property of $\tilde{\mathbf{P}}_1$,

$$\text{i.e. } \pi_1^*\tilde{\mathbf{P}}_1 = \pi_2^*\tilde{\mathbf{P}}_1.$$

Third step: Prove (4.8) and (4.9).
Let us first remark that the support of $\mathbf{P}_1 + \mathbf{P}_2$ is included in

$$\left\{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d \mid v = \Lambda_{\mathbf{x}}(u)\right\} \cup \left\{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d \mid v = \Lambda_{\mathbf{x}}^{-1}(u)\right\} \cup \left\{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d \mid v = u\right\}.$$

Therefore, by (2.12) in $(\mathbf{H_2})$ and the fact that

$$\left(\forall u \in \mathbb{R}^d, \left|\Lambda_{\mathbf{x}}(u) - u\right| \le M_K\right) \quad \Longleftrightarrow \quad \left(\forall u \in \mathbb{R}^d, \left|\Lambda_{\mathbf{x}}^{-1}(u) - u\right| \le M_K\right)$$

since $\Lambda_{\mathbf{x}}$ is invertible on $\mathbb{R}^d$, we finally get (4.9).

Then, using again $(\mathbf{H_2})$ where $\tilde{K}$ is defined and the definition of the subprobability $\mathbf{P}_1$ we get

$$\mathbb{P}\left(\tilde{F}(x, Z_1, y) = \tilde{F}(x', Z_2, y')\right) \ge \underbrace{\mathbf{P}_1\left(B(0, \tilde{K}) \times \Lambda\left(B(0, \tilde{K})\right)\right)}_{=\mathbb{P}(Z_2 = \Lambda_{\mathbf{x}}(Z_1), |Z_1| \le \tilde{K})} \tag{4.13}$$

and

$$\mathbb{P}\left(Z_2 = \Lambda_{\mathbf{x}}(Z_1), |Z_1| \le \tilde{K}\right) = \frac{1}{2} \int_{B(0,\tilde{K})} \left(D_{\Lambda_{\mathbf{x}}}(u) \wedge 1\right) \mu(du).$$

It just remains to use (2.11) and (2.12) in $(\mathbf{H_2})$ to conclude. Indeed,

$$\mathbb{P}\left(Z_2 = \Lambda_{\mathbf{x}}(Z_1), |Z_1| \le \tilde{K}\right) = \frac{1}{2} \int_{B(0,\tilde{K})} \left(\exp\left(\frac{|u|^2}{2} - \frac{|\Lambda_{\mathbf{x}}(u)|^2}{2}\right) \left|\det(J_{\Lambda_{\mathbf{x}}}(u))\right|\right) \wedge 1 \mu(du)$$

$$\ge \frac{1}{2}\mu\left(B(0, \tilde{K})\right)\left[\left(\exp\left(-\frac{(M_{\tilde{K}} + \tilde{K})^2}{2}\right) C_{\tilde{K}}\right) \wedge 1\right] =: \delta_{\tilde{K}} > 0$$

which concludes the proof. $\qquad\square$

4.2.2. *Sticking step*

Now, if the positions X_{n+1}^1 and X_{n+1}^2 are stuck together, we want that they remain fastened together for all $k > n + 1$ which means that:

$$\forall k \ge n + 1, \qquad F\left(X_k^1, \Delta_{k+1}^1\right) = F\left(X_k^2, \Delta_{k+1}^2\right)$$

$$\Longleftrightarrow \quad \forall k \ge n + 1, \qquad F\left(X_k^1, \xi_{k+1}^1 + \sum_{l=1}^{+\infty} a_l \xi_{k+1-l}^1\right) = F\left(X_k^1, \xi_{k+1}^2 + \sum_{l=1}^{+\infty} a_l \xi_{k+1-l}^2\right) \tag{4.14}$$

since $X_k^1 = X_k^2$. Recall that for all $k \in \mathbb{Z}$, $g_k = \xi_{k+1}^1 - \xi_{k+1}^2$ is the drift between the underlying noises. Then, if we have

$$\forall k \geq n+1, \qquad \xi_{k+1}^1 + \sum_{l=1}^{+\infty} a_l \xi_{k+1-l}^1 = \xi_{k+1}^2 + \sum_{l=1}^{+\infty} a_l \xi_{k+1-l}^2$$

$$\Longleftrightarrow \quad \forall k \geq n+1, \qquad g_k = -\sum_{l=1}^{+\infty} a_l g_{k-l},$$

(4.15)

the identity (4.14) is automatically satisfied.

Remark 4.2. The successful g_k defined by relation (4.15) is $\mathcal{F}_k$-measurable. This explains why we chose to index it by k even if it represents the drift between ξ_{k+1}^1 and ξ_{k+1}^2.

Hence, we will try to get (4.15) on successive finite intervals to finally get a bound on the successful-coupling probability. The size choice of those intervals will be important according to the hypothesis ($\mathbf{H_{poly}}$) or ($\mathbf{H_{exp}}$) that we made. The two next results will be our tools to get (4.15) on Section 5.2. For the sake of simplicity, we set out these results on $\mathbb{R}$. On $\mathbb{R}^d$ we just have to apply them on every marginal. Lemma 4.2 is almost the statement of Lemma 5.13 of [12] or Lemma 3.2 of [9].

Lemma 4.2. *Let $\mu := \mathcal{N}(0,1)$. Let $a \in \mathbb{R}$, $b \geq |a|$ and $M_b := \max(4b, -2\log(b/8))$.*

(i) *For all $b \geq |a|$, there exist δ_b^1 and $\delta_b^2 \in (0,1)$, such that we can build a probability measure $\mathcal{N}_{a,b}^2$ on $\mathbb{R}^2$ with every marginal equal to μ and such that:*

$$\mathcal{N}_{a,b}^2(\{(x,y) \mid y = x+a\}) \geq \delta_b^1 \quad and \quad \mathcal{N}_{a,b}^2(\{(x,y) \mid |y-x| \leq M_b\}) = 1.$$

(ii) *Moreover, if $b \in (0,1)$, the previous statement holds with $\delta_b^1 = 1 - b$.*

The following corollary is an adapted version of Lemma 3.3 of [9] to our discrete context.

Corollary 4.1. *Let $T > 0$ be an integer, $b > 0$, $g = (g_0, g_1, \ldots, g_T) \in \mathbb{R}^{T+1}$ such that $\|g\| \leq b$ where $\|\cdot\|$ is the euclidian norm on $\mathbb{R}^{T+1}$ and set $M_b := \max(4b, -2\log(b/8))$.*

(i) *Then, there exists $\delta_b^1 \in (0,1)$, for which we can build a random variable $((\xi_{k+1}^1)_{k \in [\![0,T]\!]}, (\xi_{k+1}^2)_{k \in [\![0,T]\!]})$ with values in $(\mathbb{R}^{T+1})^2$, with marginal distribution $\mathcal{N}(0, I_{T+1})$ and satisfying:*

$$\mathbb{P}(\xi_{k+1}^1 = \xi_{k+1}^2 + g_k \ \forall k \in [\![0,T]\!]) \geq \delta_b^1$$

and

$$\mathbb{P}(\|\xi^1 - \xi^2\| \leq M_b) = 1.$$

(ii) *Moreover, if $b \in (0,1)$, the previous statement holds with $\delta_b^1 = 1 - b$.*

Proof. Let $(u_k)_{k \in [\![0,T]\!]}$ be an orthonormal basis of $\mathbb{R}^{T+1}$ with $u_0 = \frac{g}{\|g\|}$. We denote by (U_1, U_2) a random variable which has distribution $\mathcal{N}_{a,b}^2$ (with $a = \|g\|$) given in the lemma 4.2. Let $(\varepsilon_k)_{k \in [\![1,T]\!]}$ be an i.i.d. random variable sequence with $\varepsilon_1 \sim \mathcal{N}(0,1)$ and independent from (U_1, U_2). Then, for $i = 1, 2$ we define the isometry:

$$\mathcal{W}^i : \quad \mathbb{R}^{T+1} \to \mathcal{W}^i(\mathbb{R}^{T+1}) \subset L^2(\Omega, \mathcal{F}, \mathbb{P}),$$

$$u_0 \mapsto U_i, \tag{4.16}$$

$$u_k \mapsto \varepsilon_k \qquad \text{for } k \in [\![1, T]\!].$$

And we set for all $n \in [\![0, T]\!]$, $\xi_{n+1}^i := \mathcal{W}^i(e_n)$ where e_n is the vector of $\mathbb{R}^{T+1}$ for which every coordinate is 0 except the $(n + 1)^{\text{th}}$ which is 1. Since $(u_k)_{k \in [\![0,T]\!]}$ is an orthonormal basis of $\mathbb{R}^{T+1}$, we then have:

$$e_n = \sum_{k=0}^{T} \langle e_n, u_k \rangle u_k.$$

Hence,

$$\xi_{n+1}^i = \mathcal{W}^i \left(\sum_{k=0}^{T} \langle e_n, u_k \rangle u_k \right) = U_i \frac{g_n}{\|g\|} + \sum_{k=1}^{T} \langle e_n, u_k \rangle \varepsilon_k.$$

ξ_{n+1}^i is clearly centered and Gaussian as a linear combination of independent centered Gaussian random variables and using that $\mathcal{W}^i$ is an isometry, we get that $(\xi_{k+1}^i)_{k \in [\![0,T]\!]}$ has distribution $\mathcal{N}(0, I_{T+1})$ for $i = 1, 2$. Therefore, we built ξ^1 and ξ^2 as announced. Indeed, by Lemma 4.2

$$\mathbb{P}\big(\xi_{n+1}^1 = \xi_{n+1}^2 + g_n \ \forall n \in [\![0, T]\!]\big) = \mathbb{P}\big(U_1 = U_2 + \|g\|\big) \geq \delta_b^1$$

and

$$\mathbb{P}\big(\|\xi^1 - \xi^2\| \leq M_b\big) = \mathbb{P}\big(|U_1 - U_2| \leq M_b\big) = 1.$$

(ii) also follows immediately from Lemma 4.2. $\qquad \square$

5. Coupling under ($\mathbf{H}_{\text{poly}}$) or ($\mathbf{H}_{\text{exp}}$)

We can now move on the real coupling procedure to finally get a lower-bound for the successful-coupling probability. In a first subsection, we explain exactly what we called "controlled cost" and in a second subsection we spell out our bound.

