The Annals of Probability
2005, Vol. 33, No. 2, 703–758
DOI 10.1214/009117904000000892

ERGODICITY OF STOCHASTIC DIFFERENTIAL EQUATIONS DRIVEN BY FRACTIONAL BROWNIAN MOTION[1]

BY MARTIN HAIRER

University of Warwick

We study the ergodic properties of finite-dimensional systems of SDEs driven by nondegenerate additive fractional Brownian motion with arbitrary Hurst parameter $H \in (0,1)$. A general framework is constructed to make precise the notions of "invariant measure" and "stationary state" for such a system. We then prove under rather weak dissipativity conditions that such an SDE possesses a unique stationary solution and that the convergence rate of an arbitrary solution toward the stationary one is (at least) algebraic. A lower bound on the exponent is also given.

Contents

Received August 2003; revised March 2004.
[1]Supported by the Fonds National Suisse.
AMS 2000 subject classifications. 60H10, 26A33.
Key words and phrases. Ergodicity, fractional Brownian motion, memory.

1. Introduction and main result. In this paper, we investigate the long-time behavior of stochastic differential equations driven by fractional Brownian motion. Fractional Brownian motion (or FBM for short) is a centered Gaussian process satisfying $B_H(0) = 0$ and

$$\mathbf{E}|B_H(t) - B_H(s)|^2 = |t - s|^{2H}, \qquad t, s > 0, \tag{1.1}$$

where H, the Hurst parameter, is a real number in the range $H \in (0, 1)$. When $H = \frac{1}{2}$, one recovers of course the usual Brownian motion, so this is a natural one-parameter family of generalizations of the "standard" Brownian motion. It follows from (1.1) that FBM is also self-similar, but with the scaling law

$$t \mapsto B_H(at) \approx t \mapsto a^H B_H(t),$$

where "$\approx$" denotes equivalence in law. Also, the sample paths of B_H are α-Hölder continuous for every $\alpha < H$. The main difference between FBM and the usual Brownian motion is that it is neither Markovian nor a semimartingale, so most standard tools from stochastic calculus cannot be applied to its analysis.

Our main motivation is to tackle the problem of ergodicity in non-Markovian systems. Such systems arise naturally in several situations. In physics, stochastic forces are used to describe the interaction between a (small) system and its (large) environment. There is no a priori reason to assume that the forces applied by the environment to the system are independent over disjoint time intervals. In statistical mechanics, for example, a non-Markovian noise term appears when one attempts to derive the Langevin equation from first principles [12, 23]. Self-similar stochastic processes like FBM appear naturally in hydrodynamics [17]. It appears that FBM is also useful to model long-time correlations in stock markets [7, 22].

Little seems to be known about the long-time behavior of non-Markovian systems. In the case of the non-Markovian Langevin equation (which is *not* covered by the results in this paper due to the presence of a delay term), the stationary solution is explicitly known to be distributed according to the usual equilibrium Gibbs measure. The relaxation toward equilibrium is a very hard problem that was solved in [12, 13]. It is, however, still open in the nonequilibrium case, where the invariant state cannot be guessed a priori. One well-studied general framework for the study of systems driven by noise with extrinsic memory like the ones considered in this paper is given by the theory of random dynamical systems (see the monograph [1] and the reference list therein). In that framework, the existence of random attractors, and therefore the *existence* of invariant measures, seems to be well understood. On the other hand, the problem of *uniqueness* (in an appropriate sense; see the comment following Theorem 1.3 below) of the invariant measure on the random attractor seems to be much harder, unless one can show that the system possesses a unique stochastic fixed point. The latter situation was studied in [19] for infinite-dimensional evolution equations driven by FBM.

The reasons for choosing FBM as driving process for (SDE) below are twofold. First, in particular when $H > \frac{1}{2}$, FBM presents genuine long-time correlations that persist even under rescaling. The second reason is that there exist simple, explicit formulae that relate FBM to "standard" Brownian motion, which simplifies our analysis. We will limit ourselves to the case where the memory of the system comes entirely from the driving noise process, so we do not consider stochastic delay equations.

We will only consider equations driven by nondegenerate additive noise, that is, we consider equations of the form

$$dx_t = f(x_t)\,dt + \sigma\,dB_H(t), \qquad x_0 \in \mathbf{R}^n, \tag{SDE}$$

where $x_t \in \mathbf{R}^n$, $f:\mathbf{R}^n \to \mathbf{R}^n$, B_H is an n-dimensional FBM with Hurst parameter H and σ is a constant and invertible $n \times n$ matrix. Of course, (SDE) should be interpreted as an integral equation.

In order to ensure the existence of globally bounded solutions and in order to have some control on the speed at which trajectories separate, we make throughout the paper the following assumptions on the components of (SDE):

(A1) *Stability.* There exist constants $C_i^{(\mathrm{A1})} > 0$ such that

$$\langle f(x) - f(y), x - y\rangle \le \min\{C_1^{(\mathrm{A1})} - C_2^{(\mathrm{A1})}\|x-y\|^2, C_3^{(\mathrm{A1})}\|x-y\|^2\},$$

for every $x, y \in \mathbf{R}^n$.

(A2) *Growth and regularity.* There exist constants $C, N > 0$ such that f and its derivative satisfy

$$\|f(x)\| \le C(1 + \|x\|)^N, \qquad \|Df(x)\| \le C(1+\|x\|)^N,$$

for every $x \in \mathbf{R}^n$.

(A3) *Nondegeneracy*. The $n \times n$ matrix σ is invertible.

REMARK 1.1. We can assume that $\|\sigma\| \le 1$ without any loss of generality. This assumption will be made throughout the paper in order to simplify some expressions.

One typical example that we have in mind is given by $f(x) = x - x^3$, $x \in \mathbf{R}$, or any polynomial of odd degree with negative leading coefficient. Notice that f satisfies (A1) and (A2), but that it is not globally Lipschitz continuous.

When the Hurst parameter H of the FBM driving (SDE) is bigger than $\frac{1}{2}$, more regularity for f is required, and we will then sometimes assume that the following stronger condition holds instead of (A2):

(A2$'$) *Strong regularity.* The derivative of f is globally bounded.

Our main result is that (SDE) possesses a *unique* stationary solution. Furthermore, we obtain an explicit bound showing that every (adapted) solution to (SDE) converges toward this stationary solution, and that this convergence is at least algebraic. We make no claim concerning the optimality of this bound for the class of systems under consideration. Our results are slightly different for small and for large values of H, so we state them separately.

THEOREM 1.2 (Small Hurst parameter). *Let* $H \in (0, \frac{1}{2})$ *and let* f *and* σ *satisfy* (A1)–(A3). *Then, for every initial condition, the solution to* (SDE) *converges toward a unique stationary solution in the total variation norm. Furthermore, for every* $\gamma < \max_{\alpha < H} \alpha(1 - 2\alpha)$, *the difference between the solution and the stationary solution is bounded by* $C_\gamma t^{-\gamma}$ *for large* t.

THEOREM 1.3 (Large Hurst parameter). *Let* $H \in (\frac{1}{2}, 1)$ *and let* f *and* σ *satisfy* (A1)–(A3) *and* (A2$'$). *Then, for every initial condition, the solution to* (SDE) *converges toward a unique stationary solution in the total variation norm. Furthermore, for every* $\gamma < \frac{1}{8}$, *the difference between the solution and the stationary solution is bounded by* $C_\gamma t^{-\gamma}$ *for large* t.

REMARK 1.4. The "uniqueness" part of these statements should be understood as uniqueness in law in the class of stationary solutions adapted to the natural filtration induced by the two-sided FBM that drives the equation. There could in theory be other stationary solutions, but they would require knowledge of the future to determine the present, so they are usually discarded as unphysical.

Even in the context of Markov processes, similar situations do occur. One can well have uniqueness of the invariant measure, but nonuniqueness of the stationary state, although other stationary states would have to foresee the future. In this sense, the notion of uniqueness appearing in the above statements is similar to the notion of uniqueness of the invariant measure for Markov processes. (See, e.g., [1, 4, 5] for discussions on invariant measures that are not necessarily measurable with respect to the past.)

REMARK 1.5. The case $H = \frac{1}{2}$ is not covered by these two theorems, but it is well known that the convergence toward the stationary state is exponential in this case (see, e.g., [21]). In both cases, the word "total variation" refers to the total variation distance between measures on the space of paths; see also Theorem 6.1 below for a rigorous formulation of the results above.

1.1. *Idea of proof and structure of the paper.* Our first task is to make precise the notions of "initial condition," "invariant measure," "uniqueness" and "convergence" appearing in the formulation of Theorems 1.2 and 1.3. This will be achieved in Section 2, where we construct a general framework for the study of systems driven by non-Markovian noise. Section 3 shows how (SDE) fits into that framework.

The main tool used in the proof of Theorems 1.2 and 1.3 is a coupling construction similar in spirit to the ones presented in [11, 20]. More precisely, we first show by some compactness argument that there exists at least one invariant measure μ_* for (SDE). Then, given an initial condition distributed according to some arbitrary measure μ, we construct a "coupling process" (x_t, y_t) on $\mathbf{R}^n \times \mathbf{R}^n$ with the following properties:

1. The process x_t is a solution to (SDE) with initial condition μ_*.
2. The process y_t is a solution to (SDE) with initial condition μ.
3. The random time $\tau_\infty = \min\{t \mid x_s = y_s \ \forall s \geq t\}$ is almost surely finite.

The challenge is to introduce correlations between x_s and y_s in precisely such a way that τ_∞ is finite. If this is possible, the uniqueness of the invariant measure follows immediately. Bounds on the moments of τ_∞ furthermore translate into bounds on the rate of convergence toward this invariant measure. In Section 4, we expose the general mechanism by which we construct this coupling. Section 5 is then devoted to the precise formulation of the coupling process and to the study of its properties, which will be used in Section 6 to prove Theorems 1.2 and 1.3. We conclude this paper with a few remarks on possible extensions of our results to situations that are not covered here.

2. General theory of stochastic dynamical systems. In this section, we first construct an abstract framework that can be used to model a large class of physically relevant models where the driving noise is stationary. Our framework is very closely related to the framework of random dynamical systems with, however, one fundamental difference. In the theory of random dynamical systems (RDS), the abstract space Ω used to model the noise part typically encodes the *future* of the noise process. In our framework of "stochastic dynamical systems" (SDS) the noise space $\mathcal{W}$ typically encodes the *past* of the noise process. As a consequence, the evolution on $\mathcal{W}$ will be stochastic, as opposed to the deterministic evolution on Ω one encounters in the theory of RDS. This distinction may seem futile at first sight, and one could argue that the difference between RDS and SDS is nonexistent by adding the past of the noise process to Ω and its future to $\mathcal{W}$.

The additional structure we require is that the evolution on $\mathcal{W}$ possesses a *unique* invariant measure. Although this requirement may sound very strong, it is actually not, and most natural examples satisfy it, as long as $\mathcal{W}$ is chosen in such a way that it does not contain information about the future of the noise. In very loose terms, this requirement of having a unique invariant measure states that the noise process driving our system is stationary and that the Markov process modeling its evolution captures all its essential features in such a way that it could not be used to describe a noise process different from the one at hand. In particular, this means that there is a continuous inflow of "new randomness" into the system, which is a crucial feature when trying to apply probabilistic methods to the study of ergodic properties of the system. This is in opposition to the RDS formalism, where the noise is "frozen," as soon as an element of Ω is chosen.

From the mathematical point of view, we will consider that the physical process we are interested in lives on a "state space" $\mathcal{X}$ and that its driving noise belongs to a "noise space" $\mathcal{W}$. In both cases, we only consider Polish (i.e., complete, separable, and metrizable) spaces. One should think of the state space as a relatively small space which contains all the information accessible to a physical observer of the process. The noise space should be thought of as a much bigger abstract space containing all the information needed to construct a mathematical model of the driving noise up to a certain time. The information contained in the noise space is not accessible to the physical observer.

Before we state our definition of an SDS, we will recall several notation and definitions, mainly for the sake of mathematical rigor. The reader can safely skip the next section and come back to it for reference concerning the notation and the mathematically precise definitions of the concepts that are used.

2.1. *Preliminary definitions and notation.* First of all, recall he definition of a transition semigroup:

DEFINITION 2.1. Let $(\mathcal{E}, \mathscr{E})$ be a Polish space endowed with its Borel σ-field. A *transition semigroup* $\mathcal{P}_t$ on $\mathcal{E}$ is a family of maps $\mathcal{P}_t : \mathcal{E} \times \mathscr{E} \to [0, 1]$ indexed by $t \in [0, \infty)$ such that

(i) for every $x \in \mathcal{E}$, the map $A \mapsto \mathcal{P}_t(x, A)$ is a probability measure on $\mathcal{E}$ and, for every $A \in \mathscr{E}$, the map $x \mapsto \mathcal{P}_t(x, A)$ is $\mathscr{E}$-measurable,

(ii) one has the identity

$$\mathcal{P}_{s+t}(x, A) = \int_{\mathcal{E}} \mathcal{P}_s(y, A)\mathcal{P}_t(x, dy),$$

for every $s, t > 0$, every $x \in \mathcal{E}$ and every $A \in \mathscr{E}$,

(iii) $\mathcal{P}_0(x, \cdot) = \delta_x$ for every $x \in \mathcal{E}$.

We will freely use the notation $(\mathcal{P}_t\psi)(x) = \int_{\mathcal{E}} \psi(y)\mathcal{P}_t(x, dy)$, $(\mathcal{P}_t\mu)(A) = \int_{\mathcal{E}} \mathcal{P}_t(x, A)\mu(dx)$, where ψ is a measurable function on $\mathcal{E}$ and μ is a measure on $\mathcal{E}$.

Since we will always work with topological spaces, we will require our transition semigroups to have good topological properties. Recall that a sequence $\{\mu_n\}$ of measures on a topological space $\mathcal{E}$ is said to converge toward a limiting measure μ in the weak topology if

$$\int_{\mathcal{E}} \psi(x)\mu_n(dx) \to \int_{\mathcal{E}} \psi(x)\mu(dx) \qquad \forall \psi \in \mathcal{C}_b(\mathcal{E}),$$

where $\mathcal{C}_b(\mathcal{E})$ denotes the space of bounded continuous functions from $\mathcal{E}$ into $\mathbf{R}$. In the sequel, we will use the notation $\mathscr{M}_1(\mathcal{E})$ to denote the space of probability measures on a Polish space $\mathcal{E}$, endowed with the topology of weak convergence.

DEFINITION 2.2. A *transition semigroup* $\mathcal{P}_t$ on a Polish space $\mathcal{E}$ is *Feller* if it maps $\mathcal{C}_b(\mathcal{E})$ into $\mathcal{C}_b(\mathcal{E})$.

REMARK 2.3. This definition is equivalent to the requirement that $x \mapsto \mathcal{P}_t(x, \cdot)$ is continuous from $\mathcal{E}$ to $\mathscr{M}_1(\mathcal{E})$. As a consequence, Feller semigroups preserve the weak topology in the sense that if $\mu_n \to \mu$ in $\mathscr{M}_1(\mathcal{E})$, then $\mathcal{P}_t\mu_n \to \mathcal{P}_t\mu$ in $\mathscr{M}_1(\mathcal{E})$ for every given t.

Now that we have defined the "good" objects for the "noisy" part of our construction, we turn to the trajectories on the state space. We are looking for a space which has good topological properties but which is large enough to contain most interesting examples. One such space is the space of càdlàg paths (continu à droite, limite à gauche—continuous on the right, limits on the left), which can be turned into a Polish space when equipped with a suitable topology.

DEFINITION 2.4. Given a Polish space $\mathcal{E}$ and a positive number T, the space $\mathcal{D}([0,T],\mathcal{E})$ is the set of functions $f:[0,T]\to\mathcal{E}$ that are right-continuous and whose left-limits exist at every point. A sequence $\{f_n\}_{n\in\mathbf{N}}$ converges to a limit f if and only if there exists a sequence $\{\lambda_n\}$ of continuous and increasing functions $\lambda_n:[0,T]\to[0,T]$ satisfying $\lambda_n(0)=0$, $\lambda_n(T)=T$, and such that

$$\lim_{n\to\infty}\sup_{0\le s<t\le T}\left|\log\frac{\lambda_n(t)-\lambda_n(s)}{t-s}\right|=0 \tag{2.1}$$

and

$$\lim_{n\to\infty}\sup_{0\le t\le T} d\bigl(f_n(t), f(\lambda_n(t))\bigr)=0, \tag{2.2}$$

where d is any totally bounded metric on $\mathcal{E}$ which generates its topology.

The space $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$ is the space of all functions from $\mathbf{R}_+$ to $\mathcal{E}$ such that their restrictions to $[0,T]$ are in $\mathcal{D}([0,T],\mathcal{E})$ for all $T>0$. A sequence converges in $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$ if there exists a sequence $\{\lambda_n\}$ of continuous and increasing functions $\lambda_n:\mathbf{R}_+\to\mathbf{R}_+$ satisfying $\lambda_n(0)=0$ and such that (2.1) and (2.2) hold.