5.1. Admissibility condition

The "controlled cost" is called "admissibility" in [12]. Here, we will talk about (K, α)-admissibility, as in [9], but in the following sense:

Definition 5.1. Let $K > 0$ and $\alpha > 0$ be two constants and τ a random variable with values in $\mathbb{N}$. We say that the system is (K, α)-admissible at time τ if $\tau(\omega) < +\infty$ and if $(X_\tau^1(\omega), X_\tau^2(\omega), (\xi_n^1(\omega), \xi_n^2(\omega))_{n \leq \tau})$ satisfies

$$\forall n \geq 0, \qquad \left| \sum_{k=n+1}^{+\infty} a_k g_{\tau+n-k}(\omega) \right| \leq v_n \qquad (5.1)$$

and

$$\left| X_\tau^i(\omega) \right| \leq K, \qquad \left| \sum_{k=1}^{+\infty} a_k \xi_{\tau+1-k}^i(\omega) \right| \leq K \qquad \text{for } i = 1, 2 \qquad (5.2)$$

with

$$v_n = (n+1)^{-\alpha} \qquad \text{under } (\mathbf{H_{poly}}) \quad \text{and} \quad v_n = e^{-\alpha n} \qquad \text{under } (\mathbf{H_{exp}}). \qquad (5.3)$$

Remark 5.1. On the one hand, condition (5.1) measures the distance between the past of the noises (before time τ). On the other hand, condition (5.2) has two parts: the first one ensures that at time τ both processes are not far from each other and the second part is a constraint on the memory part of the Gaussian noise $\Delta_{\tau+1}^i$.

The aim is to prove that under those two conditions, the coupling will be successful with a probability lower-bounded by a positive constant. To this end, we will need to ensure that at every time τ_j, the system will be (K, α)-admissible with positive probability. We set:

$$\Omega_{\alpha,\tau}^1 := \left\{ \omega, \tau(\omega) < +\infty, \left| \sum_{k=n+1}^{+\infty} a_k g_{\tau+n-k}(\omega) \right| \leq v_n \ \forall n \in \mathbb{N} \right\} \qquad (5.4)$$

and

$$\Omega_{K,\tau}^2 := \left\{ \omega, \tau(\omega) < +\infty, \left| X_\tau^i(\omega) \right| \leq K \text{ and } \left| \sum_{k=1}^{+\infty} a_k \xi_{\tau+1-k}^i(\omega) \right| \leq K \text{ for } i = 1, 2 \right\}. \qquad (5.5)$$

We define

$$\Omega_{K,\alpha,\tau} = \Omega_{\alpha,\tau}^1 \cap \Omega_{K,\tau}^2. \qquad (5.6)$$

If $\omega \in \Omega_{K,\alpha,\tau}$, we will try to couple at time $\tau + 1$. Otherwise, we say that Step 1 fails and one begins Step 3. Hence, Step 1 of trial j has two ways to fail: either ω belongs to $\Omega_{K,\alpha,\tau_{j-1}}^c$ and one moves directly to Step 3 or ω belongs to $\Omega_{K,\alpha,\tau_{j-1}}$, one tries to couple and it fails.

5.2. Lower-bound for the successful-coupling probability

The main purpose of this subsection is to get a positive lower-bound for the successful-coupling probability which will be independent of j (the number of the tentative), in other words we want to prove the following proposition.

Proposition 5.1. *Assume* (**H₁**) *and* (**H₂**). *Let* $K > 0$, $\alpha > \frac{1}{2} \vee (\frac{3}{2} - \beta)$ *if we are under* (**H_poly**) *and* $\alpha > 0$ *different from* ζ *if we are under* (**H_exp**). *In both cases, there exists* δ_0 *in* $(0, 1)$ *such that for all* $j \geq 1$,

$$\delta_0 \leq \mathbb{P}(\Delta \tau_j = +\infty \mid \Omega_{K,\alpha,\tau_{j-1}}), \tag{5.7}$$

where $\Delta \tau_j := \tau_j - \tau_{j-1}$ *and* τ_j *is defined in Section* 4.1 *as the end of trial* j.
   Moreover, we can choose $\delta_1 \in (0, 1)$ *such that*

$$\forall j \geq 1, \qquad \delta_1 \leq \mathbb{P}(\tau_j < \infty \mid \tau_{j-1} < \infty). \tag{5.8}$$

The second part of Proposition 5.1 may appear of weak interest but will be of first importance in Section 6.2.

5.2.1. *Step* 1 (*hitting step*)

Lemma 5.1. *Let* $K > 0$ *and* $\alpha > 0$. *Assume* (**H₁**) *and* (**H₂**). *Let* $\tilde{K} > 0$ *be the constant appearing in* (**H₂**), $\delta_1 \in (0, 1)$ *and* τ *be a stopping time with respect to* $(\mathcal{F}_n)_{n \in \mathbb{Z}}$ *such that* $\mathbb{P}(\Omega_{K,\alpha,\tau}) > 0$.
   We can build $(\xi_{\tau+1}^1, \xi_{\tau+1}^2)$ *with* $\xi_{\tau+1}^1 \sim \mathcal{N}(0, I_d)$ *and* $\xi_{\tau+1}^2 \sim \mathcal{N}(0, I_d)$ *such that*

(i) *There exist* $K_1 \in (0, \tilde{K}]$ *and* $\delta_{K_1} \in (0, 1)$ *such that*

$$\mathbb{P}(X_{\tau+1}^1 = X_{\tau+1}^2 \mid \Omega_{K,\alpha,\tau}) \geq \mathbb{P}(\xi_{\tau+1}^2 = \Lambda_{\mathbf{x}}(\xi_{\tau+1}^1), |\xi_{\tau+1}^1| \leq K_1 \mid \Omega_{K,\alpha,\tau})$$
$$\geq \delta_{K_1} > 0 \tag{5.9}$$

and

$$\mathbb{P}(\Omega_{K,\alpha,\tau}^c \cup (\{\xi_{\tau+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau+1}^1) \text{ or } |\xi_{\tau+1}^1| > K_1\} \cap \Omega_{K,\alpha,\tau})) \geq \delta_1, \tag{5.10}$$

where $\mathbf{x} := (X_\tau^1, X_\tau^2, \sum_{k=0}^{+\infty} a_k \xi_{\tau+1-k}^1, \sum_{k=0}^{+\infty} a_k \xi_{\tau+1-k}^2)$ *and* $\Lambda_{\mathbf{x}}$ *comes from* (**H₂**).
(ii) *There exists* $M_K > 0$ *such that*

$$|g_\tau| = |\xi_{\tau+1}^1 - \xi_{\tau+1}^2| \leq M_K \quad a.s.$$

Remark 5.2. *The constant* δ_1 *is chosen independently from* K *and* α.

Before proving this result, let us explain a bit why we add the lower-bound (5.10). As we already said, we will see further (in Section 6.2) that we need the (uniform) bound on the failure-coupling probability given in (5.8). Therefore, for every $j \geq 1$, we will consider that Step 1 of trial j fails if and only if $\omega \in \Omega_{K,\alpha,\tau_{j-1}}^c \cup (\{\xi_{\tau_{j-1}+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau_{j-1}+1}^1) \text{ or } |\xi_{\tau_{j-1}+1}^1| > K_1\} \cap \Omega_{K,\alpha,\tau_{j-1}})$ and in this case one immediately begins Step 3. Hence, for all $j \geq 1$, thanks to Lemma 5.1, we get the existence of K_1 such that:

$$\mathbb{P}(\tau_j < \infty \mid \tau_{j-1} < \infty)$$
$$\geq \mathbb{P}(\Omega_{K,\alpha,\tau_{j-1}}^c \cup (\{\xi_{\tau_{j-1}+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau_{j-1}+1}^1) \text{ or } |\xi_{\tau_{j-1}+1}^1| > K_1\} \cap \Omega_{K,\alpha,\tau_{j-1}})) \geq \delta_1$$

and then (5.8) derives from Lemma 5.1. This construction may seem artificial but it is necessary to prove Proposition 6.2. Moreover, this has no impact on the computation of the rate of convergence to equilibrium since it only affects Step 1. We can now move on the proof of Lemma 5.1.