It can be shown (see, e.g., [9] for a proof) that the spaces $\mathcal{D}([0,T],\mathcal{E})$ and $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$ are Polish when equipped with the above topology (usually called the Skorohod topology). Notice that the space $\mathcal{D}([0,T],\mathcal{E})$ has a natural embedding into $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$ by setting $f(t)=f(T)$ for $t>T$ and that this embedding is continuous. However, the restriction operator from $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$ to $\mathcal{D}([0,T],\mathcal{E})$ is not continuous, since the topology on $\mathcal{D}([0,T],\mathcal{E})$ imposes that $f_n(T)\to f(T)$, which is not imposed by the topology on $\mathcal{D}(\mathbf{R}_+,\mathcal{E})$.

In many interesting situations, it is enough to work with continuous sample paths, which live in much simpler spaces:

DEFINITION 2.5. Given a Polish space $\mathcal{E}$ and a positive number T, the space $\mathcal{C}([0,T],\mathcal{E})$ is the set of continuous functions $f:[0,T]\to\mathcal{E}$ equipped with the supremum norm.

The space $\mathcal{C}(\mathbf{R}_+,\mathcal{E})$ is the space of all functions from $\mathbf{R}_+$ to $\mathcal{E}$ such that their restrictions to $[0,T]$ are in $\mathcal{C}([0,T],\mathcal{E})$ for all $T>0$. A sequence converges in $\mathcal{C}(\mathbf{R}_+,\mathcal{E})$ if all its restrictions converge.

It is a standard result that the spaces $\mathcal{C}([0,T],\mathcal{E})$ and $\mathcal{C}(\mathbf{R}_+,\mathcal{E})$ are Polish if $\mathcal{E}$ is Polish. We can now turn to the definition of the systems we are interested in.

2.2. *Definition of an SDS.* Let us recall the following standard notation. Given a product space $\mathcal{X} \times \mathcal{W}$, we denote by $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{W}}$ the maps that select the first (resp. second) component of an element. Also, given two measurable spaces $\mathcal{E}$ and $\mathcal{F}$, a measurable map $f : \mathcal{E} \to \mathcal{F}$ and a measure μ on $\mathcal{E}$, we define the measure $f^*\mu$ on $\mathcal{F}$ in the natural way by $f^*\mu = \mu \circ f^{-1}$.

We first define the class of noise processes we will be interested in:

DEFINITION 2.6. A quadruple $(\mathcal{W}, \{\mathcal{P}_t\}_{t\geq 0}, \mathbf{P}_w, \{\theta_t\}_{t\geq 0})$ is called a *stationary noise process* if it satisfies the following:

(i) $\mathcal{W}$ is a Polish space,

(ii) $\mathcal{P}_t$ is a Feller transition semigroup on $\mathcal{W}$, which accepts $\mathbf{P}_w$ as its unique invariant measure,

(iii) The family $\{\theta_t\}_{t>0}$ is a semiflow of measurable maps on $\mathcal{W}$ satisfying the property $\theta_t^* \mathcal{P}_t(x, \cdot) = \delta_x$ for every $x \in \mathcal{W}$.

This leads to the following definition of SDS, which is intentionally kept as close as possible to the definition of RDS in [1], Definition 1.1.1:

DEFINITION 2.7. A *stochastic dynamical system* on the Polish space $\mathcal{X}$ over the stationary noise process $(\mathcal{W}, \{\mathcal{P}_t\}_{t\geq 0}, \mathbf{P}_w, \{\theta_t\}_{t\geq 0})$ is a mapping

$$\varphi : \mathbf{R}_+ \times \mathcal{X} \times \mathcal{W} \to \mathcal{X}, \qquad (t, x, w) \mapsto \varphi_t(x, w),$$

with the following properties:

(SDS1) *Regularity of paths*. For every $T > 0$, $x \in \mathcal{X}$ and $w \in \mathcal{W}$, the map $\Phi_T(x, w) : [0, T] \to \mathcal{X}$ defined by

$$\Phi_T(x, w)(t) = \varphi_t(x, \theta_{T-t} w)$$

belongs to $\mathcal{D}([0, T], \mathcal{X})$.

(SDS2) *Continuous dependence*. The maps $(x, w) \mapsto \Phi_T(x, w)$ are continuous from $\mathcal{X} \times \mathcal{W}$ to $\mathcal{D}([0, T], \mathcal{X})$ for every $T > 0$.

(SDS3) *Cocycle property*. The family of mappings φ_t satisfies

$$\begin{aligned} &\varphi_0(x, w) = x, \\ &\varphi_{s+t}(x, w) = \varphi_s\bigl(\varphi_t(x, \theta_s w), w\bigr), \end{aligned} \tag{2.3}$$

for all $s, t > 0$, all $x \in \mathcal{X}$ and all $w \in \mathcal{W}$.

REMARK 2.8. The above definition is very close to the definition of *Markovian random dynamical system* introduced in [4]. Beyond the technical differences, the main difference is a shift in the viewpoint: a Markovian RDS is built on top of an RDS, so one can analyze it from both a semigroup point of view and an RDS point of view. In the case of an SDS as defined above, there is no underlying RDS (although one can always construct one), so the semigroup point of view is the only one we consider.

REMARK 2.9. The cocycle property (2.3) looks different from the cocycle property for random dynamical systems. Actually, in our case φ is a *backward cocycle* for θ_t, which is reasonable since, as a "left inverse" for $\mathcal{P}_t$, θ_t actually pushes time backward. Notice also that, unlike in the definition of RDS, we require some continuity property with respect to the noise to hold. This continuity property sounds quite restrictive, but it is actually mainly a matter of choosing a topology on $\mathcal{W}$, which is in a sense "compatible" with the topology on $\mathcal{X}$.

Similarly, we define a continuous (where "continuous" should be thought of as continuous with respect to time) SDS by

DEFINITION 2.10. An SDS is said to be *continuous* if $\mathcal{D}([0,T],\mathcal{X})$ can be replaced by $\mathcal{C}([0,T],\mathcal{X})$ in the above definition.

REMARK 2.11. One can check that the embeddings $\mathcal{C}([0,T],\mathcal{X}) \hookrightarrow \mathcal{D}([0,T],\mathcal{X})$ and $\mathcal{C}(\mathbf{R}_+,\mathcal{X}) \hookrightarrow \mathcal{D}(\mathbf{R}_+,\mathcal{X})$ are continuous, so a continuous SDS also satisfies Definition 2.7 of an SDS.

Given an SDS as in Definition 2.7 and an initial condition $x_0 \in \mathcal{X}$, we now turn to the construction of a stochastic process with initial condition x_0 constructed in a natural way from φ. First, given $t \geq 0$ and $(x,w) \in \mathcal{X} \times \mathcal{W}$, we construct a probability measure $\mathcal{Q}_t(x,w;\cdot)$ on $\mathcal{X} \times \mathcal{W}$ by

$$\mathcal{Q}_t(x,w;A \times B) = \int_B \delta_{\varphi_t(x,w')}(A)\mathcal{P}_t(w,dw'), \tag{2.4}$$

where δ_x denotes the delta measure located at x. The following result is elementary:

LEMMA 2.12. *Let φ be an SDS on $\mathcal{X}$ over $(\mathcal{W},\{\mathcal{P}_t\}_{t\geq 0},\mathbf{P}_w,\{\theta_t\}_{t\geq 0})$ and define the family of measures $\mathcal{Q}_t(x,w;\cdot)$ by* (2.4)*. Then $\mathcal{Q}_t$ is a Feller transition semigroup on $\mathcal{X} \times \mathcal{W}$. Furthermore, it has the property that if $\Pi^*_{\mathcal{W}}\mu = \mathbf{P}_w$ for a measure μ on $\mathcal{X} \times \mathcal{W}$, then $\Pi^*_{\mathcal{W}}\mathcal{Q}_t\mu = \mathbf{P}_w$.*

PROOF. The fact that $\Pi^*_{\mathcal{W}}\mathcal{Q}_t\mu = \mathbf{P}_w$ follows from the invariance of $\mathbf{P}_w$ under $\mathcal{P}_t$. We now check that $\mathcal{Q}_t$ is a Feller transition semigroup. Conditions (i) and (iii) follow immediately from the properties of φ. The continuity of $\mathcal{Q}_t(x,w;\cdot)$ with respect to (x,w) is a straightforward consequence of the facts that $\mathcal{P}_t$ is Feller and that $(x,w) \mapsto \varphi_t(x,w)$ is continuous [the latter statement follows from (SDS2) and the definition of the topology on $\mathcal{D}([0,t],\mathcal{X})$].

It thus remains only to check that the Chapman–Kolmogorov equation holds. We have from the cocycle property:

$$
\begin{aligned}
&\mathcal{Q}_{s+t}(x, w; A \times B) \\
&\quad = \int_B \delta_{\varphi_{s+t}(x,w')}(A)\mathcal{P}_{s+t}(w, dw') \\
&\quad = \int_B \int_{\mathcal{X}} \delta_{\varphi_s(y,w')}(A)\delta_{\varphi_t(x,\theta_s w')}(dy)\mathcal{P}_{s+t}(w, dw') \\
&\quad = \int_{\mathcal{W}} \int_B \int_{\mathcal{X}} \delta_{\varphi_s(y,w')}(A)\delta_{\varphi_t(x,\theta_s w')}(dy)\mathcal{P}_s(w,'' dw')\mathcal{P}_t(w, dw'').
\end{aligned}
$$

The claim then follows from the property $\theta_s^*\mathcal{P}_s(w,'' dw') = \delta_{w''}(dw')$ by exchanging the order of integration. $\square$

REMARK 2.13. Actually, (2.4) defines the evolution of the one-point process generated by φ. The n-points process would evolve according to

$$
\mathcal{Q}_t^{(n)}(x_1, \dots, x_n, w; A_1 \times \cdots \times A_n \times B) = \int_B \prod_{i=1}^n \delta_{\varphi_t(x_i,w')}(A_i)\mathcal{P}_t(w, dw').
$$

One can check as above that this defines a Feller transition semigroup on $\mathcal{X}^n \times \mathcal{W}$.

This lemma suggests the following definition:

DEFINITION 2.14. Let φ be an SDS as above. Then a probability measure μ on $\mathcal{X} \times \mathcal{W}$ is called a *generalized initial condition* for φ if $\Pi_{\mathcal{W}}^*\mu = \mathbf{P}_w$. We denote by $\mathscr{M}_\varphi$ the space of generalized initial conditions endowed with the topology of weak convergence. Elements of $\mathscr{M}_\varphi$ that are of the form $\mu = \delta_x \times \mathbf{P}_w$ for some $x \in \mathcal{X}$ will be called *initial conditions*.

Given a generalized initial condition μ, it is natural to construct a stochastic process (x_t, w_t) on $\mathcal{X} \times \mathcal{W}$ by drawing its initial condition according to μ and then evolving it according to the transition semigroup $\mathcal{Q}_t$. The marginal x_t of this process on $\mathcal{X}$ will be called the *process generated by* φ *for* μ. We will denote by $\mathscr{Q}\mu$ the law of this process [i.e., $\mathscr{Q}\mu$ is a measure on $\mathcal{D}(\mathbf{R}_+, \mathcal{X})$ in the general case and a measure on $\mathcal{C}(\mathbf{R}_+, \mathcal{X})$ in the continuous case]. More rigorously, we define for every $T > 0$ the measure $\mathscr{Q}_T\mu$ on $\mathcal{D}([0, T], \mathcal{X})$ by

$$
\mathscr{Q}_T\mu = \Phi_T^*\mathcal{P}_t\mu,
$$

where Φ_T is defined as in (SDS1). By the embedding $\mathcal{D}([0, T], \mathcal{X}) \hookrightarrow \mathcal{D}(\mathbf{R}_+, \mathcal{X})$, this actually gives a family of measures on $\mathcal{D}(\mathbf{R}_+, \mathcal{X})$. It follows from the cocycle property that the restriction to $\mathcal{D}([0, T], \mathcal{X})$ of $\mathscr{Q}_{T'}\mu$ with $T' > T$ is equal to $\mathscr{Q}_T\mu$. The definition of the topology on $\mathcal{D}(\mathbf{R}_+, \mathcal{X})$ does therefore imply that the sequence $\mathscr{Q}_T\mu$ converges weakly to a unique measure on $\mathcal{D}(\mathbf{R}_+, \mathcal{X})$ that we denote by $\mathscr{Q}\mu$. A similar argument, combined with (SDS2), yields

LEMMA 2.15. *Let φ be an SDS. Then, the operator $\mathscr{Q}$ as defined above is continuous from $\mathscr{M}_\varphi$ to $\mathscr{M}_1(\mathcal{D}(\mathbf{R}_+, \mathcal{X}))$.* $\square$

This in turn motivates the following equivalence relation:

DEFINITION 2.16. Two generalized initial conditions μ and ν of an SDS φ are *equivalent* if the processes generated by μ and ν are equal in law. In short, $\mu \sim \nu \Leftrightarrow \mathscr{Q}\mu = \mathscr{Q}\nu$.

The physical interpretation of this notion of equivalence is that the noise space contains some redundant information that is not required to construct the future of the system. Note that this does not necessarily mean that the noise space could be reduced in order to have a more "optimal" description of the system. For example, if the process x_t generated by any generalized initial condition is Markov, then all the information contained in $\mathcal{W}$ is redundant in the above sense (i.e., μ and ν are equivalent if $\Pi_{\mathcal{X}}^*\mu = \Pi_{\mathcal{X}}^*\nu$). This does of course not mean that $\mathcal{W}$ can be entirely thrown away in the above description (otherwise, since the map φ is deterministic, the evolution would become deterministic).

The main reason for introducing the notion of SDS is to have a framework in which one can study ergodic properties of physical systems with memory. It should be noted that it is designed to describe systems where the memory is *extrinsic*, as opposed to systems with *intrinsic* memory like stochastic delay equations. We present in the next section a few elementary ergodic results in the framework of SDS.

2.3. *Ergodic properties.* In the theory of Markov processes, the main tool for investigating ergodic properties is the *invariant measure*. In the setup of SDS, we say that a measure μ on $\mathcal{X} \times \mathcal{W}$ is invariant for the SDS φ if it is invariant for the Markov transition semigroup $\mathcal{Q}_t$ generated by φ. We say that a measure μ on $\mathcal{X} \times \mathcal{W}$ is *stationary* for φ if one has

$$\mathcal{Q}_t\mu \sim \mu \qquad \forall\, t > 0,$$

that is, if the process on $\mathcal{X}$ generated by μ is stationary. Following our philosophy of considering only what happens on the state space $\mathcal{X}$, we should be interested in stationary measures, disregarding completely whether they are actually invariant or not. In doing so, we could be afraid of losing many convenient results from the well-developed theory of Markov processes. Fortunately, the following lemma shows that the set of invariant measures and the set of stationary measures are actually the same, when quotiented by the equivalence relation of Definition 2.16.

PROPOSITION 2.17. *Let φ be an SDS and let μ be a stationary measure for φ. Then, there exists a measure $\mu_\star \sim \mu$ which is invariant for φ.*

PROOF. Define the ergodic averages

$$\mathcal{R}_T\mu = \frac{1}{T}\int_0^T \mathcal{Q}_t\mu\,dt. \tag{2.5}$$

Since μ is stationary, we have $\Pi_{\mathcal{X}}^*\mathcal{R}_T\mu = \Pi_{\mathcal{X}}^*\mu$ for every T. Furthermore, $\Pi_{\mathcal{W}}^*\mathcal{R}_T\mu = \mathbf{P}_w$ for every T; therefore the sequence of measures $\mathcal{R}_T\mu$ is tight on $\mathcal{X}\times\mathcal{W}$. Let $\mu_\star$ be any of its accumulation points in $\mathscr{M}_1(\mathcal{X}\times\mathcal{W})$. Since $\mathcal{Q}_t$ is Feller, $\mu_\star$ is invariant for $\mathcal{Q}_t$ and, by Lemma 2.15, one has $\mu_\star\sim\mu$. $\square$

From a mathematical point of view, it may in some cases be interesting to know whether the invariant measure $\mu_\star$ constructed in Proposition 2.17 is uniquely determined by μ. From an intuitive point of view, this uniqueness property should hold if the information contained in the trajectories on the state space $\mathcal{X}$ is sufficient to reconstruct the evolution of the noise. This intuition is made rigorous by the following proposition.

PROPOSITION 2.18. *Let φ be an SDS, define $\mathscr{W}_T^x$ as the σ-field on $\mathcal{W}$ generated by the map $\Phi_T(x,\cdot):\mathcal{W}\to\mathcal{D}([0,T],\mathcal{X})$ and set $\mathscr{W}_T = \bigwedge_{x\in\mathcal{X}}\mathscr{W}_T^x$. Assume that $\mathscr{W}_T\subset\mathscr{W}_{T'}$ for $T<T'$ and that $\mathscr{W}=\bigvee_{T\ge 0}\mathscr{W}_T$ is equal to the Borel σ-field on $\mathcal{W}$. Then, for μ_1 and μ_2 two invariant measures, one has the implication $\mu_1\sim\mu_2\Rightarrow\mu_1=\mu_2$.*

PROOF. Assume $\mu_1\sim\mu_2$ are two invariant measures for φ. Since $\mathscr{W}_T\subset\mathscr{W}_{T'}$ if $T<T'$, their equality follows if one can show that, for every $T>0$,

$$\mathbf{E}(\mu_1|\mathscr{X}\otimes\mathscr{W}_T) = \mathbf{E}(\mu_2|\mathscr{X}\otimes\mathscr{W}_T), \tag{2.6}$$

where $\mathscr{X}$ denotes the Borel σ-field on $\mathcal{X}$.

Since $\mu_1\sim\mu_2$, one has in particular $\Pi_{\mathcal{X}}^*\mu_1 = \Pi_{\mathcal{X}}^*\mu_2$, so let us call this measure ν. Since $\mathcal{W}$ is Polish, we then have the disintegration $x\mapsto\mu_i^x$, yielding formally $\mu_i(dx,dw)=\mu_i^x(dw)\nu(dx)$, where μ_i^x are probability measures on $\mathcal{W}$. (See [10], page 196, for a proof.) Fix $T>0$ and define the family $\mu_i^{x,T}$ of probability measures on $\mathcal{W}$ by

$$\mu_i^{x,T} = \int_{\mathcal{W}}\mathcal{P}_t(w,\cdot)\mu_i^x(dw).$$

With this definition, one has

$$\mathscr{Q}_T\mu_i = \int_{\mathcal{X}}\big(\Phi_T(x,\cdot)^*\mu_i^{x,T}\big)\nu(dx).$$

Let $e_0:\mathcal{D}([0,T],\mathcal{X})\to\mathcal{X}$ be the evaluation map at 0; then

$$\mathbf{E}(\mathscr{Q}_T\mu_i|e_0=x) = \big(\Phi_T(x,\cdot)^*\mu_i^{x,T}\big),$$

for ν-almost every $x \in \mathcal{X}$. Since $\mathscr{Q}_T\mu_1 = \mathscr{Q}_T\mu_2$, one therefore has

$$\mathbf{E}(\mu_1^{x,T} | \mathscr{W}_T^x) = \mathbf{E}(\mu_2^{x,T} | \mathscr{W}_T^x), \tag{2.7}$$

for ν-almost every $x \in \mathcal{X}$. On the other hand, the invariance of μ_i implies that, for every $A \in \mathscr{X}$ and every $B \in \mathscr{W}_T$, one has the equality

$$\mu_i(A \times B) = \int_{\mathcal{X}} \int_B \chi_A(\varphi_T(x, w))\mu_i^{x,T}(dw)\nu(dx).$$

Since $\varphi_T(x, \cdot)$ is $\mathscr{W}_T^x$-measurable and $B \in \mathscr{W}_T^x$, this is equal to

$$\int_{\mathcal{X}} \int_B \chi_A(\varphi_T(x, w))\mathbf{E}(\mu_i^{x,T} | \mathscr{W}_T^x)(dw)\nu(dx).$$

Thus (2.7) implies (2.6) and the proof of Proposition 2.18 is complete. $\square$

The existence of an invariant measure is usually established by finding a Lyapunov function. In this setting, Lyapunov functions are given by the following definition.

DEFINITION 2.19. Let φ be an SDS and let $F : \mathcal{X} \to [0, \infty)$ be a continuous function. Then F is a *Lyapunov function for* φ if it satisfies the following conditions:

(L1) The set $F^{-1}([0, C])$ is compact for every $C \in [0, \infty)$.
(L2) There exist constants C and $\gamma > 0$ such that

$$\int_{\mathcal{X}\times\mathcal{W}} F(x)(\mathcal{Q}_t\mu)(dx, dw) \le C + e^{-\gamma t} \int_{\mathcal{X}} F(x)(\Pi_{\mathcal{X}}^*\mu)(dx), \tag{2.8}$$

for every $t > 0$ and every generalized initial condition μ such that the right-hand side is finite.

It is important to notice that one does *not* require F to be a Lyapunov function for the transition semigroup $\mathcal{Q}_t$, since (2.8) is only required to hold for measures μ satisfying $\Pi_{\mathcal{W}}^*\mu = \mathbf{P}_w$. One nevertheless has the following result:

LEMMA 2.20. *Let φ be an SDS. If there exists a Lyapunov function F for φ, then there exists also an invariant measure $\mu_\star$ for φ, which satisfies*

$$\int_{\mathcal{X}\times\mathcal{W}} F(x)\mu_\star(dx, dw) \le C. \tag{2.9}$$

PROOF. Let $x \in \mathcal{X}$ be an arbitrary initial condition, set $\mu = \delta_x \times \mathbf{P}_w$ and define the ergodic averages $\mathcal{R}_T\mu$ as in (2.5). Combining (L1) and (L2) with the fact that $\Pi_{\mathcal{W}}^*\mathcal{R}_T\mu = \mathbf{P}_w$ one immediately gets the tightness of the sequence $\{\mathcal{R}_T\mu\}$. By the standard Krylov–Bogoloubov argument, any limiting point of $\{\mathcal{R}_T\mu\}$ is an invariant measure for φ. The estimate (2.9) follows from (2.8), combined with the fact that F is continuous. $\square$

This concludes our presentation of the abstract framework in which we analyze the ergodic properties of (SDE).

3. Construction of the SDS. In this section, we construct a continuous stochastic dynamical system which yields the solutions to (SDE) in an appropriate sense.

First of all, let us discuss what we mean by "solution" to (SDE).

DEFINITION 3.1. Let $\{x_t\}_{t\geq 0}$ be a stochastic process with continuous sample paths. We say that x_t is a *solution* to (SDE) if the stochastic process $N(t)$ defined by

$$N(t) = x_t - x_0 - \int_0^t f(x_s)\,ds \tag{3.1}$$

is equal in law to $\sigma B_H(t)$, where σ is as in (SDE) and $B_H(t)$ is an n-dimensional FBM with Hurst parameter H.

We will set up our SDS in such a way that, for every generalized initial condition μ, the canonical process associated to the measure $\mathscr{Q}\mu$ is a solution to (SDE). This will be the content of Proposition 3.11 below. In order to achieve this, our main task is to set up a noise process in a way which complies with Definition 2.6.

3.1. *Representation of the FBM.* In this section, we give a representation of the FBM $B_H(t)$ with Hurst parameter $H \in (0,1)$ which is suitable for our analysis. Recall that, by definition, $B_H(t)$ is a centered Gaussian process satisfying $B_H(0) = 0$ and

$$\mathbf{E}|B_H(t) - B_H(s)|^2 = |t-s|^{2H}. \tag{3.2}$$

Naturally, a *two-sided FBM* by requiring that (3.2) holds for all $s, t \in \mathbf{R}$. Notice that, unlike for the normal Brownian motion, the two-sided FBM is *not* obtained by gluing two independent copies of the one-sided FBM together at $t = 0$. We have the following useful representation of the two-sided FBM, which is also (up to the normalization constant) the representation used in the original paper [17].

LEMMA 3.2. *Let $w(t)$, $t \in \mathbf{R}$, be a two-sided Wiener process and let $H \in (0,1)$. Define for some constant α_H the process*

$$B_H(t) = \alpha_H \int_{-\infty}^0 (-r)^{H-1/2}\big(dw(r+t) - dw(r)\big). \tag{3.3}$$

Then there exists a choice of α_H such that $B_H(t)$ is a two-sided FBM with Hurst parameter H.

NOTATION 3.3. Given the representation (3.3) of the FBM with Hurst parameter H, we call w the "Wiener process associated to B_H." We also refer to $\{w(t) : t \le 0\}$ as the "past" of w and to $\{w(t) : t > 0\}$ as the "future" of w. We similarly refer to the "past" and the "future" of B_H. Notice the notion of future for B_H is different from the notion of future for w in terms of σ-algebras, since the future of B_H depends on the past of w.

REMARK 3.4. The expression (3.3) looks strange at first sight, but one should actually think of $B_H(t)$ as being given by $B_H(t) = \widetilde{B}_H(t) - \widetilde{B}_H(0)$, where

$$\widetilde{B}_H(t) = \alpha_H \int_{-\infty}^{t} (t-s)^{H-1/2}\, dw(s). \tag{3.4}$$

This expression is strongly reminiscent of the usual representation of the stationary Ornstein–Uhlenbeck process, but with an algebraic kernel instead of an exponential one. Of course, (3.4) does not make any sense since $(t-s)^{H-1/2}$ is not square integrable. Nevertheless, (3.4) has the advantage of explicitly showing the stationarity of the increments for the two-sided FBM.

3.2. *Noise spaces.* In this section, we introduce the family of spaces that will be used to model our noise. Denote by $\mathcal{C}_0^\infty(\mathbf{R}_-)$ the set of $\mathcal{C}^\infty$ function $w : (-\infty, 0] \to \mathbf{R}$ satisfying $w(0) = 0$ and having compact support. Given a parameter $H \in (0,1)$, we define for every $w \in \mathcal{C}_0^\infty(\mathbf{R}_-)$ the norm

$$\|w\|_H = \sup_{t,s \in \mathbf{R}_-} \frac{|w(t) - w(s)|}{|t-s|^{(1-H)/2}(1 + |t| + |s|)^{1/2}}. \tag{3.5}$$

We then define the Banach space $\mathcal{H}_H$ to be the closure of $\mathcal{C}_0^\infty(\mathbf{R}_-)$ under the norm $\|\cdot\|_H$. The following lemma is important in view of the framework exposed in Section 2:

LEMMA 3.5. *The spaces $\mathcal{H}_H$ are separable.*

PROOF. It suffices to find a norm $\|\cdot\|_\star$ which is stronger than $\|\cdot\|_H$ and such that the closure of $\mathcal{C}_0^\infty(\mathbf{R}_-)$ under $\|\cdot\|_\star$ is separable. One example of such a norm is given by $\|w\|_\star = \sup_{t<0} |t\dot{w}(t)|$. $\square$

Notice that it is crucial to define $\mathcal{H}_H$ as the closure of $\mathcal{C}_0^\infty$ under $\|\cdot\|_H$. If we defined it simply as the space of all functions with finite $\|\cdot\|_H$-norm, it would not be separable. (Think of the space of bounded continuous functions, versus the space of continuous functions vanishing at infinity.)

In view of the representation (3.3), we define the linear operator $\mathcal{D}_H$ on functions $w \in \mathcal{C}_0^\infty$ by

$$(\mathcal{D}_H w)(t) = \alpha_H \int_{-\infty}^{0} (-s)^{H-1/2} \big(\dot{w}(s+t) - \dot{w}(s)\big)\, ds, \tag{3.6}$$

where α_H is as in Lemma 3.2. We have the following result:

LEMMA 3.6. *Let $H \in (0,1)$ and let $\mathcal{H}_H$ be as above. Then the operator $\mathcal{D}_H$, formally defined by* (3.6), *is continuous from $\mathcal{H}_H$ into $\mathcal{H}_{1-H}$. Furthermore, the operator $\mathcal{D}_H$ has a bounded inverse, given by the formula*

$$\mathcal{D}_H^{-1} = \gamma_H \mathcal{D}_{1-H},$$

for some constant γ_H satisfying $\gamma_H = \gamma_{1-H}$.

REMARK 3.7. The operator $\mathcal{D}_H$ is actually (up to a multiplicative constant) a fractional integral of order $H - \frac{1}{2}$ which is renormalized in such a way that one gets rid of the divergence at $-\infty$. It is therefore not surprising that the inverse of $\mathcal{D}_H$ is $\mathcal{D}_{1-H}$.

PROOF OF LEMMA 3.6. For $H = \frac{1}{2}$, $\mathcal{D}_H$ is the identity and there is nothing to prove. We therefore assume in the sequel that $H \neq \frac{1}{2}$.

We first show that $\mathcal{D}_H$ is continuous from $\mathcal{H}_H$ into $\mathcal{H}_{1-H}$. One can easily check that $\mathcal{D}_H$ maps $\mathcal{C}_0^\infty$ into the set of $\mathcal{C}^\infty$ functions which converge to a constant at $-\infty$. This set can be seen to belong to $\mathcal{H}_{1-H}$ by a simple cutoff argument, so it suffices to show that $\|\mathcal{D}_H w\|_{1-H} \le C\|w\|_H$ for $w \in \mathcal{C}_0^\infty$. Assume without loss of generality that $t > s$ and define $h = t - s$. We then have

$$\begin{aligned}
&(\mathcal{D}_H w)(t) - (\mathcal{D}_H w)(s) \\
&\qquad = \alpha_H \int_{-\infty}^{s} \big((t-r)^{H-1/2} - (s-r)^{H-1/2}\big)\, dw(r) \\
&\qquad\quad + \alpha_H \int_s^t (t-r)^{H-1/2}\, dw(r).
\end{aligned}$$

Splitting the integral and integrating by parts yields

$$\begin{aligned}
&(\mathcal{D}_H w)(t) - (\mathcal{D}_H w)(s) \\
&\quad = -\alpha_H\big(H - \tfrac{1}{2}\big) \int_{s-h}^{s} (s-r)^{H-3/2}\big(w(r) - w(s)\big)\, dr \\
&\qquad + \alpha_H\big(H - \tfrac{1}{2}\big) \int_{t-2h}^{t} (t-r)^{H-3/2}\big(w(r) - w(t)\big)\, dr \\
&\qquad + \alpha_H\big(H - \tfrac{1}{2}\big) \int_{-\infty}^{s-h} \big((t-r)^{H-3/2} - (s-r)^{H-3/2}\big)\big(w(r) - w(s)\big)\, dr \\
&\qquad + \alpha_H (2h)^{H-1/2}\big(w(t) - w(s)\big) \\
&\quad \equiv T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

We estimate each of these terms separately. For T_1, we have

$$|T_1| \le C(1 + |s| + |t|)^{1/2} \int_0^h r^{H-3/2+(1-H)/2}\, dr \le C h^{H/2}(1 + |s| + |t|)^{1/2}.$$

The term T_2 is bounded by $Ch^{H/2}(1+|s|+|t|)^{1/2}$ in a similar way. Concerning T_3, we bound it by

$$\begin{aligned}|T_3| &\le C\int_h^\infty \bigl(r^{H-3/2}-(h+r)^{H-3/2}\bigr)\bigl(w(s-r)-w(s)\bigr)\,dr\\ &\le Ch\int_h^\infty r^{H-5/2}r^{(1-H)/2}(1+|s|+|r|)^{1/2}\,dr\\ &\le Ch^{H/2}(1+|s|)^{1/2}+Ch\int_h^\infty r^{H/2-2}(h+r)^{1/2}\,dr\\ &\le Ch^{H/2}(1+|s|+h)^{1/2}\le Ch^{H/2}(1+|s|+|t|)^{1/2}.\end{aligned}$$

The term T_4 is easily bounded by $Ch^{H/2}(1+|s|+|t|)^{1/2}$, using the fact that $w\in\mathcal{H}_H$. This shows that $\mathcal{D}_H$ is bounded from $\mathcal{H}_H$ to $\mathcal{H}_{1-H}$.

It remains to show that $\mathcal{D}_H\circ\mathcal{D}_{1-H}$ is a multiple of the identity. For this, notice that if $w\in\mathcal{C}_0^\infty$, then one has in the notation of [25], pages 94 and 95, the following identities:

$$(\mathcal{D}_H w)(t) = -\alpha_H\Gamma\bigl(H+\tfrac12\bigr)\bigl((I_+^{H-1/2}w)(t)-(I_+^{H-1/2}w)(0)\bigr),\qquad H>\tfrac12,$$

$$(\mathcal{D}_H w)(t) = -\alpha_H\Gamma\bigl(H+\tfrac12\bigr)\bigl((D_+^{1/2-H}w)(t)-(D_+^{1/2-H}w)(0)\bigr),\qquad H<\tfrac12.$$

Furthermore, (3.6) shows that $\mathcal{D}_H w=0$ if w is a constant. The claim then follows immediately from the fact that if $w\in\mathcal{C}_0^\infty$ and $\alpha\in(0,1)$, one has $D_+^\alpha I_+^\alpha w=w$ and $I_+^\alpha D_+^\alpha w=w$ (see [25], Theorem 2.4). $\square$

Since we want to use the operators $\mathcal{D}_H$ and $\mathcal{D}_{1-H}$ to switch between Wiener processes and FBMs, it is crucial to show that the sample paths of the two-sided Wiener process belong to every $\mathcal{H}_H$ with probability 1. Actually, what we show is that the Wiener measure can be constructed as a Borel measure on $\mathcal{H}_H$.