Proof. (i) Set $\mathbf{x} := (X_\tau^1, X_\tau^2, \sum_{k=1}^{+\infty} a_k \xi_{\tau+1-k}^1, \sum_{k=1}^{+\infty} a_k \xi_{\tau+1-k}^2)$. Conditionally to $\Omega_{K,\alpha,\tau}$ we have $\mathbf{x} \in B(0, K)^4$ and we can build (Z_1, Z_2) as in Lemma 4.1. Let $\xi \sim \mathcal{N}(0, 1)$ be independent from (Z_1, Z_2) and set

$$\left(\xi_{\tau+1}^1, \xi_{\tau+1}^2\right) = (\mathbb{1}_{\Omega_{K,\alpha,\tau}} Z_1 + \mathbb{1}_{\Omega_{K,\alpha,\tau}^c} \xi, \mathbb{1}_{\Omega_{K,\alpha,\tau}} Z_2 + \mathbb{1}_{\Omega_{K,\alpha,\tau}^c} \xi). \tag{5.11}$$

Therefore, we deduce by Lemma 4.1 and its proof that for all $K_1 \in (0, \tilde{K}]$,

$$\mathbb{P}\left(X_{\tau+1}^1 = X_{\tau+1}^2 \mid \Omega_{K,\alpha,\tau}\right) \geq \underbrace{\mathbb{P}\left(Z_2 = \Lambda_{\mathbf{x}}(Z_1), |Z_1| \leq K_1 \mid \Omega_{K,\alpha,\tau}\right)}_{=\mathbb{P}(\xi_{\tau+1}^2 = \Lambda_{\mathbf{x}}(\xi_{\tau+1}^1), |\xi_{\tau+1}^1| \leq K_1 \mid \Omega_{K,\alpha,\tau})} \geq \delta_{K_1} > 0. \tag{5.12}$$

And the first part of (i) is proven. It remains to choose the good $K_1 \in (0, \tilde{K}]$ to get the second part. Set $p_K := \mathbb{P}(\Omega_{K,\alpha,\tau})$ and $\mu := \mathcal{N}(0, I_d)$, then

$$\mathbb{P}\left(\Omega_{K,\alpha,\tau}^c \cup \left(\left\{\xi_{\tau+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau+1}^1) \text{ or } |\xi_{\tau+1}^1| > K_1\right\} \cap \Omega_{K,\alpha,\tau}\right)\right)$$
$$= 1 - p_K + p_K \mathbb{P}\left(\left\{\xi_{\tau+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau+1}^1) \text{ or } |\xi_{\tau+1}^1| > K_1\right\} \mid \Omega_{K,\alpha,\tau}\right)$$
$$\geq 1 - p_K + p_K \mathbb{P}\left(|\xi_{\tau+1}^1| > K_1 \mid \Omega_{K,\alpha,\tau}\right)$$
$$\geq 1 - p_K + p_K \mu\left(B(0, K_1)^c\right) = 1 - p_K + p_K(1 - \mu(B(0, K_1))),$$

where the last inequality is due to Lemma 4.1 one more time. Finally, it remains to choose $K_1 \in (0, \tilde{K}]$ small enough in order to get $1 - p_K + p_K(1 - \mu(B(0, K_1))) \geq \delta_1$.

(ii) If $\omega \in \Omega_{K,\alpha,\tau}$, by the previous construction and Lemma 4.1, we have $|g_\tau(\omega)| = |Z_1(\omega) - Z_2(\omega)| \leq M_K$. And if $\omega \in \Omega_{K,\alpha,\tau}^c$ then $|g_\tau(\omega)| = |\xi(\omega) - \xi(\omega)| = 0$ which concludes the proof of (ii). □

To fix the ideas let us recall what we mean by "success of Step 1" and "failure of Step 1" of trial j ($j \geq 1$):

$$\{\text{success of Step 1}\} = \Omega_{K,\alpha,\tau_{j-1}} \cap \left\{\xi_{\tau_{j-1}+1}^2 = \Lambda_{\mathbf{x}}(\xi_{\tau_{j-1}+1}^1), |\xi_{\tau_{j-1}+1}^1| \leq K_1\right\}, \tag{5.13}$$

$$\{\text{failure of Step 1}\} = \Omega_{K,\alpha,\tau_{j-1}}^c$$
$$\cup \left(\left\{\xi_{\tau_{j-1}+1}^2 \neq \Lambda_{\mathbf{x}}(\xi_{\tau_{j-1}+1}^1) \text{ or } |\xi_{\tau_{j-1}+1}^1| > K_1\right\} \cap \Omega_{K,\alpha,\tau_{j-1}}\right), \tag{5.14}$$

where $\mathbf{x} := (X_{\tau_{j-1}}^1, X_{\tau_{j-1}}^2, \sum_{k=0}^{+\infty} a_k \xi_{\tau_{j-1}+1-k}^1, \sum_{k=0}^{+\infty} a_k \xi_{\tau_{j-1}+1-k}^2)$.

5.2.2. *Step 2 (sticking step)*

Step 2 of trial j consists in trying to keep the paths fastened together on successive intervals $I_{j,\ell}$. More precisely, during trial j, we set

$$I_{j,0} := \{\tau_{j-1} + 1\}, \qquad I_{j,1} := [\![\tau_{j-1} + 2, \tau_{j-1} + 2c_2 - 1]\!] \quad \text{and}$$
$$\forall \ell \geq 2, \qquad I_{j,\ell} := [\![\tau_{j-1} + c_2 s_\ell, \tau_{j-1} + c_2 s_{\ell+1} - 1]\!], \tag{5.15}$$

where $c_2 \geq 2$ will be chosen further and with

$$\forall \ell \geq 2, \qquad s_\ell = \begin{cases} 2^\ell & \text{under } (\mathbf{H_{poly}}), \\ \ell & \text{under } (\mathbf{H_{exp}}). \end{cases} \tag{5.16}$$

We denote

$$\ell_j^* := \sup\{\ell \geq 1 \mid \forall n \in I_{j,\ell-1}, g_{n-1} = g_{n-1}^{(s)}\}, \tag{5.17}$$

where $g_{n-1}^{(s)}$ is the successful-coupling drift defined by (4.15), that is, $g_{n-1}^{(s)} = -\sum_{l=1}^{+\infty} a_l g_{n-1-l}$. In other words, I_{j,ℓ_j^*} is the interval where the failure occurs. If $\{\ell \geq 1 \mid \forall n \in I_{j,\ell-1}, g_{n-1} = g_{n-1}^{(s)}\} = \varnothing$, we adopt the convention $\ell_j^* = 0$, it corresponds to the case where the failure occurs at Step 1. When $\ell_j^* = +\infty$, trial j is successful. For a given positive α and $K > 0$, we set

$$\mathcal{B}_{j,\ell} := \Omega_{K,\alpha,\tau_{j-1}} \cap \{\ell_j^* > \ell\} \qquad \forall j \geq 1, \ell \geq 0, \tag{5.18}$$

which means that failure of Step 2 may occur at most after ℓ trials. With this notations we get

$$\mathbb{P}(\Delta\tau_j = +\infty \mid \Omega_{K,\alpha,\tau_{j-1}}) = \mathbb{P}(\text{success of Step 1} \mid \Omega_{K,\alpha,\tau_{j-1}}) \prod_{\ell=1}^{+\infty} \mathbb{P}(\mathcal{B}_{j,\ell} \mid \mathcal{B}_{j,\ell-1}), \tag{5.19}$$

where the event {success of Step 1} is defined by (5.13).

Remark 5.3. There is an infinite product in this expression of the successful-coupling probability. Hence, the size choice of the intervals $I_{j,\ell}$ defined in (5.16) will play a significant role in the convergence of the product to a positive limit.

In the following lemma, similarly to the above definitions, we consider for a stopping time τ the intervals $(I_{\tau,\ell})_{\ell \geq 1}$, the integer ℓ_τ^* and the events $\mathcal{B}_{\tau,\ell}$, replacing τ_{j-1} by τ.

Lemma 5.2. *Let $K > 0$, assume $(\mathbf{H_1})$ and $(\mathbf{H_2})$. Let $\alpha > \frac{1}{2} \vee (\frac{3}{2} - \beta)$ under $(\mathbf{H_{poly}})$ or $\alpha > 0$ different from ζ under $(\mathbf{H_{exp}})$. Let τ be a stopping time with respect to $(\mathcal{F}_n)_{n \in \mathbb{Z}}$ (defined in Section 4.1) and assume that the system is (K, α)-admissible at time τ, then there exists $C_K > 0$ such that for $c_2 \geq 2$ large enough the successful drift $g^{(s)}$ satisfies*

$$\text{for } \ell = 1, \qquad \|g^{(s)}\|_{I_{\tau,1}} \leq C_K$$

and

$$\forall \ell \geq 2, \qquad \left\| g^{(s)} \right\|_{I_{\tau,\ell}} = \left(\sum_{k=\tau+c_2 s_\ell}^{\tau+c_2 s_{\ell+1}-1} \left| g_{k-1}^{(s)} \right|^2 \right)^{1/2} \leq 2^{-\tilde{\alpha}\ell}$$

where

$$\tilde{\alpha} := \begin{cases} \min\{\alpha, \beta, \alpha+\beta-1\} - 1/2 - \varepsilon & \text{for all } \varepsilon > 0 & \text{under } (\mathbf{H_{poly}}), \\ \min(\alpha, \zeta) & & \text{under } (\mathbf{H_{exp}}). \end{cases}$$

Therefore, for all $\ell \geq 1$, we can build thanks to Corollary 4.1 $((\xi_k^1)_{k \in I_{\tau,\ell}}, (\xi_k^2)_{k \in I_{\tau,\ell}})$ *during Step* 2 *in such a way that*

$$\mathbb{P}(\mathcal{B}_{\tau,1} \mid \mathcal{B}_{\tau,0}) \geq \delta_K^1 \quad \text{and} \quad \forall \ell \geq 2, \qquad \mathbb{P}(\mathcal{B}_{\tau,\ell} \mid \mathcal{B}_{\tau,\ell-1}) \geq 1 - 2^{-\tilde{\alpha}\ell}$$

where $\delta_K^1 \in (0,1)$.

Moreover, if $2 \leq \ell_\tau^* < +\infty$, there exists $C_\alpha > 0$ independent from K such that

$$\left(\sum_{k=\tau+c_2 s_{\ell_\tau^*}}^{\tau+c_2 s_{\ell_\tau^*+1}-1} |g_{k-1}|^2 \right)^{1/2} \leq C_\alpha \left(\ell_\tau^* + 1 \right)$$

and if $\ell_\tau^* = 1$, $\|g\|_{I_{\tau,1}} \leq C_K'$ for some constant $C_K' > 0$.

Remark 5.4.

▷ Under hypothesis $(\mathbf{H_{poly}})$ the condition $\alpha > \frac{1}{2} \vee (\frac{3}{2} - \beta)$ will ensure that $\min\{\alpha, \beta, \alpha+\beta - 1\} - 1/2 > 0$.

▷ In the polynomial case, for technical reasons $\tilde{\alpha}$ depends on $\varepsilon > 0$. This expression allows us to put together different cases and simplify the lemma. Indeed, if $(\alpha, \beta) \notin \{1\} \times (0,1] \cup (0,1] \times \{1\}$, we can take $\varepsilon = 0$.

To prove this lemma we will use in the *polynomial case* the following technical result which a more precise statement and a proof are given in the supplementary material [23], Section 6.

Lemma 5.3 (Technical lemma). *Let $\alpha > 0$ and $\beta > 0$ such that $\alpha + \beta > 1$. Then, there exists $C(\alpha, \beta) > 0$ such that for every $\varepsilon > 0$,*

$$\forall n \geq 0, \qquad \sum_{k=0}^{n} (k+1)^{-\beta} (n+1-k)^{-\alpha} \leq C(\alpha, \beta)(n+1)^{-\min\{\alpha, \beta, \alpha+\beta-1\}+\varepsilon}.$$

When $(\alpha, \beta) \notin \{1\} \times (0,1] \cup (0,1] \times \{1\}$, we can take $\varepsilon = 0$ in the previous inequality.

We can now move on the proof of Lemma 5.2.

Proof. Let us prove the first part of the lemma, namely the upper-bound of the ℓ^2 norm for the successful-coupling drift term on the intervals I_ℓ^τ. Indeed, the second part is just an application of Corollary 4.1. Since the system is (K, α)-admissible at time τ, we get by (5.1)

$$\forall n \geq 0, \qquad \left| \sum_{k=n+1}^{+\infty} a_k g_{\tau+n-k} \right| \leq v_n.$$

But, if Step 2 is successful, we recall that by (4.15) the successful drift satisfies $g_{\tau+n}^{(s)} = -\sum_{k=1}^{+\infty} a_k g_{\tau+n-k}$ for all $n \geq 1$, hence

$$\forall n \geq 1, \qquad \sum_{k=0}^{n} a_k g_{\tau+n-k}^{(s)} = \underbrace{- \sum_{k=n+1}^{+\infty} a_k g_{\tau+n-k}}_{=:u_n} \quad \text{and we set } u_0 := g_\tau^{(s)}.$$

Therefore thanks to Remark 2.2, this is equivalent to $g_{\tau+n}^{(s)} = \sum_{k=0}^{n} b_k u_{n-k}$ for all $n \geq 0$. By the admissibility assumption, we have $\forall k \in \{0, \ldots, n-1\}$, $|u_{n-k}| \leq v_{n-k}$ and by Lemma 5.1(ii), $|u_0| \leq M_K$. Hence, we get

$$\forall n \geq 0, \qquad \left| g_{\tau+n}^{(s)} \right| \leq M_K \sum_{k=0}^{n} |b_k| v_{n-k}. \tag{5.20}$$

• *Polynomial case*: Assume ($\mathbf{H_{poly}}$). Then for all $n \geq 0$, $v_n = (n+1)^{-\alpha}$ with $\alpha > \frac{1}{2} \vee (\frac{3}{2} - \beta)$. Here, (5.20) is equivalent to

$$\forall n \geq 0, \qquad \left| g_{\tau+n}^{(s)} \right| \leq M_K C_\beta \sum_{k=0}^{n} (k+1)^{-\beta} (n+1-k)^{-\alpha}$$

$$\leq M_K' (n+1)^{-\min\{\alpha, \beta, \alpha+\beta-1\}+\varepsilon},$$

for all $\varepsilon > 0$ by applying the technical lemma 5.3 and setting $M_K' := C(\alpha, \beta) M_K C_\beta$.

We then set $\tilde{\alpha} := \min\{\alpha, \beta, \alpha + \beta - 1\} - 1/2 - \varepsilon$.

Hence, for all $\ell \geq 2$,

$$\left\| g^{(s)} \right\|_{I_{\tau, \ell}} = \left(\sum_{k=\tau+c_2 2^\ell}^{\tau+c_2 2^{\ell+1}-1} \left| g_{k-1}^{(s)} \right|^2 \right)^{1/2}$$

$$\leq M_K' \left(\sum_{k=c_2 2^\ell}^{c_2 2^{\ell+1}-1} k^{-2\tilde{\alpha}-1} \right)^{1/2}$$

$$\leq M_K' \times \left(c_2 2^\ell \times \left(c_2 2^\ell \right)^{-2\tilde{\alpha}-1} \right)^{1/2} = M_K' c_2^{-\tilde{\alpha}} 2^{-\tilde{\alpha}\ell}.$$

It remains to choose $c_2 \geq 2 \vee (M_K')^{1/\tilde{\alpha}}$ to get the desired bound.

• *Exponential case*: Assume (**H$_{\text{exp}}$**). Then for all $n \geq 0$, $v_n = e^{-\alpha n}$ with $\alpha > 0$ and $\alpha \neq \zeta$. Here, (5.20) is equivalent to

$$\forall n \geq 0, \qquad \left| g^{(s)}_{\tau+n} \right| \leq M_K C_\zeta \sum_{k=0}^{n} e^{-\zeta k} e^{-\alpha(n-k)}$$

$$\leq M'_K e^{-\min\{\alpha, \zeta\} n},$$

where we have set $M'_K := M_K C_\zeta$. We then define $\tilde{\alpha} := \min(\alpha, \zeta)$.

Hence, for all $\ell \geq 2$,

$$\left\| g^{(s)} \right\|_{I^{(\tau)}_\ell} = \left(\sum_{k=\tau+c_2\ell}^{\tau+c_2(\ell+1)-1} \left| g^{(s)}_{k-1} \right|^2 \right)^{1/2}$$

$$\leq M'_K \left(\sum_{k=c_2\ell}^{c_2(\ell+1)-1} e^{-2\tilde{\alpha}(k-1)} \right)^{1/2}$$

$$\leq M'_K \times \left(\frac{e^{-2\tilde{\alpha}(c_2\ell-2)}}{2\tilde{\alpha}} \left[1 - e^{-2\tilde{\alpha}c_2} \right] \right)^{1/2} \qquad \text{(by integral upper-bound)}$$

$$\leq e^{-\tilde{\alpha}\ell} \qquad \text{(by choosing } c_2 \geq 2 \text{ large enough)}$$

$$\leq 2^{-\tilde{\alpha}\ell}.$$

For $\ell = 1$, in both polynomial and exponential cases, the same approach gives us the existence of $C_K > 0$ such that

$$\left\| g^{(s)} \right\|_{I_{\tau,1}} \leq C_K. \qquad \qquad \square$$

By combining Lemma 5.1, Lemma 5.2 and the expression (5.19) we finally get Proposition 5.1.

6. (K, α)-Admissibility

6.1. On condition (5.1)

Let $\Delta t^{(j)}_3$ be the duration of Step 3 of trial j for $j \geq 1$. The purpose of the next proposition is to prove that thanks to a calibration of $\Delta t^{(j)}_3$, one satisfies almost surely condition (5.1) at time τ_j.

Proposition 6.1. *Assume* (**H$_1$**) *and* (**H$_2$**). *Let* $\alpha \in (\frac{1}{2} \vee (\frac{3}{2} - \beta), \rho)$ *if we are under* (**H$_{\text{poly}}$**) *and* $\alpha \in (0, \lambda)$ *different from* ζ *if we are under* (**H$_{\text{exp}}$**). *Assume that for all* $j \geq 1$,

$$\Delta t^{(j)}_3 = \begin{cases} t_* \varsigma^j 2^{\theta \ell^*_j} & \text{with } \theta > \left(2(\rho - \alpha) \right)^{-1} & \text{under } (\mathbf{H_{\text{poly}}}), \\ t_* + \varsigma^j + \theta \ell^*_j & \text{with } \theta > 0 & \text{under } (\mathbf{H_{\text{exp}}}), \end{cases}$$

where ℓ_j^* is defined in (5.17) and with $\varsigma > 1$ arbitrary. Then for every $K > 0$, there exists a choice of t_* such that, for all $j \geq 0$, condition (5.1) is a.s. true at time τ_j on the event $\{\tau_j < +\infty\}$. In other words, for all $j \geq 0$,

$$\mathbb{P}\big(\Omega^1_{\alpha,\tau_j} \mid \{\tau_j < +\infty\}\big) = 1.$$

Remark 6.1.

$\triangleright$ With Proposition 6.1 in hand, we are now in position to discuss the statement of Theorem 2.1(iii) as mentioned in Remark 2.6. Adapting the proof of (7.3) (by taking an exponential Markov inequality), it appears that a fundamental tool to get exponential rate of convergence to equilibrium would be: for all $j \geq 1$, there exist $\lambda, C > 0$ such that

$$\mathbb{E}\big[e^{\lambda \Delta \tau_j} \mid \mathcal{F}_{\tau_{j-1}}\big] \leq C, \tag{6.1}$$

where $\Delta \tau_j := \tau_j - \tau_{j-1}$. But, under $(\mathbf{H_{exp}})$, Proposition 6.1 shows that $\Delta \tau_j \geq \varsigma^j$ with $\varsigma > 1$. This dependency on j conflicts with the necessary control of the conditional expectation previously cited.

$\triangleright$ Now, let us focus on the particular case of a finite memory: there exists $m \geq 1$ such that $a_k = 0$ for all $k > m$. As we saw in the second part of Section 2.6, $(|b_k|)$ decays exponentially fast. Then, an adjustment[1] of the proof of Proposition 6.1 would lead to (6.1) and perhaps in such case, our robust but general approach could be simplified.