LEMMA 3.8. *There exists a unique Gaussian measure* W *on* $\mathcal{H}_H$ *which is such that the canonical process associated to it is a time-reversed Brownian motion.*

PROOF. We start by showing that the $\mathcal{H}_H$-norm of the Wiener paths has bounded moments of all orders. It follows from a generalization of the Kolmogorov criterion ([24], Theorem 2.1) that

$$\mathbf{E}\biggl(\sup_{s,t\in[0,2]}\frac{|w(s)-w(t)|}{|s-t|^{(1-H)/2}}\biggr)^p<\infty \tag{3.7}$$

for all $p>0$. Since the increments of w are independent, this implies that, for every $\varepsilon>0$, there exists a random variable C_1 such that

$$\sup_{|s-t|\le1}\frac{|w(s)-w(t)|}{|s-t|^{(1-H)/2}(1+|t|+|s|)^\varepsilon}<C_1, \tag{3.8}$$

with probability 1, and that all the moments of C_1 are bounded. We can therefore safely assume in the sequel that $|t-s| > 1$. It follows immediately from (3.8) and the triangle inequality that there exists a constant C such that

$$|w(s) - w(t)| \le CC_1 |t-s|(1+|t|+|s|)^{\varepsilon}, \tag{3.9}$$

whenever $|t-s| > 1$. Furthermore, it follows from the time-inversion property of the Brownian motion, combined with (3.7), that $|w|$ does not grow much faster than $|t|^{1/2}$ for large values of t. In particular, for every $\varepsilon' > 0$, there exists a random variable C_2 such that

$$|w(t)| \le C_2 (1+|t|)^{1/2+\varepsilon'} \qquad \forall t \in \mathbf{R}, \tag{3.10}$$

and that all the moments of C_2 are bounded. Combining (3.9) and (3.10), we get (for some other constant C)

$$\begin{aligned} &|w(s) - w(t)| \\ &\quad \le C C_1^{(1-H)/2} \\ &\qquad \times C_2^{(1+H)/2} |t-s|^{(1-H)/2} (1+|s|+|t|)^{(H+1)/4+\varepsilon(1-H)/2+\varepsilon'(1+H)/2}. \end{aligned}$$

The claim follows by choosing, for example, $\varepsilon = \varepsilon' = (1-H)/4$.

This is not quite enough, since we want the sample paths to belong to the closure of $\mathcal{C}_0^\infty$ under the norm $\|\cdot\|_H$. Define the function

$$(s,t) \mapsto \Gamma(s,t) = \frac{(1+|t|+|s|)^2}{|t-s|}.$$

By looking at the above proof, we see that we actually proved the stronger statement that for every $H \in (0,1)$, one can find a $\gamma > 0$ such that

$$\|w\|_{H,\gamma} = \sup_{s,t} \frac{\Gamma(s,t)^\gamma |w(s) - w(t)|}{|s-t|^{(1-H)/2}(1+|t|+|s|)^{1/2}} < \infty$$

with probability 1. Let us call $\mathcal{H}_{H,\gamma}$ the Banach space of functions with finite $\|\cdot\|_{H,\gamma}$-norm. We will show that one has the continuous inclusions:

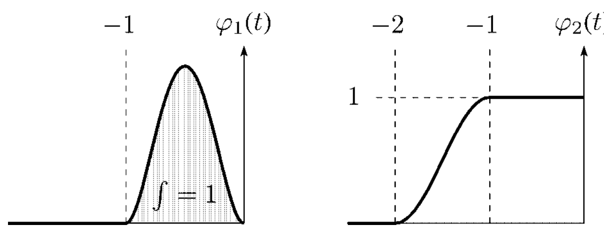
$$\mathcal{H}_{H,\gamma} \hookrightarrow \mathcal{H}_H \hookrightarrow \mathcal{C}(\mathbf{R}_-, \mathbf{R}). \tag{3.11}$$

Let us call $\widetilde{\mathsf{W}}$ the usual time-reversed Wiener measure on $\mathcal{C}(\mathbf{R}_-, \mathbf{R})$ equipped with the σ-field $\mathscr{R}$ generated by the evaluation functions. Since $\mathcal{H}_{H,\gamma}$ is a measurable subset of $\mathcal{C}(\mathbf{R}_-, \mathbf{R})$ and $\widetilde{\mathsf{W}}(\mathcal{H}_{H,\gamma}) = 1$, we can restrict $\widetilde{\mathsf{W}}$ to a measure on $\mathcal{H}_H$, equipped with the restriction $\widetilde{\mathscr{R}}$ of $\mathscr{R}$. It remains to show that $\widetilde{\mathscr{R}}$ is equal to the Borel σ-field $\mathscr{B}$ on $\mathcal{H}_H$. This follows from the fact that the evaluation functions are $\mathscr{B}$-measurable (since they are actually continuous) and that a countable number of function evaluations suffices to determine the $\|\cdot\|_H$-norm of a function. The proof of Lemma 3.8 is thus complete if we show (3.11).

Notice first that the function $\Gamma(s,t)$ becomes large when $|t-s|$ is small or when either $|t|$ or $|s|$ is large; more precisely, we have

$$\Gamma(s,t) > \max\{|s|,|t|,|t-s|^{-1}\}. \tag{3.12}$$

Therefore, functions $w \in \mathcal{H}_{H,\gamma}$ are actually more regular and have better growth properties than what is needed to have finite $\|\cdot\|_H$-norm. Given w with $\|w\|_{H,\gamma} < \infty$ and any $\varepsilon > 0$, we will construct a function $\tilde{w} \in \mathcal{C}_0^\infty$ such that $\|w - \tilde{w}\|_H < \varepsilon$. Take two $\mathcal{C}^\infty$ functions φ_1 and φ_2 with the following shape:



Furthermore, we choose them such that

$$\int_{\mathbf{R}_-} \varphi_1(s)\,ds = 1, \qquad \left|\frac{d\varphi_2(t)}{dt}\right| \le 2.$$

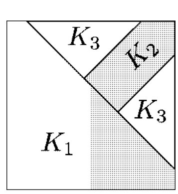
For two positive constants $r < 1$ and $R > 1$ to be chosen later, we define

$$\tilde{w}(t) = \varphi_2(t/R) \int_{\mathbf{R}_-} w(t+s) \frac{\varphi_1(s/r)}{r}\,ds,$$

that is, we smoothen out w at length scales smaller than r and we cut it off at distances bigger than R. A straightforward estimate shows that there exists a constant C such that

$$\|\tilde{w}\|_{H,\gamma} \le C\|w\|_{H,\gamma},$$

independently of $r < 1/4$ and $R > 1$. For $\delta > 0$ to be chosen later, we then divide the quadrant $K = \{(t,s) | t,s < 0\}$ into three regions:



$$K_1 = \{(t,s) | |t| + |s| \ge R\} \cap K,$$
$$K_2 = \{(t,s) | |t-s| \le \delta\} \cap K \setminus K_1,$$
$$K_3 = K \setminus (K_1 \cup K_2).$$

We then bound $\|w - \tilde{w}\|_H$ by

$$\begin{aligned}\|w - \tilde{w}\|_H &\le \sup_{(s,t)\in K_1\cup K_2} \frac{C\|w\|_{H,\gamma}}{\Gamma(t,s)^\gamma} + \sup_{(s,t)\in K_3} \frac{|w(s) - \tilde{w}(s)| + |w(t) - \tilde{w}(t)|}{|t-s|^{(1-H)/2}(1+|t|+|s|)^{1/2}} \\ &\le C(\delta^\gamma + R^{-\gamma})\|w\|_{H,\gamma} + 2\delta^{(H-1)/2} \sup_{0<t<R} |w(t) - \tilde{w}(t)|.\end{aligned}$$

By choosing δ small enough and R large enough, the first term can be made arbitrarily small. One can then choose r small enough to make the second term arbitrarily small as well. This shows that (3.11) holds and therefore the proof of Lemma 3.8 is complete. $\square$

3.3. *Definition of the SDS.* The results shown so far in this section are sufficient to construct the required SDS. We start by considering the pathwise solutions to (SDE). Given a time $T > 0$, an initial condition $x \in \mathbf{R}^n$ and a noise $b \in \mathcal{C}_0([0,T],\mathbf{R}^n)$, we look for a function $\Phi_T(x,b) \in \mathcal{C}([0,T],\mathbf{R}^n)$ satisfying

$$\Phi_T(x,b)(t) = \sigma b(t) + x + \int_0^t f\big(\Phi_T(x,b)(s)\big)\,ds. \tag{3.13}$$

We have the following standard result:

LEMMA 3.9. *Let $f:\mathbf{R}^n \to \mathbf{R}^n$ satisfy assumptions* (A1) *and* (A2). *Then, there exists a unique map $\Phi_T:\mathbf{R}^n \times \mathcal{C}([0,T],\mathbf{R}^n) \to \mathcal{C}([0,T],\mathbf{R}^n)$ satisfying* (3.13). *Furthermore, Φ_T is locally Lipschitz continuous.*

PROOF. The local (i.e., small T) existence and uniqueness of continuous solutions to (3.13) follow from a standard contraction argument. In order to show the global existence and the local Lipschitz property, fix x, b and T, and define $y(t) = x + \sigma b(t)$. Define $z(t)$ as the solution to the differential equation

$$\dot z(t) = f\big(z(t) + y(t)\big), \qquad z(0) = 0. \tag{3.14}$$

Writing down the differential equation satisfied by $\|z(t)\|^2$ and using (A1) and (A2), one sees that (3.14) possesses a (unique) solution up to time T. One can then set $\Phi_T(x,b)(t) = z(t) + y(t)$ and check that it satisfies (3.13). The local Lipschitz property of Φ_T then immediately follows from the local Lipschitz property of f. $\square$

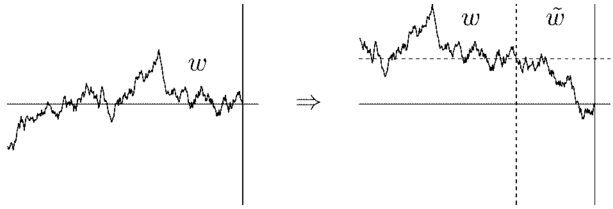
We now define the stationary noise process. For this, we define $\theta_t:\mathcal{H}_H \to \mathcal{H}_H$ by

$$(\theta_t w)(s) = w(s-t) - w(-t).$$

In order to construct the transition semigroup $\mathcal{P}_t$, we define first $\widetilde{\mathcal{H}}_H$ like $\mathcal{H}_H$, but with arguments in $\mathbf{R}_+$ instead of $\mathbf{R}_-$, and we write $\widetilde{\mathsf{W}}$ for the Wiener measure on $\widetilde{\mathcal{H}}_H$, as constructed in Lemma 3.8 above. Define the function $P_t:\mathcal{H}_H \times \widetilde{\mathcal{H}}_H \to \mathcal{H}_H$ by

$$\big(P_t(w,\tilde w)\big)(s) = \begin{cases} \tilde w(t+s) - \tilde w(t), & \text{for } s > -t, \\ w(t+s) - \tilde w(t), & \text{for } s \le -t, \end{cases} \tag{3.15}$$

and set $\mathcal{P}_t(w,\cdot) = P_t(w,\cdot)^*\widetilde{\mathsf{W}}$. This construction can be visualized by the following picture:



One then has the following.

LEMMA 3.10. *The quadruple* $(\mathcal{H}_H, \{\mathcal{P}_t\}_{t\geq 0}, \mathsf{W}, \{\theta_t\}_{t\geq 0})$ *is a stationary noise process.*

PROOF. We already know from Lemma 3.5 that $\mathcal{H}_H$ is Polish. Furthermore, one has $\theta_t \circ P_t(w, \cdot) = w$, so it remains to show that $\mathcal{P}_t$ is a Feller transition semigroup with W as its unique invariant measure. It is straightforward to check that it is a transition semigroup and the Feller property follows from the continuity of $P_t(w, \tilde{w})$ with respect to w. By the definition (3.15) and the time-reversal invariance of the Wiener process, every invariant measure for $\{\mathcal{P}_t\}_{t\geq 0}$ must have its finite-dimensional distributions coincide with those of W. Since the Borel σ-field on $\mathcal{H}_H$ is generated by the evaluation functions, this shows that W is the only invariant measure. □

We now construct an SDS over n copies of the above noise process. With a slight abuse of notation, we denote that noise process by $(\mathcal{W}, \{\mathcal{P}_t\}_{t\geq 0}, \mathsf{W}, \{\theta_t\}_{t\geq 0})$. We define the (continuous) shift operator $R_T : \mathcal{C}((-\infty, 0], \mathbf{R}^n) \to \mathcal{C}_0([0, T], \mathbf{R}^n)$ by $(R_T b)(t) = b(t - T) - b(-T)$ and set

$$
\begin{aligned}
\varphi : \mathbf{R}_+ \times \mathbf{R}^n \times \mathcal{W} &\to \mathbf{R}^n, \\
(t, x, w) &\mapsto \Phi_t(x, R_t \mathcal{D}_H w)(t).
\end{aligned}
\tag{3.16}
$$

From the above results, the following is straightforward:

PROPOSITION 3.11. *The function* φ *of* (3.16) *defines a continuous SDS over the noise process* $(\mathcal{W}, \{\mathcal{P}_t\}_{t\geq 0}, \mathsf{W}, \{\theta_t\}_{t\geq 0})$. *Furthermore, for every generalized initial condition* μ, *the process generated by* φ *from* μ *is a solution to* (SDE) *in the sense of Definition* 3.1.

PROOF. The regularity properties of φ have already been shown in Lemma 3.9. The cocycle property is an immediate consequence of the composition property for solutions of ODEs. The fact that the processes generated by φ are solutions to (SDE) is a direct consequence of (3.13), combined with Lemma 3.2, the definition of $\mathcal{D}_H$ and the fact that W is the Wiener measure. □

To conclude this section, we show that, thanks to the dissipativity condition imposed on the drift term f, the SDS defined above admits any power of the Euclidean norm on $\mathbf{R}^n$ as a Lyapunov function:

PROPOSITION 3.12. *Let φ be the continuous SDS defined above and assume that* (A1) *and* (A2) *hold. Then, for every $p \geq 2$, the map $x \mapsto \|x\|^p$ is a Lyapunov function for φ.*

PROOF. Fix $p \geq 2$ and let μ be an arbitrary generalized initial condition satisfying

$$\int_{\mathbf{R}^n} \|x\|^p (\Pi^*_{\mathbf{R}^n}\mu)(dx) < \infty.$$

Let $\tilde{\varphi}$ be the continuous SDS associated by Proposition 3.11 to the equation

$$dy(t) = -y\,dt + \sigma\,dB_H(t). \tag{3.17}$$

Notice that both φ and $\tilde{\varphi}$ are defined over the same stationary noise process.

We define x_t as the process generated by φ from μ and y_t as the process generated by $\tilde{\varphi}$ from $\delta_0 \times \mathsf{W}$ (in other words $y_0 = 0$). Since both SDS are defined over the same stationary noise process, x_t and y_t are defined over the same probability space. The process y_t is obviously Gaussian, and a direct (but lengthy) calculation shows that its variance is given by

$$\mathbf{E}\|y_t\|^2 = 2H\,\mathrm{tr}(\sigma\sigma^*)e^{-t}\int_0^t s^{2H-1}\cosh(t-s)\,ds.$$

In particular, one has for all t,

$$\mathbf{E}\|y_t\|^2 \leq 2H\,\mathrm{tr}(\sigma\sigma^*)\int_0^\infty s^{2H-1}e^{-s}\,ds = \Gamma(2H+1)\,\mathrm{tr}(\sigma\sigma^*) \equiv C_\infty. \tag{3.18}$$

Now define $z_t = x_t - y_t$. The process z_t is seen to satisfy the random differential equation given by

$$\frac{dz_t}{dt} = f(z_t + y_t) + y_t, \qquad z_0 = x_0.$$

Furthermore, one has the following equation for $\|z_t\|^2$:

$$\frac{d\|z_t\|^2}{dt} = 2\langle z_t, f(z_t + y_t)\rangle + 2\langle z_t, y_t\rangle.$$

Using (A2) and (A3) and the Cauchy–Schwarz inequality, we can estimate the right-hand side of this expression by:

$$\begin{aligned} \frac{d\|z_t\|^2}{dt} &\leq 2C_1^{(\mathrm{A1})} - 2C_2^{(\mathrm{A1})}\|z_t\|^2 + 2\langle z_t, y_t + f(y_t)\rangle \\ &\leq -2C_2^{(\mathrm{A1})}\|z_t\|^2 + \widetilde{C}(1 + \|y_t\|^2)^N, \end{aligned} \tag{3.19}$$

for some constant $\widetilde{C}$. Therefore,

$$\|z_t\|^2 \le e^{-2C_2^{(\mathrm{A1})}t}\|x_0\|^2 + \widetilde{C}\int_0^t e^{-2C_2^{(\mathrm{A1})}(t-s)}(1+\|y_s\|^2)^N\,ds.$$

It follows immediately from (3.18) and the fact that y_s is Gaussian with bounded covariance (3.18) that there exists a constant C_p such that

$$\mathbf{E}\|z_t\|^p \le C_p e^{-pC_2^{(\mathrm{A1})}t}\mathbf{E}\|x_0\|^p + C_p,$$

for all times $t > 0$. Therefore (2.8) holds and the proof of Proposition 3.12 is complete. □

4. Coupling construction.

We do now have the necessary formalism to study the long-time behavior of the SDS φ we constructed from (SDE). The main tool that will allow us to do that is the notion of self-coupling for stochastic dynamical systems.