Proof. Let us begin by the first coupling trial, in other words for $j = 0$. We recall that $g_k = 0$ for all $k < \tau_0$ (see (4.6)), therefore

$$\forall n \geq 0, \qquad \left| \sum_{k=n+1}^{+\infty} a_k g_{\tau_0 + n - k} \right| = 0 \leq v_n$$

and then condition (5.1) is a.s. true at time τ_0. Now, we assume $j \geq 1$ and we work on the event

$$\{\tau_j < +\infty\} \qquad \big(\supset \{\tau_m < +\infty\} \text{ for all } 0 \leq m \leq j-1 \big).$$

Let us prove that on this event we have for all $n \geq 0$, $|\sum_{k=n+1}^{+\infty} a_k g_{\tau_j + n - k}| \leq v_n$. Set $u_n := \sum_{k=n+1}^{+\infty} a_k g_{\tau_j + n - k}$. Since $g_k = 0$ for all $k < \tau_0$, we get $u_n = \sum_{k=n+1}^{n + \tau_j - \tau_0} a_k g_{\tau_j + n - k}$. Let us now separate the right term into the contributions of the different coupling trials. We get

$$u_n = \sum_{m=1}^{j} \left(\sum_{k=n+\tau_j - \tau_m + 1}^{n+\tau_j - \tau_{m-1}} a_k g_{\tau_j + n - k} \right) = \sum_{m=1}^{j} \Big(\underbrace{\sum_{k=\tau_{m-1}}^{\tau_m - 1} a_{n+\tau_j - k} g_k}_{(\star)_m} \Big)$$

---

[1] In this case, we can fix the duration of Step 3 equal to m since the memory only involves the m previous times.

and $(\star)_m$ corresponds to the contribution of trial m, divided into two parts: success and failure. We have now to distinguish two cases:

First case: $\ell_m^* \geq 1$, in other words the failure occurs during Step 2. We recall that in this case the system was automatically (K, α)-admissible at time τ_{m-1}, which will allow us to use Lemma 5.2 on τ_{m-1}.

Then, since $g_k = 0$ on $[\![\tau_{m-1} + c_2 s_{\ell_m^*+1}, \tau_m - 1]\!]$ by definition of Step 3 of the coupling procedure,

$$
(\star)_m = \sum_{k=\tau_{j_{m-1}}}^{\tau_{m-1}+c_2 s_{\ell_m^*+1}-1} a_{n+\tau_j-k} g_k
$$

$$
= \underbrace{\sum_{k=0}^{c_2 s_{\ell_m^*}-1} a_{n+\tau_j-\tau_{m-1}-k} g_{\tau_{m-1}+k}}_{\text{success}} + \underbrace{\sum_{k=c_2 s_{\ell_m^*}}^{c_2 s_{\ell_m^*+1}-1} a_{n+\tau_j-\tau_{m-1}-k} g_{\tau_{m-1}+k}}_{\text{failure}} \,.
$$

We have now to make the distinction between the polynomial and the exponential case.
▷ Under $(\mathbf{H_{poly}})$: $s_\ell = 2^\ell$, $|a_k| \leq C_\rho (k+1)^{-\rho}$ and then $a_k^2 \leq C_\rho^2 (k+1)^{-2\rho}$.
Using Cauchy–Schwarz inequality, the domination assumption on (a_k), and the fact that

$$
n + \tau_j - \tau_{m-1} - k + 1 \geq n + \tau_m - \tau_{m-1} - k + 1 = \Delta t_3^{(m)} + c_2 2^{\ell_m^*+1} + n - k
$$

we get,

$$
|\text{success}| \leq \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} a_{n+\tau_j-\tau_{m-1}-k}^2 \right)^{1/2} \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}
$$

$$
\leq C_\rho \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} (n+\tau_j-\tau_{m-1}-k+1)^{-2\rho} \right)^{1/2} \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}
$$

$$
\leq C_\rho \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} \left(\Delta t_3^{(m)} + c_2 2^{\ell_m^*+1} + n - k \right)^{-2\rho} \right)^{1/2} \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}
$$

$$
= C_\rho \left(\sum_{k=\Delta t_3^{(m)}+c_2 2^{\ell_m^*}+n+1}^{\Delta t_3^{(m)}+c_2 2^{\ell_m^*+1}+n} k^{-2\rho} \right)^{1/2} \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}
$$

$$
\leq C_\rho \left(c_2 2^{\ell_m^*} \left(n + \Delta t_3^{(m)} \right)^{-2\rho} \right)^{1/2} \left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2} \,.
$$

By the same arguments, we obtain

$$|\text{failure}| \leq C_\rho \big(c_2 2^{\ell_m^*}\big(n + \Delta t_3^{(m)}\big)^{-2\rho}\big)^{1/2} \left(\sum_{k=c_2 2^{\ell_m^*}}^{c_2 2^{\ell_m^*+1}-1} |g_{\tau_{m-1}+k}|^2\right)^{1/2}.$$

Hence, by the triangular inequality, we have

$$\big|(\star)_m\big| \leq C_\rho \sqrt{c_2}\, 2^{\ell_m^*/2}\big(n + \Delta t_3^{(m)}\big)^{-\rho} \left[\underbrace{\left(\sum_{k=0}^{c_2 2^{\ell_m^*}-1} |g_{\tau_{m-1}+k}|^2\right)^{1/2}}_{(1)} + \underbrace{\left(\sum_{k=c_2 2^{\ell_m^*}}^{c_2^{\ell_m^*+1}-1} |g_{\tau_{m-1}+k}|^2\right)^{1/2}}_{(2)}\right].$$

Since $\ell_j^* \geq 1$, by Lemma 5.1(ii) and Lemma 5.2, we have

$$(1) \leq M_K + \sum_{\ell=1}^{\ell_m^*-1} \big\|g^{(s)}\big\|_{I_\ell} \leq M_K + C_K + \sum_{\ell=2}^{+\infty} 2^{-\tilde{\alpha}\ell} =: \tilde{C}_K$$

and

$$(2) \leq \max\big(C_\alpha, C_K'\big)\big(\ell_m^* + 1\big).$$

Therefore

$$\big|(\star)_m\big| \leq C_K^{(2)} 2^{\ell_m^*/2}\big(\ell_m^* + 1\big)\big(n + \Delta t_3^{(m)}\big)^{-\rho} \tag{6.2}$$

where $C_K^{(2)} := \max(C_\alpha, C_K', \tilde{C}_K)$.

Moreover, recall that under $(\mathbf{H_{poly}})$

$$\Delta t_3^{(m)} = t_* \varsigma^m 2^{\theta \ell_m^*} \qquad \text{with } \theta > \big(2(\rho - \alpha)\big)^{-1} \text{ and } \varsigma > 1. \tag{6.3}$$

Plugging the definition of $\Delta t_3^{(m)}$ into (6.2) and using that for all $x, y > 0$, $(x + y)^{-\rho} \leq x^{-(\rho-\alpha)} y^{-\alpha}$

$$\big|(\star)_m\big| \leq C_K^{(2)} 2^{\ell_m^*/2}\big(\ell_m^* + 1\big)\big(n + \Delta t_3^{(m)}\big)^{-\rho}$$

$$\leq C_K^{(2)}\big(\ell_m^* + 1\big)2^{(1/2-\theta(\rho-\alpha))\ell_m^*}\big(t_* \varsigma^m\big)^{-(\rho-\alpha)}(n + 1)^{-\alpha}. \tag{6.4}$$

Since $\theta > (2(\rho - \alpha))^{-1}$ we have

$$C_{\alpha,K} = \sup_{\ell^* > 0} C_K^{(2)}\big(\ell^* + 1\big)2^{(1/2-\theta(\rho-\alpha))\ell^*} < +\infty,$$

and (6.4) yields

$$\big|(\star)_m\big| \leq C_{\alpha,K}\big(t_* \varsigma^m\big)^{-(\rho-\alpha)}(n + 1)^{-\alpha} \qquad \text{under } (\mathbf{H_{poly}}). \tag{6.5}$$

$\triangleright$ Under ($\mathbf{H_{exp}}$): $s_\ell = \ell$, $|a_k| \le C_\lambda e^{-\lambda k}$ and then $a_k^2 \le C_\lambda^2 e^{-2\lambda k}$.

Since the proof is almost the same in the exponential case, we will go faster and skip some details.

Using again Cauchy–Schwarz inequality, the domination assumption on (a_k), and the fact that

$$n + \tau_j - \tau_{m-1} - k \ge n + \tau_m - \tau_{m-1} - k = \Delta t_3^{(m)} + c_2(\ell_m^* + 1) + n - k - 1$$

we get

$$|\text{success}| \le \frac{C_\lambda e^\lambda}{\sqrt{2\lambda}} e^{-\lambda(n+\Delta t_3^{(m)})} \left(\sum_{k=0}^{c_2 2^{\ell_m^*} - 1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}$$

and by the same arguments,

$$|\text{failure}| \le \frac{C_\lambda e^\lambda}{\sqrt{2\lambda}} e^{-\lambda(n+\Delta t_3^{(m)})} \left(\sum_{k=c_2\ell_m^*}^{c_2(\ell_m^*+1)-1} |g_{\tau_{m-1}+k}|^2 \right)^{1/2}.$$

As in the polynomial case, by using Lemma 5.1 and 5.2 we get the existence of $C_K^{(3)} > 0$ such that

$$\left| (\star)_m \right| \le C_K^{(3)} (\ell_m^* + 1) e^{-\lambda(n+\Delta t_3^{(m)})}. \tag{6.6}$$

Moreover, recall that under ($\mathbf{H_{exp}}$)

$$\Delta t_3^{(m)} = t_* + \varsigma^m + \theta \ell_m^* \qquad \text{with } \theta > 0 \text{ and } \varsigma > 1. \tag{6.7}$$

Plugging the definition of $\Delta t_3^{(m)}$ into (6.6), we get

$$\left| (\star)_m \right| \le C_K^{(3)} (\ell_m^* + 1) e^{-\lambda(n+\Delta t_3^{(m)})}$$
$$\le C_K^{(3)} (\ell_m^* + 1) e^{-(\lambda-\alpha)\Delta t_3^{(m)}} e^{-\alpha n}$$
$$= C_K^{(3)} (\ell_m^* + 1) e^{-(\lambda-\alpha)(t_*+\varsigma^m+\theta\ell_m^*)} e^{-\alpha n}.$$

We set

$$C'_{\alpha,K} = \sup_{\ell^*>0} C_K^{(3)} (\ell^* + 1) e^{-\theta(\lambda-\alpha)\ell^*} < +\infty.$$

And this gives us

$$\left| (\star)_m \right| \le C'_{\alpha,K} e^{-(\lambda-\alpha)(t_*+\varsigma^m)} e^{-\alpha n} \qquad \text{under } (\mathbf{H_{exp}}). \tag{6.8}$$

Second case: $\ell_m^* = 0$, in other words, failure occurs during Step 1. This includes the case when the system is not (K, α)-admissible at time τ_{m-1}.

We have

$$(\star)_m = a_{n+\tau_j - \tau_{m-1}} g_{\tau_{m-1}}.$$

By Lemma 5.1(ii), $|g_{\tau_{m-1}}| \leq M_K$.

Moreover, since $n + \tau_j - \tau_{m-1} \geq n + \tau_m - \tau_{m-1} = n + \Delta t_3^{(m)}$, we obtain by using the same method as in the first case,

$$|(\star)_m| \leq \begin{cases} M_K \big(t_* \varsigma^m\big)^{-(\rho-\alpha)} (n+1)^{-\alpha} & \text{under } (\mathbf{H_{poly}}), \\ M_K e^{-(\lambda-\alpha)(t_*+\varsigma^m)} e^{-\alpha n} & \text{under } (\mathbf{H_{exp}}). \end{cases} \tag{6.9}$$

By putting (6.5), (6.8) and (6.9) together, we finally get

$$|(\star)_m| \leq \begin{cases} \max(M_K, C_{\alpha,K})\big(t_* \varsigma^m\big)^{-(\rho-\alpha)} (n+1)^{-\alpha} & \text{under } (\mathbf{H_{poly}}), \\ \max\big(M_K, C'_{\alpha,K}\big)e^{-(\lambda-\alpha)(t_*+\varsigma^m)} e^{-\alpha n} & \text{under } (\mathbf{H_{exp}}). \end{cases}$$

Set $S_1 = \sum_{m=1}^{+\infty} \varsigma^{-(\rho-\alpha)m}$ and $S_2 = \sum_{m=1}^{+\infty} e^{-(\lambda-\alpha)\varsigma^m}$. By choosing t_* large enough, we obtain for all $1 \leq m \leq j$:

$$|(\star)_m| \leq \begin{cases} \dfrac{1}{S_1} \varsigma^{-(\rho-\alpha)m} (n+1)^{-\alpha} & \text{under } (\mathbf{H_{poly}}), \\[2mm] \dfrac{1}{S_2} e^{-(\lambda-\alpha)\varsigma^m} e^{-\alpha n} & \text{under } (\mathbf{H_{exp}}). \end{cases} \tag{6.10}$$

Finally, by adding (6.10) for $m = 1, .., j$ we have

$$\forall n \geq 0, \qquad |u_n| \leq \begin{cases} (n+1)^{-\alpha} & \text{under } (\mathbf{H_{poly}}), \\ e^{-\alpha n} & \text{under } (\mathbf{H_{exp}}) \end{cases}$$

which concludes the proof. $\square$

6.2. Compact return condition (5.2)

In the sequel, we set

$$\mathcal{E}_j := \{\tau_j < \infty\} \qquad \big(= \{\tau_1 < \infty, \ldots, \tau_j < \infty\}\big). \tag{6.11}$$

The aim of this subsection is to prove the following proposition.

Proposition 6.2. *Assume* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$. *For all* $\varepsilon > 0$, *there exists* $K_\varepsilon > 0$ *such that*

$$\mathbb{P}\big(\Omega^2_{K_\varepsilon, \tau_j} \mid \mathcal{E}_j\big) \geq 1 - \varepsilon. \tag{6.12}$$

At this stage, we assume that ($\mathbf{H_{poly}}$) is true. Indeed, the exponential case will immediately follow from the polynomial one since ($\mathbf{H_{exp}}$) implies ($\mathbf{H_{poly}}$).

Since for every events A_1, A_2, A_3 and A_4, we have $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) \geq \sum_{i=1}^{4} \mathbb{P}(A_i) - 3$, it is enough to prove that for all $\varepsilon > 0$, there exists $K_\varepsilon > 0$ such that

$$\mathbb{P}\big(|X_{\tau_j}^i| \leq K_\varepsilon \mid \mathcal{E}_j\big) \geq 1 - \varepsilon \quad \text{and}$$

$$\mathbb{P}\left(\left|\sum_{k=1}^{+\infty} a_k \xi_{\tau_j+1-k}^i\right| \leq K_\varepsilon \,\Big|\, \mathcal{E}_j\right) \geq 1 - \varepsilon \qquad \text{for } i = 1, 2 \tag{6.13}$$

to get (6.12). Let us first focus on the first part of (6.13) concerning $|X_{\tau_j}^i|$ for $i = 1, 2$. Recall that the function $V : \mathbb{R}^d \to \mathbb{R}_+^*$ appearing in ($\mathbf{H_1}$) is such that $\lim_{|x| \to +\infty} V(x) = +\infty$. For $K > 0$ large enough we then have: $|x| \geq K \Rightarrow V(x) \geq K$. Therefore, for $i = 1, 2$ and K_ε large enough, using Markov inequality we get

$$\mathbb{P}\big(|X_{\tau_j}^i| \geq K_\varepsilon \mid \mathcal{E}_j\big) \leq \mathbb{P}\big(V\big(X_{\tau_j}^i\big) \geq K_\varepsilon \mid \mathcal{E}_j\big) \leq \frac{\mathbb{E}(V(X_{\tau_j}^i) \mid \mathcal{E}_j)}{K_\varepsilon}. \tag{6.14}$$

Hence, the first part of (6.13) is true if there exists a constant C such that for every $j \in \mathbb{N}$ and for every $K > 0$,

$$\mathbb{E}\big(V\big(X_{\tau_j}^i\big) \mid \mathcal{E}_j\big) \leq C \qquad \text{for } i = 1, 2. \tag{6.15}$$

Indeed, plugging (6.15) into (6.14) and taking $K_\varepsilon \geq \frac{C}{\varepsilon}$ yield the desired inequality. We see here that the independence of C with respect to K is essential.

For the sake of simplicity, we we will first use the following hypothesis to prove (6.15):

($\mathbf{H_1'}$) Let $\gamma \in (0, 1)$. There exists $C_\gamma > 0$ such that for all $j \in \mathbb{N}$, for every $K > 0$ and for $i = 1, 2$,

$$\mathbb{E}\left[\sum_{l=1}^{\Delta\tau_j} \gamma^{\Delta\tau_j - l} |\Delta_{\tau_{j-1}+l}^i| \,\Big|\, \mathcal{E}_j\right] < C_\gamma$$

where $\Delta\tau_j := \tau_j - \tau_{j-1}$ and Δ^i is the stationary Gaussian sequence defined by (2.2).

Proposition 6.3. *Assume* ($\mathbf{H_1}$), ($\mathbf{H_2}$) *and* ($\mathbf{H_1'}$). *Let* $(X_n^1, X_n^2)_{n \in \mathbb{N}}$ *be a solution of* (4.2) *with initial condition* (X_0^1, X_0^2) *satisfying* $\mathbb{E}(V(X_0^i)) < \infty$ *for* $i = 1, 2$. *Moreover, assume that* $\tau_0 = 0$ *and that* $(\tau_j)_{j \geq 1}$ *is built in such a way that for all* $j \geq 1$, $\mathbb{P}(\mathcal{E}_j \mid \mathcal{E}_{j-1}) \geq \delta_1 > 0$ *(where* δ_1 *is not depending on* j*) and* $\Delta\tau_j \geq \frac{\log(\delta_1/2)}{\log(\gamma)}$. *Then, there exists a constant* C *such that for all* $j \in \mathbb{N}$ *and for every* $K > 0$,

$$\mathbb{E}\big(V\big(X_{\tau_j}^i\big) \mid \mathcal{E}_j\big) \leq C \qquad \text{for } i = 1, 2.$$

Remark 6.2.

▷ Actually, hypothesis ($\mathbf{H'_1}$) is true under ($\mathbf{H_{poly}}$) and will be proven in the supplementary material [23], Section 7.

▷ The existence of $\delta_1 > 0$ independent from j is proven in Section 5.2.

▷ To get $\Delta \tau_j \geq \frac{\log(\delta_1/2)}{\log(\gamma)}$, it is sufficient to choose t_* large enough in the expression of $\Delta t_3^{(j)}$ (see Proposition 6.1).

▷ Since in Theorem 2.1 we made the assumption $\int_{\mathcal{X}} V(x) \Pi_{\mathcal{X}}^* \mu_0(dx) < +\infty$ and since an invariant distribution (extracted thanks to Theorem 3.1) also satisfies $\int_{\mathcal{X}} V(x) \Pi_{\mathcal{X}}^* \mu_\star(dx) < +\infty$, we get that $\mathbb{E}(V(X_0^i)) < \infty$ for $i = 1, 2$. Hence, we can set $\tau_0 = 0$.