4.1. *Self-coupling of SDS.* The main goal of this paper is to show that the asymptotic behavior of the solutions of (SDE) does not depend on its initial condition. This will then imply that the dynamics converges to a stationary state (in a suitable sense). We therefore look for a suitable way of comparing solutions to (SDE). In general, two solutions starting from different initial points in $\mathbf{R}^n$ and driven with the same realization of the noise B_H have no reason of getting close to each other as time goes by. Condition (A1) indeed only ensures that they will tend to approach each other as long as they are sufficiently far apart. This is reasonable, since by comparing only solutions driven by the same realization of the noise process, one completely forgets about the randomness of the system and the "blurring" this randomness induces.

It is therefore important to compare probability measures (e.g., on path-space) induced by the solutions rather than the solution themselves. More precisely, given an SDS φ and two generalized initial conditions μ and ν, we want to compare the measures $\mathscr{Q}\mathfrak{Q}_t\mu$ and $\mathscr{Q}\mathfrak{Q}_t\nu$ as t goes to infinity. The distance we will work with is the total variation distance, henceforth denoted by $\|\cdot\|_{\mathrm{TV}}$. We will actually use the following useful representation of the total variation distance. Let Ω be a measurable space and let $\mathbf{P}_1$ and $\mathbf{P}_2$ be two probability measures on Ω. We denote by $C(\mathbf{P}_1, \mathbf{P}_2)$ the set of all probability measures on $\Omega \times \Omega$ which are such that their marginals on the two components are equal to $\mathbf{P}_1$ and $\mathbf{P}_2$, respectively. Let furthermore $\Delta \subset \Omega \times \Omega$ denote the diagonal, that is, the set of elements of the form (ω, ω). We then have

$$\|\mathbf{P}_1 - \mathbf{P}_2\|_{\mathrm{TV}} = 2 - \sup_{\mathbf{P}\in C(\mathbf{P}_1,\mathbf{P}_2)} 2\mathbf{P}(\Delta). \tag{4.1}$$

Elements of $C(\mathbf{P}_1, \mathbf{P}_2)$ will be referred to as *couplings* between $\mathbf{P}_1$ and $\mathbf{P}_2$. This leads naturally to the following definition:

DEFINITION 4.1. Let φ be an SDS with state space $\mathcal{X}$ and let $\mathscr{M}_\varphi$ be the associated space of generalized initial conditions. A *self-coupling* for φ is a measurable map $(\mu,\nu)\mapsto \mathscr{Q}(\mu,\nu)$ from $\mathscr{M}_\varphi\times\mathscr{M}_\varphi$ into $\mathcal{D}(\mathbf{R}_+,\mathcal{X})\times\mathcal{D}(\mathbf{R}_+,\mathcal{X})$, with the property that for every pair (μ,ν), $\mathscr{Q}(\mu,\nu)$ is a coupling for $\mathscr{Q}\mu$ and $\mathscr{Q}\nu$.

Define the shift map $\Sigma_t:\mathcal{D}(\mathbf{R}_+,\mathcal{X})\to\mathcal{D}(\mathbf{R}_+,\mathcal{X})$ by

$$(\Sigma_t x)(s)=x(t+s).$$

It follows immediately from the cocycle property and the stationarity of the noise process that $\mathscr{Q}\mathcal{Q}_t\mu=\Sigma_t^*\mathscr{Q}\mu$. Therefore, the measure $\Sigma_t^*\mathscr{Q}(\mu,\nu)$ is a coupling for $\mathscr{Q}\mathcal{Q}_t\mu$ and $\mathscr{Q}\mathcal{Q}_t\nu$ [which is in general different from the coupling $\mathscr{Q}(\mathcal{Q}_t\mu,\mathcal{Q}_t\nu)$]. Our aim in the remainder of this paper is to construct a self-coupling $\mathscr{Q}(\mu,\nu)$ for the SDS associated to (SDE) which has the property that

$$\lim_{t\to\infty}\bigl(\Sigma_t^*\mathscr{Q}(\mu,\nu)\bigr)(\Delta)=1,$$

where Δ denotes as before the diagonal of the space $\mathcal{D}(\mathbf{R}_+,\mathcal{X})\times\mathcal{D}(\mathbf{R}_+,\mathcal{X})$. We will then use the inequality

$$\|\mathscr{Q}\mathcal{Q}_t\mu-\mathscr{Q}\mathcal{Q}_t\nu\|_{\mathrm{TV}}\le 2-2\bigl(\Sigma_t^*\mathscr{Q}(\mu,\nu)\bigr)(\Delta) \tag{4.2}$$

to deduce the uniqueness of the stationary state for (SDE).

In the remainder of the paper, the general way of constructing such a self-coupling will be the following. First, we fix a Polish space $\mathcal{A}$ that contains some auxiliary information on the dynamics of the coupled process we want to keep track of. We also define a "future" noise space $\mathcal{W}_+$ to be equal to $\widetilde{\mathcal{H}}_H^n$, where $\widetilde{\mathcal{H}}_H$ is as in (3.15). There is a natural continuous time-shift operator on $\mathbf{R}\times\mathcal{W}\times\mathcal{W}_+$ defined for $t>0$ by

$$(s,w,\tilde w)\mapsto\bigl(s-t,P_t(w,\tilde w),S_t\tilde w\bigr),\qquad (S_t\tilde w)(r)=\tilde w(r+t)-\tilde w(t), \tag{4.3}$$

where P_t was defined in (3.15). We then construct a (measurable) map

$$\begin{aligned}\mathscr{C}:\mathcal{X}^2\times\mathcal{W}^2\times\mathcal{A}&\to\mathbf{R}\times\mathscr{M}_1(\mathcal{A}\times\mathcal{W}_+^2),\\ (x,y,w_x,w_y,a)&\mapsto\bigl(T(x,y,w_x,w_y,a),\mathsf{W}_2(x,y,w_x,w_y,a)\bigr),\end{aligned} \tag{4.4}$$

with the properties that, for all (x,y,w_x,w_y,a),

(C1) the time $T(x,y,w_x,w_y,a)$ is positive and greater than 1,
(C2) the marginals of $\mathsf{W}_2(x,y,w_x,w_y,a)$ onto the two copies of $\mathcal{W}_+$ are both equal to the Wiener measure W.

We call the map $\mathscr{C}$ the "coupling map," since it yields a natural way of constructing a self-coupling for the SDS φ. The remainder of this section explains how to achieve this.

Given the map $\mathscr{C}$, we can construct a Markov process on the augmented space $\mathscr{X}=\mathcal{X}^2\times\mathcal{W}^2\times\mathbf{R}_+\times\mathcal{A}\times\mathcal{W}_+^2$ in the following way. As long as the component

$\tau \in \mathbf{R}_+$ is positive, we just time-shift the elements in $\mathcal{W}^2 \times \mathcal{W}_+^2 \times \mathbf{R}_+$ according to (4.3) and we evolve in $\mathcal{X}^2$ by solving (SDE). As soon as τ becomes 0, we redraw the future of the noise up to time $T(x, y, a)$ according to the distribution W_2, which may at the same time modify the information stored in $\mathcal{A}$.

To shorten notation, we denote elements of $\mathscr{X}$ by

$$X = (x, y, w_x, w_y, \tau, a, \tilde{w}_x, \tilde{w}_y).$$

With this notation, the transition function $\widetilde{\mathcal{Q}}_t$ for the process we just described is defined by:

(a) For $t < \tau$, we define $\widetilde{\mathcal{Q}}_t(X; \cdot)$ by

$$\widetilde{\mathcal{Q}}_t(X; \cdot) = \delta_{\varphi_t(x, P_t(w_x, \tilde{w}_x))} \times \delta_{\varphi_t(y, P_t(w_y, \tilde{w}_y))} \times \delta_{P_t(w_x, \tilde{w}_x)} \\ \times \delta_{P_t(w_y, \tilde{w}_y)} \times \delta_{\tau - t} \times \delta_a \times \delta_{S_t \tilde{w}_x} \times \delta_{S_t \tilde{w}_y}.$$

(b) For $t = \tau$, we define $\widetilde{\mathcal{Q}}_t(X; \cdot)$ by

$$\begin{aligned} \widetilde{\mathcal{Q}}_t(X; \cdot) &= \delta_{\varphi_t(x, P_t(w_x, \tilde{w}_x))} \times \delta_{\varphi_t(y, P_t(w_y, \tilde{w}_y))} \times \delta_{P_t(w_x, \tilde{w}_x)} \\ &\quad \times \delta_{P_t(w_y, \tilde{w}_y)} \times \delta_{T(x, y, P_t(w_x, \tilde{w}_x), P_t(w_y, \tilde{w}_y), a)} \\ &\quad \times \mathsf{W}_2(x, y, P_t(w_x, \tilde{w}_x), P_t(w_y, \tilde{w}_y), a). \end{aligned} \tag{4.5}$$

(c) For $t > \tau$, we define $\widetilde{Q}_t$ by imposing that the Chapman–Kolmogorov equations hold. Since we assumed that $T(x, y, w_x, w_y, a)$ is always greater than 1, this procedure is well defined.

We now construct an initial condition for this process, given two generalized initial conditions μ_1 and μ_2 for φ. We do this in such a way that, in the beginning, the noise component of our process lives on the diagonal of the space $\mathcal{W}^2$. In other words, the two copies of the two-sided FBM driving our coupled system have the same past. This is possible since the marginals of μ_1 and μ_2 on $\mathcal{W}$ coincide. Concerning the components of the initial condition in $\mathbf{R}_+ \times \mathcal{A} \times \mathcal{W}_+^2$, we just draw them according to the map $\mathscr{C}$, with some distinguished element $a_0 \in \mathcal{A}$.

We call $\mathscr{Q}_0(\mu_1, \mu_2)$ the measure on $\mathscr{X}$ constructed by this procedure. Consider a cylindrical subset of $\mathscr{X}$ of the form

$$X = X_1 \times X_2 \times W_1 \times W_2 \times F,$$

where F is a measurable subset of $\mathbf{R}_+ \times \mathcal{A} \times \mathcal{W}_+^2$. We make use of the disintegration $w \mapsto \mu_i^w$, yielding formally $\mu_i(dx, dw) = \mu_i^w(dx)\mathsf{W}(dw)$, and we define $\mathscr{Q}_0(\mu_1, \mu_2)$ by

$$\begin{aligned} &\mathscr{Q}_0(\mu_1, \mu_2)(X) \\ &\quad = \int_{W_1 \cap W_2} \int_{X_1} \int_{X_2} \big(\delta_{T(x_1, x_2, w, w, a_0)} \times \mathsf{W}_2(x_1, x_2, w, w, a_0)\big)(F) \\ &\qquad \times \mu_2^w(dx_2)\mu_1^w(dx_1)\mathsf{W}(dw). \end{aligned} \tag{4.6}$$

With this definition, we finally construct the self-coupling $\mathscr{Q}(\mu_1, \mu_2)$ of φ corresponding to the function $\mathscr{C}$ as the marginal on $\mathcal{C}(\mathbf{R}_+, \mathcal{X}) \times \mathcal{C}(\mathbf{R}_+, \mathcal{X})$ of the process generated by the initial condition $\mathscr{Q}_0(\mu_1, \mu_2)$ evolving under the semigroup given by $\widetilde{\mathcal{Q}}_t$. Condition (C2) ensures that this is indeed a coupling for $\mathscr{Q}\mu_1$ and $\mathscr{Q}\mu_2$.

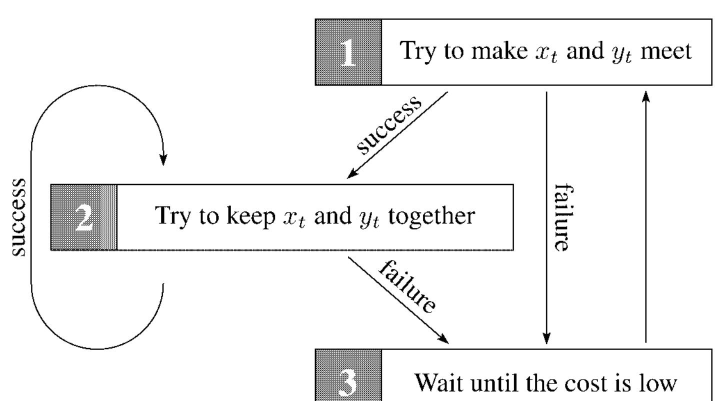
The following section gives an overview of the way the coupling function $\mathscr{C}$ is constructed.

4.2. *Construction of the coupling function.* Let us consider that the initial conditions μ_1 and μ_2 are fixed once and for all and denote by x_t and y_t the two $\mathcal{X}$-valued processes obtained by considering the marginals of $\mathscr{Q}(\mu_1, \mu_2)$ on its two $\mathcal{X}$ components. Define the random (but not stopping) time τ_∞ by

$$\tau_\infty = \inf\{t > 0 | x_s = y_s \text{ for all } s > t\}.$$

Our aim is to find a space $\mathcal{A}$ and a function $\mathscr{C}$ satisfying (C1) and (C2) such that the processes x_t and y_t eventually meet and stay together for all times, that is, such that $\lim_{T\to\infty} \mathbf{P}(\tau_\infty < T) = 1$. If the noise process driving the system was Markov, the "stay together" part of this statement would not be a problem, since it would suffice to start driving x_t and y_t with identical realizations of the noise as soon as they meet. Since the FBM is not Markov, it is possible to make the future realizations of two copies coincide with probability 1 only if the past realizations also coincide. If the past realizations do not coincide for some time, we interpret this as introducing a "cost" into the system, which we need to master. (This notion of cost will be made precise in Definition 5.3 below.) Fortunately, the memory of past events becomes smaller and smaller as time goes by, which can be interpreted as a natural tendency of the cost to decrease. This way of interpreting our system leads to the following algorithm that should be implemented by the coupling function $\mathscr{C}$:

(4.7)



The precise meaning of the statements appearing in this diagram will be made clear in the sequel, but the general idea of the construction should be clear by now.

One step in (4.7) corresponds to the time between two jumps of the τ-component of the coupled process. Our aim is to construct the coupling function $\mathscr{C}$ in such a way that, with probability 1, there is a time after which step 2 always succeeds. This time is then precisely the random time τ_∞ we want to estimate.

It is clear from what has just been exposed that we will actually never need to consider the continuous-time process on the space $\mathscr{X}$ given by the self-coupling described in the previous section, but it is sufficient to describe what happens at the beginning of each step in (4.7). We will therefore only consider the discrete-time dynamic obtained by sampling the continuous-time system just before each step. The discrete-time dynamic will take place on the space $\mathcal{Z} = (\mathcal{X}^2 \times \mathcal{W}^2 \times \mathcal{A}) \times \mathbf{R}_+$ and we will denote its elements by

$$(Z, \tau), \qquad Z = (x, y, w_x, w_y, a), \qquad \tau \in \mathbf{R}_+.$$

Since the time steps of the discrete dynamic are not equally spaced, the time τ is required to keep track of how much time really elapsed. The dynamic of the discrete process (Z_n, τ_n) on $\mathcal{Z}$ is determined by the function $\Phi : \mathbf{R}_+ \times \mathcal{Z} \times (\mathcal{A} \times \mathcal{W}_+^2) \to \mathcal{Z}$ given by

$$\begin{aligned}\Phi\bigl(t, (Z, \tau), (\tilde{w}_x, \tilde{w}_y, \tilde{a})\bigr) = \bigl(&\varphi_t\bigl(x, P_t(w_x, \tilde{w}_x)\bigr), \varphi_t\bigl(y, P_t(w_y, \tilde{w}_y)\bigr),\\ &P_t(w_x, \tilde{w}_x), P_t(w_y, \tilde{w}_y), \tilde{a}, \tau + t\bigr).\end{aligned}$$

(The notation are the same as in the definition of $\tilde{\mathcal{Q}}_t$ above.) With this definition at hand, the transition function for the process (Z_n, τ_n) is given by

$$\mathscr{P}(Z, \tau) = \Phi\bigl(T(Z), (Z, \tau), \cdot\bigr)^* \mathsf{W}_2(Z), \tag{4.8}$$

where T and W_2 are defined in (4.4). Given two generalized initial conditions μ_1 and μ_2 for the original SDS, the initial condition (Z_0, τ_0) is constructed by choosing $\tau_0 = 0$ and by drawing Z_0 according to the measure

$$\mu_0(X) = \delta_{a_0}(A) \int_{W_1 \cap W_2} \int_{X_1} \int_{X_2} \mu_2^w(dx_2)\mu_1^w(dx_1)\mathsf{W}(dw),$$

where X is a cylindrical set of the form $X = X_1 \times X_2 \times W_1 \times W_2 \times A$. It follows from the definitions (4.5) and (4.6) that if we define τ_n as the nth jump of the process on $\mathscr{X}$ constructed above and Z_n as (the component in $\mathcal{X}^2 \times \mathcal{W}^2 \times \mathcal{A}$ of) its left-hand limit at τ_n, the process we obtain is equal in law to the Markov chain that we just constructed.

Before carrying further on with the construction of $\mathscr{C}$, we make a few preliminary computations to see how changes in the past of the FBM affect its future. The formulae and estimates obtained in Section 4.3 are crucial for the construction of $\mathscr{C}$ and for the obtainment of the bounds that lead to Theorems 1.2 and 1.3. In particular, Proposition 4.4 is the main estimate that leads to the coherence of the coupling construction and to the bounds on the convergence rate toward the stationary state.

4.3. *Influence of the past on the future.* Let $w_x \in \mathcal{H}_H$ and set $B_x = \mathcal{D}_H w_x$. Consider furthermore two functions g_w and g_B satisfying

$$t \mapsto \int_0^t g_w(s)\,ds \in \mathcal{H}_H, \qquad t \mapsto \int_0^t g_B(s)\,ds \in \mathcal{H}_{1-H}, \tag{4.9}$$

and define B_y and w_y by $B_y(0) = w_y(0) = 0$ and

$$dB_y = dB_x + g_B\,dt, \qquad dw_y = dw_x + g_w\,dt. \tag{4.10}$$

As an immediate consequence of the definition of $\mathcal{D}_H$, the following relations between g_w and g_B will ensure that $B_y = \mathcal{D}_H w_y$.

LEMMA 4.2. *Let B_x, B_y, w_x, w_y, g_B and g_w be as in* (4.9), (4.10) *and assume that $B_x = \mathcal{D}_H w_x$ and $B_y = \mathcal{D}_H w_y$. Then, g_w and g_B satisfy the following relation*:

$$g_w(t) = \alpha_H \frac{d}{dt} \int_{-\infty}^t (t-s)^{1/2-H} g_B(s)\,ds, \tag{4.11a}$$

$$g_B(t) = \gamma_H \alpha_{1-H} \frac{d}{dt} \int_{-\infty}^t (t-s)^{H-1/2} g_w(s)\,ds. \tag{4.11b}$$

If $g_w(t) = 0$ for $t > t_0$, one has

$$g_B(t) = \bigl(H - \tfrac{1}{2}\bigr)\gamma_H \alpha_{1-H} \int_{-\infty}^{t_0} (t-s)^{H-3/2} g_w(s)\,ds, \tag{4.11c}$$

for $t \geq t_0$. Similarly, if $g_B(t) = 0$ for $t > t_0$, one has

$$g_w(t) = \bigl(\tfrac{1}{2} - H\bigr)\alpha_H \int_{-\infty}^{t_0} (t-s)^{-H-1/2} g_B(s)\,ds, \tag{4.11d}$$

for $t \geq t_0$. If g_w is differentiable for $t > t_0$ and $g_w(t) = 0$ for $t < t_0$, one has

$$g_B(t) = \frac{\gamma_H \alpha_{1-H} g_w(t_0)}{(t-t_0)^{1/2-H}} + \gamma_H \alpha_{1-H} \int_{t_0}^t \frac{g_w'(s)}{(t-s)^{1/2-H}}\,ds, \tag{4.11e}$$

for $t \geq t_0$. Similarly, if g_B is differentiable for $t > t_0$ and $g_B(t) = 0$ for $t < t_0$, one has

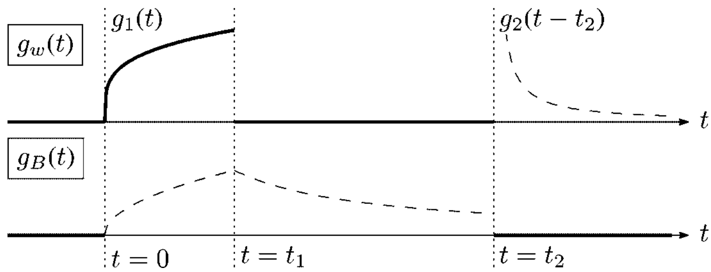
$$g_w(t) = \frac{\alpha_H g_B(t_0)}{(t-t_0)^{H-1/2}} + \alpha_H \int_{t_0}^t \frac{g_B'(s)}{(t-s)^{H-1/2}}\,ds, \tag{4.11f}$$

for $t \geq t_0$.

PROOF. The claims (4.11a) and (4.11b) follow immediately from (4.10), using the linearity of $\mathcal{D}_H$ and the inversion formula. The other claims are simply obtained by differentiating under the integral; see [25] for a justification. $\square$

We will be led in the sequel to consider the following situation, where t_1, t_2 and g_1 are assumed to be given:

(4.12)



In this figure, g_w and g_B are related by (4.11a) and (4.11b) as before. The boldfaced regions indicate that we consider the corresponding parts of g_w or g_B to be given. The dashed regions indicate that those parts of g_w and g_B are computed from the boldfaced regions by using the relations (4.11a) and (4.11b). The picture is coherent since (4.11a) and (4.11b) in both cases only use information about the past to compute the present. One should think of the interval $[0, t_1]$ as representing the time spent on steps 1 and 2 of the algorithm (4.7). The interval $[t_1, t_2]$ corresponds to the waiting time, that is, step 3. Let us first give an explicit formula for g_2 in terms of g_1:

LEMMA 4.3. *Consider the situation of Proposition* 4.4. *Then, g_2 is given by*

$$g_2(t) = C \int_0^{t_1} \frac{t^{1/2-H}(t_2-s)^{H-1/2}}{t+t_2-s} g_1(s)\, ds, \tag{4.13}$$

with a constant C depending only on H.

PROOF. We extend $g_1(t)$ to the whole real line by setting it equal to 0 outside of $[0, t_1]$. Using Lemma 4.2, we see that, for some constant C and for $t > t_2$,

$$\begin{aligned}
g_2(t-t_2) &= C \int_0^{t_2} (t-s)^{-H-1/2} g_B(s)\, ds \\
&= C \int_0^{t_2} (t-s)^{-H-1/2} \frac{d}{ds} \int_0^s (s-r)^{H-1/2} g_1(r)\, dr\, ds \\
&= C(t-t_2)^{-H-1/2} \int_0^{t_2} (t_2-r)^{H-1/2} g_1(r)\, dr \\
&\quad - C\big(H+\tfrac{1}{2}\big) \int_0^{t_2} (t-s)^{-H-3/2} \int_0^s (s-r)^{H-1/2} g_1(r)\, dr\, ds \\
&\equiv C \int_0^{t_1} K(t,r) g_1(r)\, dr,
\end{aligned}$$

where the integration stops at t_1 because g_1 is equal to 0 for larger values of t. The kernel K is given by

$$\begin{aligned}K(t,r) &= (t-t_2)^{-H-1/2}(t_2-r)^{H-1/2}\\ &\qquad - \big(H+\tfrac{1}{2}\big)\int_r^{t_2}(t-s)^{-H-3/2}(s-r)^{H-1/2}\,ds\\ &= \frac{(t-t_2)^{-H-1/2}(t_2-r)^{H+1/2}}{t-r}\left(\frac{t-r}{t_2-r}-1\right)\\ &= \frac{(t-t_2)^{1/2-H}(t_2-r)^{H-1/2}}{t-r},\end{aligned}$$

and the claim follows. $\square$

We give now estimates on g_2 in terms of g_1. To this end, given $\alpha > 0$, we introduce the following norm on functions $g:\mathbf{R}_+ \to \mathbf{R}^n$:

$$\|g\|_\alpha^2 = \int_0^\infty (1+t)^{2\alpha}\|g(t)\|^2\,dt.$$

The following proposition is essential to the coherence of our coupling construction:

PROPOSITION 4.4. *Let $t_2 > 2t_1 > 0$, let $g_1:[0,t_1] \to \mathbf{R}^n$ be a square integrable function, and define $g_2:\mathbf{R}_+ \to \mathbf{R}_+$ by*

$$g_2(t) = \int_0^{t_1} \frac{t^{1/2-H}(t_2-s)^{H-1/2}}{t+t_2-s}\|g_1(s)\|\,ds.$$

Then, for every α satisfying

$$0 < \alpha < \min\{\tfrac{1}{2}; H\},$$

there exists a constant $\kappa > 0$ depending only on α and H such that the estimate

$$\|g_2\|_\alpha \le \kappa \left|\frac{t_2}{t_1}\right|^{\alpha-1/2}\|g_1\|_\alpha \tag{4.14}$$

holds.

REMARK 4.5. The important features of this proposition are that the constant κ *does not* depend on t_1 or t_2 and that the exponent in (4.14) is negative.

PROOF OF PROPOSITION 4.4. We define $r = t_2/t_1$ to shorten notation. Using (4.13) and Cauchy–Schwarz, we then have

$$\begin{aligned}\|g_2(t)\| &\le C\|g_1\|_\alpha \sqrt{\int_0^{t_1} (1+s)^{-2\alpha} \frac{(rt_1 - s)^{2H-1} t^{1-2H}}{(t + rt_1 - s)^2}\, ds} \\ &= C\|g_1\|_\alpha {t_1}^{H-\alpha} t^{1-H-1/2} \sqrt{\int_0^1 s^{-2\alpha} \frac{(r-s)^{2H-1}}{(t + rt_1 - t_1 s)^2}\, ds} \\ &\le C\|g_1\|_\alpha \frac{{t_1}^{H-\alpha} t^{1/2-H} r^{H-1/2}}{t + (r-1)t_1},\end{aligned}$$

where we made use of the assumptions that $2\alpha < 1$ and $r \ge 2$. Therefore, $\|g_2\|_\alpha$ is bounded by

$$\begin{aligned}\|g_2\|_\alpha &\le \kappa \|g_1\|_\alpha {t_1}^{H-\alpha} r^{H-1/2} \sqrt{\int_0^\infty \frac{(1+t)^{2\alpha} t^{1-2H}}{(t + (r-1)t_1)^2}\, dt} \\ &\le \kappa \|g_1\|_\alpha r^{\alpha - 1/2} \sqrt{\int_0^\infty \frac{t^{2\alpha} t^{1-2H}}{(t+1)^2}\, dt},\end{aligned}$$

for some constant κ, where the last inequality was obtained through the change of variables $t \mapsto (r-1)t_1 t$ and used the fact that $r \ge 2$. The convergence of the integral is obtained under the condition $\alpha < H$ which is verified by assumption, so the proof of Proposition 4.4 is complete. □

We will construct our coupling function $\mathscr{C}$ in such a way that there always exist functions g_w and g_B satisfying (4.9) and (4.10), where w_x and w_y denote the noise components of our coupling process, and B_x and B_y are obtained by applying the operator $\mathcal{D}_H$ to them. We have now all the necessary ingredients for the construction of $\mathscr{C}$.

5. Definition of the coupling function. Our coupling construction depends on a parameter $\alpha < \min\{\frac{1}{2}, H\}$ which we fix once and for all. This parameter will then be tuned in Section 6.

First of all, we define the auxiliary space $\mathcal{A}$:

$$\mathcal{A} = \{0, 1, 2, 3\} \times \mathbf{N} \times \mathbf{N} \times \mathbf{R}_+. \tag{5.1}$$

Elements of $\mathcal{A}$ will be denoted by

$$a = (S, N, \widetilde{N}, T_3). \tag{5.2}$$

The component S denotes which step of (4.7) is going to be performed next (the value 0 will be used only for the initial value a_0). The counter N is incremented every time step 2 is performed and is reset to 0 every time another step is performed. The counter $\widetilde{N}$ is incremented every time step 1 or step 2 fails. If steps 1 or 2 fail, the time T_3 contains the duration of the upcoming step 3. We take

$$a_0 = (0, 1, 1, 0)$$

as initial condition for our coupling construction.

Remember that the coupling function $\mathscr{C}$ is a function from $\mathcal{X}^2 \times \mathcal{W}^2 \times \mathcal{A}$, representing the state of the system at the end of a step, into $\mathbf{R} \times \mathscr{M}_1(\mathcal{A} \times \mathcal{W}_+^2)$, representing the duration and the realization of the noise for the next step. We now define $\mathscr{C}$ for the four possible values of S.

5.1. *Initial stage* ($S = 0$). Notice first that (A1) implies that

$$\frac{\langle f(y) - f(x), y - x\rangle}{\|y - x\|} \le C_4^{(\mathrm{A1})} - C_2^{(\mathrm{A1})}\|y - x\|, \tag{5.3}$$

where we set $C_4^{(\mathrm{A1})} = \sqrt{C_1^{(\mathrm{A1})}(C_2^{(\mathrm{A1})} + C_3^{(\mathrm{A1})})}$.

In the beginning, we just wait until the two copies of our process are within distance $1 + (C_4^{(\mathrm{A1})}/C_2^{(\mathrm{A1})})$ of each other. If x_t and y_t satisfy (SDE) with the same realization of the noise process B_H, and $\rho_t = y_t - x_t$, we have for $\|\rho_t\|$ the differential inequality

$$\frac{d\|\rho_t\|}{dt} = \frac{\langle f(y_t) - f(x_t), \rho_t\rangle}{\|\rho_t\|} \le C_4^{(\mathrm{A1})} - C_2^{(\mathrm{A1})}\|\rho_t\|,$$

and therefore by Gronwall's lemma

$$\|\rho_t\| \le \|y_0 - x_0\| e^{-C_2^{(\mathrm{A1})}t} + \frac{C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}}\bigl(1 - e^{-C_2^{(\mathrm{A1})}t}\bigr).$$

It is enough to wait for a time $t = (\log\|y_0 - x_0\|)/C_2^{(\mathrm{A1})}$ to ensure that $\|\rho_t\| \le 1 + (C_4^{(\mathrm{A1})}/C_2^{(\mathrm{A1})})$, so we define the coupling function $\mathscr{C}$ in this case by

$$T(Z, a_0) = \max\left\{\frac{\log\|y_0 - x_0\|}{C_2^{(\mathrm{A1})}}, 1\right\}, \qquad \mathsf{W}_2(Z, a_0) = \Delta^*\mathsf{W} \times \delta_{a'}, \tag{5.4}$$

where the map $\Delta : \mathcal{W}_+ \to \mathcal{W}_+^2$ is defined by $\Delta(w) = (w, w)$ and the element a' is given by

$$a' = (1, 0, 0, 0).$$

In other terms, we wait until the two copies of the process are close to each other, and then we proceed to step 1.

5.2. *Waiting stage* ($S = 3$). In this stage, both copies evolve with the same realization of the underlying Wiener process. Using notation (5.2) and (4.4), we therefore define the coupling function $\mathscr{C}$ in this case by

$$T(Z, a) = T_3, \qquad \mathsf{W}_2(Z, a) = \Delta^* \mathsf{W} \times \delta_{a'}, \tag{5.5}$$

where the map Δ is defined as above and the element a' is given by

$$a' = (1, N, \widetilde{N}, 0).$$

Notice that this definition is in accordance with (4.7); that is, the counters N and $\widetilde{N}$ remain unchanged, the dynamic evolves for a time T_3 with two identical realizations of the Wiener process (note that the realizations of the FBM driving the two copies of the system are in general different, since the *pasts* of the Wiener processes may differ), and then proceeds to step 1.

5.3. *Hitting stage* ($S = 1$). In this section, we construct and then analyze the map $\mathscr{C}$ corresponding to the step 1, which is the most important step for our construction. We start with a few preliminary computations. Define $W_{1,1}$ as the space of almost everywhere differentiable functions g, such that the quantity

$$\|g\|_{1,1} = \int_0^1 \left\| \frac{dg_B(t)}{dt} \right\| dt + \|g(0)\|,$$

is finite.

LEMMA 5.1. *Let* $g_B : [0, 1] \to \mathbf{R}^n$ *be in* $W_{1,1}$ *and define* g_w *by* (4.11a) *with* $H \in (\frac{1}{2}, 1)$. (*The function* g_B *is extended to* $\mathbf{R}$ *by setting it equal to* 0 *outside of* $[0, 1]$ *and* g_w *is considered as a function from* $\mathbf{R}_+$ *to* $\mathbf{R}^n$.) *Then, for every* $\alpha \in (0, H)$, *there exists a constant* C *such that*

$$\|g_w\|_\alpha \le C \|g_B\|_{1,1}.$$

PROOF. We first bound the L^2-norm of g_w on the interval $[0, 2]$. Using (4.11f), we can bound $\|g_w(t)\|$ by

$$\|g_w(t)\| \le C \|g_B(0)\| t^{1/2-H} + C \int_0^t \|\dot{g}_B(s)\| (t-s)^{1/2-H} \, ds.$$

Since $t^{1/2-H}$ is square integrable at the origin, it remains to bound the terms I_1 and I_2 given by

$$I_1 = \int_0^2 \left(\int_0^t (t-s)^{1/2-H} \|\dot{g}_B(s)\| \, ds \int_0^t (t-r)^{1/2-H} \|\dot{g}_B(r)\| \, dr \right) dt,$$

$$I_2 = \|g_B(0)\| \int_0^2 t^{1/2-H} \int_0^t (t-s)^{1/2-H} \|\dot{g}_B(s)\| \, ds \, dt.$$

We only show how to bound I_1, as I_2 can be bounded in a similar fashion. Writing $r \vee s = \max\{r, s\}$, one has

$$I_1 = \int_0^1 \int_0^1 \int_{r \vee s}^2 (t-s)^{1/2-H} (t-r)^{1/2-H} \, dt \, \|\dot{g}_B(s)\| \|\dot{g}_B(r)\| \, dr \, ds.$$

Since

$$\int_{r \vee s}^2 (t-s)^{1/2-H} (t-r)^{1/2-H} \, dt \le \int_{r \vee s}^2 \bigl(t - (r \vee s)\bigr)^{1-2H} dt \le \frac{2^{2-2H}}{2-2H},$$

I_1 is bounded by $C \|g_B\|_{1,1}^2$.