Proof. By ($\mathbf{H_1}$), there exist $\gamma \in (0, 1)$ and $C > 0$ such that for all $n \geq 0$ and for $i = 1, 2$ we have

$$V(X_{n+1}^i) \leq \gamma V(X_n^i) + C(1 + |\Delta_{n+1}^i|).$$

By applying this inequality at time $n = \tau_j - 1$, and by induction, we immediately get:

$$V(X_{\tau_j}^i) \leq \gamma^{\Delta \tau_j} V(X_{\tau_{j-1}}^i) + C \sum_{l=1}^{\Delta \tau_j} \gamma^{\Delta \tau_j - l} (1 + |\Delta_{\tau_{j-1}+l}^i|). \tag{6.16}$$

By assumption $\Delta \tau_j \geq \frac{\log(\delta_1/2)}{\log(\gamma)}$ then $\gamma^{\Delta \tau_j} \leq \frac{\delta_1}{2}$. Moreover, since $\mathcal{E}_j \subset \mathcal{E}_{j-1}$ and $\mathbb{P}(\mathcal{E}_j \mid \mathcal{E}_{j-1}) \geq \delta_1$, we get

$$\mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_j] = \frac{1}{\mathbb{P}(\mathcal{E}_j)} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mathbb{1}_{\mathcal{E}_j}] \leq \underbrace{\frac{\mathbb{P}(\mathcal{E}_{j-1})}{\mathbb{P}(\mathcal{E}_j)}}_{=\mathbb{P}(\mathcal{E}_j | \mathcal{E}_{j-1})^{-1}} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}]$$

$$\leq \delta_1^{-1} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}].$$

Therefore,

$$\mathbb{E}[\gamma^{\Delta \tau_j} V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_j] \leq \frac{\delta_1}{2} \delta_1^{-1} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}] = \frac{1}{2} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}].$$

Hence, by taking (6.16), we have

$$\mathbb{E}[V(X_{\tau_j}^i) \mid \mathcal{E}_j] \leq \frac{1}{2} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}] + C \sum_{l=0}^{+\infty} \gamma^l + C \mathbb{E}\left[\sum_{l=1}^{\Delta \tau_j} \gamma^{\Delta \tau_j - l} |\Delta_{\tau_{j-1}+l}^i| \,\Big|\, \mathcal{E}_j \right]$$

$$= \frac{1}{2} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}] + \frac{C}{1 - \gamma} + C \mathbb{E}\left[\sum_{l=1}^{\Delta \tau_j} \gamma^{\Delta \tau_j - l} |\Delta_{\tau_{j-1}+l}^i| \,\Big|\, \mathcal{E}_j \right].$$

Hypothesis ($\mathbf{H'_1}$) allows us to say

$$\mathbb{E}[V(X_{\tau_j}^i) \mid \mathcal{E}_j] \leq \frac{1}{2} \mathbb{E}[V(X_{\tau_{j-1}}^i) \mid \mathcal{E}_{j-1}] + C\left(\frac{1}{1 - \gamma} + C_\gamma \right).$$

By induction, we get the existence of a constant $\tilde{C}_\gamma > 0$ such that

$$\mathbb{E}\big[V(X^i_{\tau_j}) \mid \mathcal{E}_j\big] \le \left(\frac{1}{2}\right)^j \mathbb{E}\big[V(X^i_{\tau_0}) \mid \mathcal{E}_0\big] + \tilde{C}_\gamma.$$

Since $\mathbb{P}(\mathcal{E}_0) = 1$, it comes

$$\mathbb{E}\big[V(X^i_{\tau_j}) \mid \mathcal{E}_j\big] \le \left(\frac{1}{2}\right)^j \mathbb{E}\big[V(X^i_{\tau_0})\big] + \tilde{C}_\gamma.$$

Since $\tau_0 = 0$ and we assumed that $\mathbb{E}[V(X^i_0)] < \infty$, the proof is over. $\qquad\square$

As a result of the proof of $(\mathbf{H'_1})$ (see the supplementary material [23], Section 7), we get the first part of (6.13). The second part can be also deduced from the proof of $(\mathbf{H'_1})$ thanks to [23], Remark 7.1.

7. Proof of Theorem 2.1

Now we have all the necessary elements to prove the second part of the main theorem 2.1 concerning the convergence in total variation to the unique invariant distribution (where the uniqueness will immediately follow from this convergence).

We recall that $\Delta\tau_j$ denotes the duration of coupling trial j and we set

$$j^{(s)} := \inf\{j > 0, \Delta\tau_j = +\infty\}. \tag{7.1}$$

$j^{(s)}$ corresponds to the trial where the coupling procedure is successful. The aim of this section is to bound $\mathbb{P}(\tau_\infty > n)$ where $\tau_\infty = \inf\{n \ge 0 \mid X^1_k = X^2_k, \forall k \ge n\}$, since $\|\mathcal{L}((X^1_k)_{k \ge n}) - \mathcal{S}\mu_\star\|_{\mathrm{TV}} \le \mathbb{P}(\tau_\infty > n)$. But, we have

$$\mathbb{P}(\tau_\infty > n) = \mathbb{P}\left(\sum_{k=1}^{+\infty} \Delta\tau_k \mathbb{1}_{\{j^{(s)} > k\}} > n\right),$$

where $j^{(s)}$ is defined in (7.1). It remains to bound the right term. Let $p \in (0, +\infty)$.

If $p \in (0, 1)$, then by the Markov inequality and the simple inequality $|a + b|^p \le |a|^p + |b|^p$, we get

$$\mathbb{P}\left(\sum_{k=1}^{+\infty} \Delta\tau_k \mathbb{1}_{\{j^{(s)} > k\}} > n\right)$$

$$\le \frac{1}{n^p} \sum_{k=1}^{+\infty} \mathbb{E}\big[|\Delta\tau_k|^p \mathbb{1}_{\{j^{(s)} > k\}}\big]$$

$$\le \frac{1}{n^p} \sum_{k=1}^{+\infty} \mathbb{E}\big[\mathbb{E}\big[|\Delta\tau_k|^p \mathbb{1}_{\{\Delta\tau_k < +\infty\}} \mid \{\tau_{k-1} < +\infty\}\big] \mathbb{1}_{\{\tau_{k-1} < +\infty\}}\big]. \tag{7.2}$$

Else, if $p \geq 1$, by Markov inequality and Minkowski inequality, we have

$$
\mathbb{P}\left(\sum_{k=1}^{+\infty} \Delta \tau_k \mathbb{1}_{\{j^{(s)} > k\}} > n\right)
$$

$$
\leq \frac{1}{n^p} \mathbb{E}\left[\left(\sum_{k=1}^{+\infty} \Delta \tau_k \mathbb{1}_{\{j^{(s)} > k\}}\right)^p\right]
$$

$$
\leq \frac{1}{n^p}\left(\sum_{k=1}^{+\infty} \mathbb{E}\left[|\Delta \tau_k|^p \mathbb{1}_{\{j^{(s)} > k\}}\right]^{1/p}\right)^p
$$

$$
\leq \frac{1}{n^p}\left(\sum_{k=1}^{+\infty} \mathbb{E}\left[\mathbb{E}\left[|\Delta \tau_k|^p \mathbb{1}_{\{\Delta \tau_k < +\infty\}} \mid \{\tau_{k-1} < +\infty\}\right] \mathbb{1}_{\{\tau_{k-1} < +\infty\}}\right]^{1/p}\right)^p. \tag{7.3}
$$

We define the event $\mathcal{A}_{k,\ell} := \mathcal{B}_{k,\ell}^c \cap \mathcal{B}_{k,\ell-1}$ which corresponds to the failure of Step 2 after ℓ attempts at trial k. Both in (7.2) and (7.3), we separate the term $\mathbb{E}[|\Delta \tau_k|^p \mathbb{1}_{\{\Delta \tau_k < +\infty\}} \mid \{\tau_{k-1} < +\infty\}]$ through the events $\mathcal{A}_{k,\ell}$ which gives

$$
\mathbb{E}\left[|\Delta \tau_k|^p \mathbb{1}_{\{\Delta \tau_k < +\infty\}} \mid \{\tau_{k-1} < +\infty\}\right]
$$

$$
= \sum_{\ell=1}^{+\infty} \mathbb{E}\left[\mathbb{1}_{\mathcal{A}_{k,\ell}} |\Delta \tau_k|^p \mathbb{1}_{\{\Delta \tau_k < +\infty\}} \mid \{\tau_{k-1} < +\infty\}\right]. \tag{7.4}
$$

Moreover, thanks to Lemma 5.2 and the definition of the events $\mathcal{A}_{k,\ell}$, we deduce that for $\ell \geq 2$,

$$
\mathbb{P}\left(\mathcal{A}_{k,\ell} \mid \{\tau_{k-1} < +\infty\}\right) \leq 2^{-\tilde{\alpha}\ell}, \tag{7.5}
$$

where

$$
\tilde{\alpha} := \begin{cases} \min\{\alpha, \beta, \alpha + \beta - 1\} - 1/2 - \varepsilon & \text{for all } \varepsilon > 0 \quad \text{under } (\mathbf{H_{poly}}), \\ \min(\alpha, \zeta) & \text{under } (\mathbf{H_{exp}}). \end{cases}
$$

We have now to distinguish the polynomial case from the exponential one.