It remains to bound the large-time tail of g_w. For $t \ge 2$, one has, again by Lemma 4.2,

$$\|g_w(t)\| \le (t-1)^{-H-1/2} \sup_{s \in [0,1]} \|g_B(s)\| \le C(t-1)^{-H-1/2} \|g_B\|_{1,1}. \tag{5.6}$$

It follows from the definition that the $\| \cdot \|_\alpha$-norm of this function is bounded if $\alpha < H$. The proof of Lemma 5.1 is complete. □

In the case $H < \frac{1}{2}$, one has a similar result, but the regularity of g_B can be weakened.

LEMMA 5.2. *Let $g_B : [0,1] \to \mathbf{R}^n$ be a continuous function and define g_w as in Lemma* 5.1, *but with $H \in (0, \frac{1}{2})$. Then, for every $\alpha \in (0, H)$, there exists a constant C such that*

$$\|g_w\|_\alpha \le C \sup_{t \in [0,1]} \|g_B(t)\|.$$

PROOF. Since $H < \frac{1}{2}$, one can move the derivative under the integral of the first equation in Lemma 4.2 to get

$$\|g_w(t)\| \le C \int_0^t (t-s)^{-H-1/2} \|g_B(s)\| \, ds \le C \sup_{t \in [0,1]} \|g_B(t)\|.$$

This shows that the restriction of g_w to $[0, 2]$ is square integrable. The large-time tail can be bounded by (5.6) as before. □

We already hinted several times toward the notion of a "cost function" that measures the difficulty of coupling the two copies of the process. This notion is now made precise. Denote by $Z = (x_0, y_0, w_x, w_y)$ an element of $\mathcal{X}^2 \times \mathcal{W}^2$ and assume that there exists a square integrable function $g_w : \mathbf{R}_- \to \mathbf{R}^n$ such that

$$w_y(t) = w_x(t) + \int_t^0 g_w(s) \, ds \qquad \forall t < 0. \tag{5.7}$$

In regard of (4.13), we introduce for $T > 0$ the operator $\mathcal{R}_T$ given by

$$(\mathcal{R}_T g)(t) = C \int_{-\infty}^0 \frac{t^{1/2-H}(T-s)^{H-1/2}}{t+T-s} \|g(s)\| \, ds,$$

where C is the constant appearing in (4.13). The cost is then defined as follows.

DEFINITION 5.3. The cost function $\mathcal{K}_\alpha : \mathsf{L}^2(\mathbf{R}_-) \to [0, \infty]$ is defined by

$$\mathcal{K}_\alpha(g) = \sup_{T>0} \|\mathcal{R}_T g\|_\alpha + C_K \int_{-\infty}^0 (-s)^{H-3/2} \|g(s)\| \, ds, \tag{5.8}$$

where, for convenience, we define $C_K = |(2H-1)\gamma_H \alpha_{1-H}|$. Given Z as above, $\mathcal{K}_\alpha(Z)$ is defined as $\mathcal{K}_\alpha(g_w)$ if there exists a square integrable function g_w satisfying (5.7) and as ∞ otherwise.

REMARK 5.4. The cost function $\mathcal{K}_\alpha$ defined above has the important property that

$$\mathcal{K}_\alpha(\theta_t g) \le \mathcal{K}_\alpha(g) \qquad \text{for all } t \ge 0, \tag{5.9}$$

where the shifted function $\theta_t g$ is given by

$$(\theta_t g)(s) = \begin{cases} g(s+t), & \text{if } s < -t, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, it is a norm, and thus satisfies the triangle inequality.

REMARK 5.5. By (4.13), the first term in (5.8) measures by how much the two realizations of the Wiener process have to differ in order to obtain identical increments for the associated FBMs. By (4.11c), the second term in (5.8) measures by how much the two realizations of the FBM differ if one lets the system evolve with two identical realizations of the Wiener process.

We now turn to the construction of the process (x_t, y_t) during step 1. We will set up our coupling construction in such a way that, whenever step 1 is to be performed, the initial condition Z is admissible in the following sense:

DEFINITION 5.6. Let α satisfy $0 < \alpha < \min\{\frac{1}{2}; H\}$. We say that $Z = (x_0, y_0, w_x, w_y)$ is *admissible* if one has

$$\|x_0 - y_0\| \le 1 + \frac{1 + C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}}, \tag{5.10}$$

[the constants $C_i^{(\mathrm{A1})}$ are as in (A1) and in (5.3)], and its cost satisfies $\mathcal{K}_\alpha(Z) \le 1$.

Denote now by Ω the space of continuous functions $\omega : [0,1] \to \mathbf{R}^n$ which are the restriction to $[0,1]$ of an element of $\widetilde{\mathcal{H}}_H$. Our aim is construct two measures $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ on $\Omega \times \Omega$ satisfying the following conditions:

(B1) The marginals of $\mathbf{P}_Z^1 + \mathbf{P}_Z^2$ onto the two components Ω of the product space are both equal to the Wiener measure W.

(B2) Let $\mathscr{B}_\kappa \subset \Omega \times \Omega$ denote the set of pairs $(\tilde{w}_x, \tilde{w}_y)$ such that there exists a function $g_w : [0,1] \to \mathbf{R}^n$ satisfying

$$\tilde{w}_y(t) = \tilde{w}_x(t) + \int_0^t g_w(s)\, ds, \qquad \int_0^1 \|g_w(s)\|^2\, ds \le \kappa.$$

Then, there exists a value of κ such that, for every admissible initial condition Z_0, we have $\mathbf{P}_Z^1(\mathscr{B}_\kappa) + \mathbf{P}_Z^2(\mathscr{B}_\kappa) = 1$.

(B3) Let (x_t, y_t) be the process constructed by solving (SDE) with respective initial conditions x_0 and y_0, and with respective noise processes $P_t(w_x, \tilde{w}_x)$ and $P_t(w_y, \tilde{w}_y)$. Then, one has $x_1 = y_1$ for $\mathbf{P}_Z^1$-almost every noise $(\tilde{w}_x, \tilde{w}_y)$. Furthermore, there exists a constant $\delta > 0$ such that $\mathbf{P}_Z^1(\Omega \times \Omega) \ge \delta$ for every admissible initial condition Z.

REMARK 5.7. Both measures $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ can easily be extended to measures on $\mathcal{W}_+^2$ in such a way that (B1) holds. Since the dynamic constructed from the coupling function $\mathscr{C}$ will not depend on this extension, we just choose one arbitrarily and denote again by $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ the corresponding measures on $\mathcal{W}_+^2$.

Given $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$, we construct the coupling function $\mathscr{C}$ in the following way, using notation (5.2) and (4.4):

$$T(Z,a) = 1, \qquad \mathsf{W}_2(Z,a) = \mathbf{P}_Z^1 \times \delta_{a_1} + \mathbf{P}_Z^2 \times \delta_{a_2}, \tag{5.11}$$

where the two elements a_1 and a_2 are defined as

$$a_1 = (2, 0, \widetilde{N}, 0), \tag{5.12a}$$

$$a_2 = \bigl(3, 0, \widetilde{N} + 1, t_* \widetilde{N}^{4/(1-2\alpha)}\bigr), \tag{5.12b}$$

for some constant t_* to be determined later in this section. Notice that this definition reflects the algorithm (4.7) and the explanation following (5.2). The reason behind the particular choice of the waiting time in (5.12b) will become clear in Remark 5.11.

The way the construction of $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ works is very close to the binding construction in [11]. The main difference is that the construction presented in [11] does not allow to satisfy (B2) above. We will therefore introduce a symmetrized version of the binding construction that allows to gain a better control over g_w. If μ_1 and μ_2 are two positive measures with densities D_1 and D_2 with respect to some common measure μ, we define the measure $\mu_1 \wedge \mu_2$ by

$$(\mu_1 \wedge \mu_2)(dw) = \min\{D_1(w), D_2(w)\}\mu(dw).$$

The key ingredient for the construction of $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ is the following lemma, the proof of which will be given later in this section.

LEMMA 5.8. *Let $Z = (x_0, y_0, w_x, w_y)$ be an admissible initial condition and let H, σ and f satisfy the hypotheses of either Theorem* 1.2 *or Theorem* 1.3. *Then, there exists a measurable map $\Psi_Z : \Omega \to \Omega$ with measurable inverse, having the following properties.*

(B1′) *There exists a constant $\delta > 0$ such that $\mathsf{W} \wedge \Psi_Z^* \mathsf{W}$ has mass bigger than 2δ for every admissible initial condition Z.*

(B2′) *There exists a constant κ such that $\{(\tilde{w}_x, \tilde{w}_y) | \tilde{w}_y = \Psi_Z(\tilde{w}_x)\} \subset \mathscr{B}_\kappa$ for every admissible initial condition Z.*

(B3′) *Let (x_t, y_t) be the process constructed by solving* (SDE) *with respective initial conditions x_0 and y_0, and with noise processes $P_t(w_x, \tilde{w}_x)$ and $P_t(w_y, \Psi_Z(\tilde{w}_x))$. Then, one has $x_1 = y_1$ for every $\tilde{w}_x \in \Omega$ and every admissible initial condition Z.*

Furthermore, the maps Ψ_Z and Ψ_Z^{-1} are measurable with respect to Z.

Given such a Ψ_Z, we first define the maps $\Psi_\uparrow$ and $\Psi_\to$ from Ω to $\Omega \times \Omega$ by

$$\Psi_\uparrow(\tilde{w}_x) = \bigl(\tilde{w}_x, \Psi_Z(\tilde{w}_x)\bigr), \qquad \Psi_\to(\tilde{w}_y) = \bigl(\Psi_Z^{-1}(\tilde{w}_y), \tilde{w}_y\bigr).$$

(See also Figure 1.) We also define the "switch map" $S : \Omega \times \Omega \to \Omega \times \Omega$ by $S(\tilde{w}_x, \tilde{w}_y) = (\tilde{w}_y, \tilde{w}_x)$.

With these definitions at hand, we construct two measures $\mathbf{P}_Z^1$ and $\widetilde{\mathbf{P}}_Z^1$ on $\Omega \times \Omega$ by

$$\mathbf{P}_Z^1 = \tfrac{1}{2}(\Psi_\uparrow^* \mathsf{W} \wedge \Psi_\to^* \mathsf{W}), \qquad \widetilde{\mathbf{P}}_Z^1 = \mathbf{P}_Z^1 + S^* \mathbf{P}_Z^1. \tag{5.13}$$

In Figure 1, $\mathbf{P}_Z^1$ lives on the boldfaced curve and $\widetilde{\mathbf{P}}_Z^1$ is its symmetrized version which lives on both the boldfaced and the dashed curve. Denote by

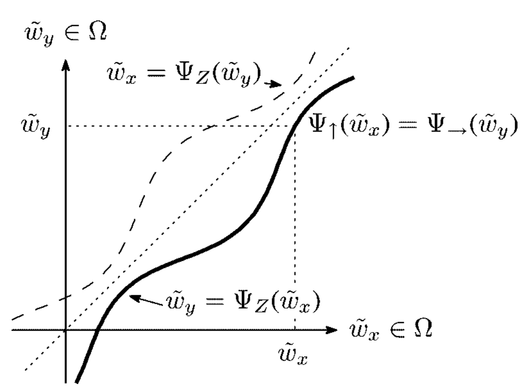


FIG. 1. *Construction of* $\mathbf{P}_Z$.

$\Pi_i : \Omega \times \Omega \to \Omega$ the projectors onto the ith component and by $\Delta : \Omega \to \Omega \times \Omega$ the lift onto the diagonal $\Delta(w) = (w, w)$. Then, we define the measure $\mathbf{P}_Z^2$ by

$$\mathbf{P}_Z^2 = S^* \mathbf{P}_Z^1 + \Delta^*(\mathsf{W} - \Pi_1^* \widetilde{\mathbf{P}}_Z^1). \tag{5.14}$$

By (5.13), $\mathsf{W} > \Pi_1 \widetilde{\mathbf{P}}_Z^1$, so $\mathbf{P}_Z^1$ and $\mathbf{P}_Z^2$ are both positive and their sum is a probability measure. Furthermore, one has by definition

$$\mathbf{P}_Z^1 + \mathbf{P}_Z^2 = \widetilde{\mathbf{P}}_Z^1 + \Delta^*(\mathsf{W} - \Pi_1^* \widetilde{\mathbf{P}}_Z^1).$$

Since $\Pi_1^* \Delta^*$ is the identity, this immediately implies

$$\Pi_1^* \mathbf{P}_Z^1 + \Pi_1^* \mathbf{P}_Z^2 = \mathsf{W}.$$

The symmetry $S^* \widetilde{\mathbf{P}}_Z^1 = \widetilde{\mathbf{P}}_Z^1$ then implies that the second marginal is also equal to W, that is, (B1) is satisfied. Furthermore, the set $\{(\tilde{w}_x, \tilde{w}_y) | \tilde{w}_y = \Psi_Z(\tilde{w}_x)\}$ has $\mathbf{P}_Z$-measure bigger than δ by (B1′), so (B3) is satisfied as well. Finally, (B2) is an immediate consequence of (B2′). It remains to construct the function Ψ_Z.

PROOF OF LEMMA 5.8. As previously, we write Z as

$$Z = (x_0, y_0, w_x, w_y). \tag{5.15}$$

In order to construct Ψ_Z, we proceed as in [11], Section 5, except that we want the solutions x_t and y_t to become equal after time 1. Let $\tilde{w}_x \in \Omega$ be given and define

$$B_H(t) = \big(\mathcal{D}_H P_1(w_x, \tilde{w}_x)\big)(t-1), \tag{5.16}$$

where W denotes the corresponding part of the initial condition Z_0 in (5.15). We write the solutions to (SDE) as

$$dx_t = f(x_t)\, dt + \sigma\, dB_H(t), \tag{5.17a}$$

$$dy_t = f(y_t)\, dt + \sigma\, dB_H(t) + \sigma \tilde{g}_B(t)\, dt, \tag{5.17b}$$

where $\tilde{g}_B(t)$ is a function to be determined. Notice that x_t is completely determined by $\tilde{w}_x$ and by the initial condition Z. We introduce the process $\rho_t = y_t - x_t$, so we get

$$\frac{d\rho_t}{dt} = f(x_t + \rho_t) - f(x_t) + \sigma \tilde{g}_B(t). \tag{5.18}$$

We now define $\tilde{g}_B(t)$ by

$$\tilde{g}_B(t) = -\sigma^{-1}\left(\kappa_1 \rho_t + \kappa_2 \frac{\rho_t}{\sqrt{\|\rho_t\|}}\right), \tag{5.19}$$

for two constants κ_1 and κ_2 to be specified. This yields for the norm of ρ_t the estimate

$$\frac{d\|\rho_t\|^2}{dt} \le 2\big(C_3^{(\mathrm{A1})} - \kappa_1\big)\|\rho_t\|^2 - 2\kappa_2 \|\rho_t\|^{3/2}.$$

We choose $\kappa_1 = C_3^{(\mathrm{A1})}$ and so

$$\|\rho_t\| \le \begin{cases} \big(6\kappa_2 t - \sqrt{\|\rho_0\|}\big)^2, & \text{for } t < \sqrt{\|\rho_0\|}/(6\kappa_2), \\ 0, & \text{for } t \ge \sqrt{\|\rho_0\|}/(6\kappa_2). \end{cases} \tag{5.20}$$

We can then choose κ_2 sufficiently large, so that $\|\rho_t\| = 0$ for $t > 1/2$. Since the initial condition was admissible by assumption, the constant κ_2 can be chosen as a function of the constants $C_i^{(\mathrm{A1})}$ only. Notice also that the preceding construction yields $\tilde{g}_B$ as a function of Z and $\tilde{w}_x$ only.

We then construct $\tilde{w}_y = \Psi_Z(\tilde{w}_x)$ in such a way that (5.17) is satisfied with the function $\tilde{g}_B$ we just constructed. Define g_w by (5.7) and construct g_B by applying (4.11b). Then, we extend $\tilde{g}_B$ to $(-\infty, 1]$ by simply putting it equal to g_B on $(-\infty, 0]$. Applying the inverse formula (4.11a), we obtain a function $\tilde{g}_w$ on $(-\infty, 1]$, which is equal to g_w on $(-\infty, 0]$ and which is such that

$$\big(\Psi_Z(\tilde{w}_x)\big)(t) \equiv \tilde{w}_x(t) + \int_0^t \tilde{g}_w(s)\, ds$$

has precisely the required property.

It remains to check that the family of maps Ψ_Z constructed this way has the properties stated in Lemma 5.8. The inverse of Ψ_Z is constructed in the following way. Choose $\tilde{w}_y \in \Omega$ and consider the solution to the equation

$$dy_t = f(y_t)\, dt + \sigma\, dB'_H(t),$$

where B_H is defined as in (5.16) with x replaced by y. Once y_t is obtained, one can construct the process ρ_t as before, but this time by solving

$$\frac{d\rho_t}{dt} = f(y_t) - f(y_t - \rho_t) - \left(\kappa_1 \rho_t + \kappa_2 \frac{\rho_t}{\sqrt{\|\rho_t\|}}\right).$$

This allows to define $\tilde{g}_B$ as in (5.19). The element $\tilde{w}_x \equiv \Psi_Z^{-1}(\tilde{w}_y)$ is then obtained by the same procedure as before.