$\triangleright$ Under $(\mathbf{H_{poly}})$:

We have a bound of type $\Delta \tau_k \leq C_1 \varsigma^k 2^{\theta \ell}$ (due to the value of $\Delta t_3^{(k)}$, see Proposition 6.1) on the event $\mathcal{A}_{k,\ell}$ where $\varsigma > 1$ is arbitrary. Indeed, on $\mathcal{A}_{k,\ell}$, we have

$$
\Delta \tau_k = \tau_k - \tau_{k-1} \leq c_2 2^{\ell+1} + \Delta t_3^{(k)}
$$

$$
= c_2 2^{\ell+1} + t_* \varsigma^k 2^{\theta \ell}
$$

$$
\leq C_1 \varsigma^k 2^{(\theta \vee 1)\ell} \qquad \text{(for } C_1 \text{ large enough)}.
$$

Hence, in (7.4) we get

$$\mathbb{E}\big[\big|\Delta\tau_k\big|^p \mathbb{1}_{\{\Delta\tau_k<+\infty\}} \mid \{\tau_{k-1}<+\infty\}\big] \le C_1^p \varsigma^{kp}\left(\sum_{\ell=1}^{+\infty} 2^{(\theta\vee 1)p\ell}\mathbb{P}\big(\mathcal{A}_{k,\ell} \mid \{\tau_{k-1}<+\infty\}\big)\right)$$

$$\le C_1^p \varsigma^{kp}\left(2^{(\theta\vee 1)p} + \sum_{\ell=2}^{+\infty} 2^{((\theta\vee 1)p-\tilde{\alpha})\ell}\right) \qquad \text{using (7.5)}$$

$$\le C\varsigma^{kp} \qquad \text{if } p\in\left(0,\frac{\tilde{\alpha}}{\theta\vee 1}\right).$$

Then for $p\in(0,\frac{\tilde{\alpha}}{\theta\vee 1})$,

$$\mathbb{E}\big[\mathbb{E}\big[\big|\Delta\tau_k\big|^p \mathbb{1}_{\{\Delta\tau_k<+\infty\}} \mid \{\tau_{k-1}<+\infty\}\big]\mathbb{1}_{\{\tau_{k-1}<+\infty\}}\big] \le C\varsigma^{kp}\mathbb{P}\big(j^{(s)}>k-1\big) \qquad (7.6)$$

and it remains to control $\mathbb{P}(j^{(s)}>k-1)$. We have

$$\mathbb{P}\big(j^{(s)}>k-1\big) = \prod_{j=1}^{k-1}\mathbb{P}(\mathcal{E}_j \mid \mathcal{E}_{j-1}) = \prod_{j=1}^{k-1}\big(1-\mathbb{P}\big(\mathcal{E}_j^c \mid \mathcal{E}_{j-1}\big)\big),$$

where $\mathcal{E}_j$ is defined in (6.11). By Proposition 5.1 and 6.2 applied for $\varepsilon=1/2$, we get for every $j\ge 2$,

$$\mathbb{P}\big(\mathcal{E}_j^c \mid \mathcal{E}_{j-1}\big) \ge \mathbb{P}\big(\Delta\tau_j=+\infty \mid \Omega_{K_{1/2},\alpha,\tau_{j-1}}\big)\mathbb{P}\big(\Omega_{K_{1/2},\alpha,\tau_{j-1}} \mid \mathcal{E}_{j-1}\big) \ge \frac{\delta_0}{2},$$

where $\delta_0>0$ depends on $K_{1/2}$. Therefore, $\mathbb{P}(j^{(s)}>k-1)\le(1-\frac{\delta_0}{2})^{k-1}$ and by (7.6)

$$\mathbb{E}\big[\mathbb{E}\big[\big|\Delta\tau_k\big|^p \mathbb{1}_{\{\Delta\tau_k<+\infty\}} \mid \{\tau_{k-1}<+\infty\}\big]\mathbb{1}_{\{\tau_{k-1}<+\infty\}}\big] \le C\varsigma^{kp}\left(1-\frac{\delta_0}{2}\right)^{k-1}. \qquad (7.7)$$

Finally, by choosing $1<\varsigma<(1-\frac{\delta_0}{2})^{-1/p}$, we get using (7.2) or (7.3) that for all $p\in(0,\frac{\tilde{\alpha}}{\theta\vee 1})$, there exists $C_p>0$ such that

$$\mathbb{P}(\tau_\infty>n) \le \mathbb{P}\left(\sum_{k=1}^{+\infty}\Delta\tau_k \mathbb{1}_{\{j^{(s)}>k\}}>n\right) \le C_p n^{-p}. \qquad (7.8)$$

It remains to optimize the upper-bound $\frac{\tilde{\alpha}}{\theta\vee 1}$ for p. Since $\tilde{\alpha}:=\min\{\alpha,\beta,\alpha+\beta-1\}-1/2-\varepsilon$ with $\varepsilon>0$ as small as necessary and since by Proposition 6.1: $\theta>(2(\rho-\alpha))^{-1}$ and $\alpha\in(\frac{1}{2}\vee(\frac{3}{2}-\beta),\rho)$, we finally get (7.8) for all $p\in(0,\upsilon(\beta,\rho))$ where

$$\upsilon(\beta,\rho) = \sup_{\alpha\in(\frac{1}{2}\vee(\frac{3}{2}-\beta),\rho)}\min\{1,2(\rho-\alpha)\}\big(\min\{\alpha,\beta,\alpha+\beta-1\}-1/2\big).$$

This concludes the proof of Theorem 2.1 in the polynomial case.

▷ Under ($\mathbf{H_{exp}}$):

The proof is almost the same. The only differences are that we use the following bound

$$\Delta \tau_k = \tau_k - \tau_{k-1} \leq c_2(\ell + 1) + \Delta t_3^{(k)}$$
$$= c_2(\ell + 1) + t_* + \varsigma^k + \theta \ell$$
$$\leq C_1 \varsigma^k \theta \ell \text{ (for } C_1 \text{ large enough)}$$

on the events $\mathcal{A}_{k,\ell}$ and the upperbound $\mathbb{P}(\mathcal{A}_{k,\ell} \mid \{\tau_{k-1} < +\infty\}) \leq 2^{-\tilde{\alpha}\ell}$ given in (7.5). And then we get for all $p > 0$ the existence of $C_p > 0$ such that

$$\mathbb{P}(\tau_\infty > n) \leq \mathbb{P}\left(\sum_{k=1}^{+\infty} \Delta \tau_k \mathbb{1}_{\{j^{(s)} > k\}} > n\right) \leq C_p n^{-p}. \tag{7.9}$$

by choosing $1 < \varsigma < (1 - \frac{\delta_0}{2})^{-1/p}$ and the proof of Theorem 2.1 is over.

Acknowledgements

Supplementary Material

Supplement to "Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory" (DOI: 10.3150/18-BEJ1089SUPP; .pdf). We provide proofs of a few selected results from this paper.

References

[1] Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042

[2] Brockwell, P.J. and Davis, R.A. (2013). *Time Series*: *Theory and Methods*. Springer Science & Business Media.

[3] Brouste, A., Cai, C. and Kleptsyna, M. (2014). Asymptotic properties of the MLE for the autoregressive process coefficients under stationary Gaussian noise. *Math. Methods Statist*. **23** 103–115. MR3224634

[4] Brouste, A. and Kleptsyna, M. (2012). Kalman type filter under stationary noises. *Systems Control Lett*. **61** 1229–1234. MR2998208

[5] Cohen, S. and Panloup, F. (2011). Approximation of stationary solutions of Gaussian driven stochastic differential equations. *Stochastic Process*. *Appl*. **121** 2776–2801. MR2844540

[6] Cohen, S., Panloup, F. and Tindel, S. (2014). Approximation of stationary solutions to SDEs driven by multiplicative fractional noise. *Stochastic Process*. *Appl*. **124** 1197–1225. MR3148010

[7] Deya, A., Panloup, F. and Tindel, S. (2016). Rate of convergence to equilibrium of fractional driven stochastic differential equations with rough multiplicative noise. arXiv preprint, arXiv:1605.00880.

[8] Down, D., Meyn, S.P. and Tweedie, R.L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann*. *Probab*. **23** 1671–1691. MR1379163

[9] Fontbona, J. and Panloup, F. (2017). Rate of convergence to equilibrium of fractional driven stochastic differential equations with some multiplicative noise. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **53** 503–538. MR3634264

[10] Ford, N.J., Savostyanov, D.V. and Zamarashkin, N.L. (2014). On the decay of the elements of inverse triangular Toeplitz matrices. *SIAM J*. *Matrix Anal*. *Appl*. **35** 1288–1302. MR3272624

[11] Guasoni, P. (2006). No arbitrage under transaction costs, with fractional Brownian motion and beyond. *Math*. *Finance* **16** 569–582. MR2239592

[12] Hairer, M. (2005). Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann*. *Probab*. **33** 703–758. MR2123208

[13] Hairer, M. (2008). Ergodic properties of a class of non-Markovian processes. *Lecture Notes*.

[14] Hairer, M. and Ohashi, A. (2007). Ergodic theory for SDEs with extrinsic memory. *Ann*. *Probab*. **35** 1950–1977. MR2349580

[15] Hairer, M. and Pillai, N.S. (2013). Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths. *Ann*. *Probab*. **41** 2544–2598. MR3112925

[16] Jeon, J.-H., Tejedor, V., Burov, S., Barkai, E., Selhuber-Unkel, C., Berg-Sørensen, K., Oddershede, L. and Metzler, R. (2011). In vivo anomalous diffusion and weak ergodicity breaking of lipid granules. *Phys*. *Rev*. *Lett*. **106** 048103.

[17] Kolmogorov, A.N. and Rozanov, Ju.A. (1960). On a strong mixing condition for stationary Gaussian processes. *Theory Probab*. *Appl*. **5** 204–208. MR0133175

[18] Kou, S.C. (2008). Stochastic modeling in nanoscale biophysics: Subdiffusion within proteins. *Ann*. *Appl*. *Stat*. **2** 501–535. MR2524344

[19] Odde, D.J., Tanaka, E.M., Hawkins, S.S. and Buettner, H.M. (1996). Stochastic dynamics of the nerve growth cone and its microtubules during neurite outgrowth. *Biotechnol*. *Bioeng*. **50** 452–461.

[20] Ostry, D.I. (2006). Synthesis of accurate fractional Gaussian noise by filtering. *IEEE Trans*. *Inform*. *Theory* **52** 1609–1623. MR2241208

[21] Panloup, F. and Richard, A. (2018). Sub-exponential convergence to equilibrium for gaussian driven stochastic differential equations with semi-contractive drift. arXiv preprint, arXiv:1804.01348.

[22] Rosenblatt, M. (1972). Central limit theorem for stationary processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, *Volume* 2: *Probability Theory* 551–561. Berkeley: Univ. California Press. MR0402869

[23] Varvenne, M. (2019). Supplement to "Rate of convergence to equilibrium for discrete-time stochastic dynamics with memory." DOI:10.3150/18-BEJ1089SUPP.