Before turning to the proof of properties (B1′)–(B3′), we give some estimate on the function $\tilde{g}_w$ that we just constructed.

LEMMA 5.9. *Assume that the conditions of Lemma* 5.8 *hold. Then, there exists a constant K such that the function $\tilde{g}_w(Z, \tilde{w}_x)$ constructed above satisfies*

$$\int_0^1 \|\tilde{g}_w(Z, \tilde{w}_x)(s)\|^2\, ds < K,$$

for every admissible initial condition Z and for every $\tilde{w}_x \in \mathcal{W}_+$.

PROOF. We write $\tilde{g}_w(t)$ for $t > 0$ as

$$\begin{aligned} \tilde{g}_w(t) &= C \int_{-\infty}^0 \frac{t^{1/2-H}(-s)^{H-1/2}}{t-s} g_w(s)\, ds + \alpha_H \frac{d}{dt} \int_0^t (t-s)^{1/2-H} \tilde{g}_B(s)\, ds \\ &\equiv \tilde{g}_w^{(1)}(t) + \tilde{g}_w^{(2)}(t), \end{aligned}$$

where g_w is defined by (5.7), g_B is given by (5.19) and the constant C is the constant appearing in (4.13). The L^2-norm of $\tilde{g}_w^{(1)}$ is bounded by 1 by the assumption that Z is admissible. To bound the norm of $\tilde{g}_w^{(2)}$, we treat the cases $H < \frac{1}{2}$ and $H > \frac{1}{2}$ separately.

The case $H < \frac{1}{2}$. For this case, we simply combine Lemma 5.2 with the definition (5.19) and the estimate (5.20).

The case $H > \frac{1}{2}$. For this case, we apply Lemma 5.1, so we bound the $\|\cdot\|_{1,1}$-norm of $\tilde{g}_B$. By (5.19), one has

$$\left\|\frac{d}{dt}\tilde{g}_B(t)\right\| \le C\left\|\frac{d\rho_t}{dt}\right\|(1+\|\rho_t\|^{-1/2}), \tag{5.21}$$

for some positive constant C. Using (5.18), the assumption about the boundedness of the derivative of f and the definition (5.19), we get

$$\left\|\frac{d\rho_t}{dt}\right\| \le C\big(\|\rho_t\| + \sqrt{\|\rho_t\|}\big).$$

Combining this with (5.21) and (5.20), the required bound on $\|\tilde{g}_B\|_{1,1}$ follows. $\square$

Property (B1′) now follows from Lemma 5.9 and Girsanov's theorem in the following way. Denote by $\mathscr{D}_Z$ the density of $\Psi_Z^*\mathsf{W}$ with respect to W, that is, $(\Psi_Z^*\mathsf{W})(d\tilde{w}_x) = \mathscr{D}_Z(\tilde{w}_x)\mathsf{W}(d\tilde{w}_x)$. It is given by Girsanov's formula

$$\mathscr{D}_Z(\tilde{w}_x) = \exp\biggl(\int_0^1 \bigl\langle\bigl(\tilde{g}_w(Z,\tilde{w}_x)\bigr)(t), d\tilde{w}_x(t)\bigr\rangle - \tfrac{1}{2}\int_0^1 \|\tilde{g}_w(Z,\tilde{w}_x)\|^2(t)\,dt\biggr).$$

One can check (see, e.g., [20]) that $\|\mathsf{W}\wedge\Psi_Z^*\mathsf{W}\|_{\mathrm{TV}}$ is bounded from below by

$$\|\mathsf{W}\wedge\Psi_Z^*\mathsf{W}\|_{\mathrm{TV}} \ge \biggl(4\int_\Omega \mathscr{D}_Z(w)^{-2}\mathsf{W}(dw)\biggr).$$

Property (B1′) thus follows immediately from Lemma 5.9, using the fact that

$$\int_\Omega \exp\biggl(-2\int_0^1 \bigl\langle\bigl(\tilde{g}_w(Z,\tilde{w}_x)\bigr)(t), d\tilde{w}_x(t)\bigr\rangle - 2\int_0^1 \|\tilde{g}_w(Z,\tilde{w}_x)\|^2(t)\,dt\biggr)\mathsf{W}(dw) = 1.$$

Property (B2′) is also an immediate consequence of Lemma 5.9, and property (B3′) follows by construction from (5.20). The proof of Lemma 5.8 is complete. $\square$

Before concluding this section we show that, if step 1 fails, t_* can be chosen in such a way that the waiting time $t_*\widetilde{N}^{4/(1-2\alpha)}$ in (5.12b) is long enough so that (5.10) holds again after step 3 and so that the cost function does not increase by more than $1/(2\widetilde{N}^2)$. By the triangle inequality, the second claim follows if we show that

$$\mathcal{K}_\alpha\bigl(\theta_t\tilde{g}_w(Z,\tilde{w}_x)\bigr) \le \frac{1}{2\widetilde{N}^2}, \tag{5.22}$$

whenever t is large enough [the shift θ_t is as in (5.9)]. Combining (4.14), Lemma 5.9 and the definition of $\mathcal{K}_\alpha$, we get, for some constant C,

$$\mathcal{K}_\alpha\big(\theta_t \tilde{g}_w(Z, \tilde{w}_x)\big) \le Ct^{\alpha-1/2} + Ct^{H-3/2} \qquad \text{for } t \ge 2.$$

There thus exists a constant t_* such that the bound (5.22) is satisfied if the waiting time is longer than $t_* \widetilde{N}^{4/(1-2\alpha)}$. It remains to show that (5.10) holds after the waiting time is over. If step 1 failed, the realizations $\tilde{w}_x$ and $\tilde{w}_y$ are drawn either in the set

$$\tilde{\Delta}_1 = \{(\tilde{w}_x, \tilde{w}_y) \in \Omega^2 | \tilde{w}_x = \tilde{w}_y\}$$

or in the set

$$\tilde{\Delta}_2 = \{(\tilde{w}_x, \tilde{w}_y) \in \Omega^2 | \tilde{w}_x = \Psi_Z(\tilde{w}_y)\}$$

(see Figure 1). In order to describe the dynamics also during the waiting time (i.e., step 3), we extend those sets to $\mathcal{W}_+^2$ by

$$\Delta_i = \big\{(\tilde{w}_x, \tilde{w}_y) \in \mathcal{W}_+^2 | \big(\tilde{w}_x|_{[0,1]}, \tilde{w}_y|_{[0,1]}\big) \in \tilde{\Delta}_i \\ \text{and } \tilde{w}_x(t) - \tilde{w}_y(t) = \text{const for } t > 1\big\}.$$

Given an admissible initial condition $Z = (x_0, y_0, w_x, w_y)$ and a pair $(\tilde{w}_x, \tilde{w}_y) \in \mathcal{W}_+^2$, we consider the solutions x_t and y_t to (SDE) given by

$$\begin{aligned} dx_t &= f(x_t)\,dt + \sigma\,dB_H^x(t), \\ dy_t &= f(y_t)\,dt + \sigma\,dB_H^y(t), \end{aligned} \tag{5.23}$$

where B_H^x (and similarly for B_H^y) is constructed as usual by concatenating w_x and $\tilde{w}_x$ and applying the operator $\mathcal{D}_H$. The key observation is the following lemma.

LEMMA 5.10. *Let Z be an admissible initial condition as above, let $(\tilde{w}_x, \tilde{w}_y) \in \Delta_1 \cup \Delta_2$, and let x_t and y_t be given by* (5.23) *for $t > 0$. Then, there exists a constant $t_* > 0$ such that*

$$\|x_t - y_t\| \le 1 + \frac{1 + C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}}$$

holds again for $t > t_*$.

PROOF. Fix an admissible initial condition Z and consider the case when $(\tilde{w}_x, \tilde{w}_y) \in \Delta_2$ first. Let $g_w : \mathbf{R}_- \to \mathbf{R}^n$ be as in (5.7) and define $\tilde{g}_w : \mathbf{R}_+ \to \mathbf{R}^n$ by

$$\tilde{w}_y(t) = \tilde{w}_x(t) + \int_0^t \tilde{g}_w(s)\,ds.$$

Introducing $\rho_t = y_t - x_t$, we see that it satisfies the equation

$$\frac{d\rho_t}{dt} = f(y_t) - f(x_t) + \sigma \mathcal{G}_t, \tag{5.24}$$

where the function $\mathcal{G}_t$ is given by

$$\mathcal{G}_t = c_1 \int_{-\infty}^0 (t-s)^{H-3/2} g_w(s)\, ds + c_2 \frac{d}{dt} \int_0^t (t-s)^{H-1/2} \tilde{g}_w(s)\, ds, \tag{5.25}$$

with some constants c_1 and c_2 depending only on H. It follows from (5.24), (5.3) and Gronwall's lemma that the Euclidean norm $\|\rho_t\|$ satisfies the inequality

$$\|\rho_t\| \le e^{-C_2^{(\mathrm{A1})} t} \|\rho_0\| + \int_0^t e^{-C_2^{(\mathrm{A1})}(t-s)} \bigl(C_4^{(\mathrm{A1})} + \|\mathcal{G}_s\|\bigr)\, ds. \tag{5.26}$$

Consider first the time interval [0, 1] and define

$$\widetilde{\mathcal{G}}_t = c_1 \int_{-\infty}^0 (t-s)^{H-3/2} g_w(s)\, ds - c_2 \frac{d}{dt} \int_0^t (t-s)^{H-1/2} \tilde{g}_w(s)\, ds;$$

that is, we simply reversed the sign of $\tilde{g}_w$. This corresponds to the case where $(\tilde{w}_x, \tilde{w}_y)$ are interchanged, and thus satisfy $\tilde{w}_y = \Psi_Z(\tilde{w}_x)$ instead of $\tilde{w}_x = \Psi_Z(\tilde{w}_y)$. We thus deduce from (5.19) and (5.20) that

$$\|\widetilde{\mathcal{G}}_s\| \le \|\sigma^{-1}\| \bigl(\kappa_1 \|\rho_0\| + \kappa_2 \sqrt{\|\rho_0\|}\,\bigr), \tag{5.27}$$

for $s \in [0, 1]$. This yields for $\|\mathcal{G}_s\|$ the estimate

$$\begin{aligned} \|\mathcal{G}_s\| &\le \|\sigma^{-1}\| \bigl(\kappa_1 \|\rho_0\| + \kappa_2 \sqrt{\|\rho_0\|}\,\bigr) + 2c_1 \int_{-\infty}^0 (t-s)^{H-3/2} \|g_w(s)\|\, ds \\ &\le \|\sigma^{-1}\| \bigl(\kappa_1 \|\rho_0\| + \kappa_2 \sqrt{\|\rho_0\|}\,\bigr) + 1, \end{aligned} \tag{5.28}$$

where we used the fact that Z is admissible for the second step. Notice that (5.28) only holds for $s \in [0, 1]$, so we consider now the case $s > 1$. In this case, we can write $\mathcal{G}_t$ as

$$\mathcal{G}_t = c_1 \int_{-\infty}^0 (t-s)^{H-3/2} g_w(s)\, ds + c_1 \int_0^1 (t-s)^{H-3/2} \tilde{g}_w(s)\, ds.$$

The first term is bounded by 1 as before. In order to bound the second term, we use Lemma 5.9, so we get

$$\|\mathcal{G}_t\| \le 1 + \sqrt{\frac{K}{2H-2} \bigl((t-1)^{2H-2} - t^{2H-2}\bigr)}. \tag{5.29}$$

This function has a singularity at $t = 1$, but this singularity is always integrable. For $t > 2$, say, it behaves like $t^{H-3/2}$. Putting the estimates (5.28) and (5.29) into (5.26), we see that there exists a constant C depending only on H and on the parameters in assumption (A1) such that, for $t > 2$, one has the estimate

$$\|\rho_t\| \le e^{-C_2^{(\mathrm{A1})} t} \|\rho_0\| + \frac{1 + C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}} + C t^{H-3/2}.$$

The claim follows at once. $\square$

REMARK 5.11. To summarize, we have shown the following in this section:

1. There exists a positive constant δ such that if the state Z of the coupled system is admissible, step 1 has a probability larger than δ to succeed.
2. If step 1 fails and the waiting time for step 3 is chosen larger than $t_*\widetilde{N}^{4/(1-2\alpha)}$, then the state of the coupled system is again admissible after the end of step 3, provided the cost $\mathcal{K}_\alpha(Z)$ at the beginning of step 1 was smaller than $1-\frac{1}{2\widetilde{N}^2}$.
3. The increase in the cost given between the beginning of step 1 and the end of step 3 is smaller than $\frac{1}{2\widetilde{N}^2}$.

In the following section, we will define step 2 and so conclude the construction and the analysis of the coupling function $\mathscr{C}$.

5.4. *Coupling stage* ($S=2$). In this section, we construct and analyze the coupling map $\mathscr{C}$ corresponding to step 2. Following (4.7), we construct it in such a way that, with positive probability, the two copies of the process (SDE) are driven with the same noise. In other terms, if $Z=(x_0,y_0,w_x,w_y)$ denotes the state of our coupled system at the beginning of step 2, we construct a measure $\mathbf{P}_Z$ on $\mathcal{W}_+^2$ such that if $(\tilde{w}_x,\tilde{w}_y)$ is drawn according to $\mathbf{P}_Z$, then one has

$$\big(\mathcal{D}_H(w_x\sqcup\tilde{w}_x)\big)(t)=\big(\mathcal{D}_H(w_y\sqcup\tilde{w}_y)\big)(t),\qquad t>0, \tag{5.30}$$

with positive probability. Here, $\sqcup$ denotes the concatenation operator given by

$$(w\sqcup\tilde{w})(t)=\begin{cases} w(t), & \text{for } t<0,\\ \tilde{w}(t), & \text{for } t\geq 0.\end{cases}$$

In the notation (5.2), step 2 will have a duration 2^N and N will be incremented by 1 every time step 2 succeeds.

The construction of $\mathbf{P}_Z$ will be similar in spirit to the construction of the previous section. We therefore introduce as before the function $\tilde{g}_w$ given by

$$\tilde{w}_y(t)=\tilde{w}_x(t)+\int_0^t \tilde{g}_w(s)\,ds. \tag{5.31}$$

Our main concern is of course to get good bounds on this function $\tilde{g}_w$. This is achieved by the following lemma, which is crucial in the process of showing that step 2 will eventually succeed infinitely often.

LEMMA 5.12. *Let Z_0 be an admissible initial condition and denote by $\mathcal{T}$ the measure on $\mathcal{X}^2\times\mathcal{W}^2$ obtained by evolving Z_0 according to the successful realization of step* 1. *Then, there exists a constant $\widetilde{K}>0$ depending only on H, α and the parameters appearing in* (A1), *such that for $\mathcal{T}$-almost every $Z=(x,y,w_x,w_y)$, and for every pair $(\tilde{w}_x,\tilde{w}_y)$ satisfying* (5.30), *we have the bounds*

$$\|\tilde{g}_w\|_\alpha\leq\widetilde{K},\qquad \left\|\frac{d\tilde{g}_w}{dt}\right\|_{\alpha+1}\leq\widetilde{K}. \tag{5.32}$$

Furthermore, one has $x=y$, $\mathcal{T}$-almost surely.

PROOF. It is clear from Lemma 5.8 that $x = y$. Let now Z be an element drawn according to $\mathcal{T}$ and denote by $g_w : \mathbf{R}_- \to \mathbf{R}^n$ the function formally defined by

$$dw_y(t) = dw_x(t) + g_w(t)\,dt. \tag{5.33}$$

We also denote by $g_b : \mathbf{R}_- \to \mathbf{R}^n$ the function such that

$$dB_y(t) = dB_x(t) + g_b(t)\,dt, \tag{5.34}$$

where $B_x = \mathcal{D}_H w_x$ and $B_y = \mathcal{D}_H w_y$. (Note that g_w and g_b are almost surely well defined, so we discard elements Z for which they cannot be defined.) Since Z corresponds almost surely to a successful realization of step 1, g_b is equal on the interval $[-1, 0]$ (up to translation in time) to the function $\tilde{g}_B$ constructed in (5.19). By (5.20), there exists therefore a constant C_g such that

$$\|g_b(s)\| \le \begin{cases} C_g, & \text{for } s \in [-1, -\frac{1}{2}), \\ 0, & \text{for } s \in [-\frac{1}{2}, 0]. \end{cases} \tag{5.35}$$

Combining the linearity of $\mathcal{D}_H$ with (4.13), one can see that if $(\tilde{w}_x, \tilde{w}_y)$ satisfy (5.30), then the function $\tilde{g}_w$ is given by the formula

$$\begin{aligned} \tilde{g}_w(t) = C_1 \int_{-\infty}^{-1} \frac{|t+1|^{1/2-H}|s+1|^{H-1/2}}{t-s} g_w(s)\,ds \\ + C_2 \int_{-1}^{-1/2} (t-s)^{-H-1/2} g_b(s)\,ds, \end{aligned} \tag{5.36}$$

for some constants C_1 and C_2 depending only on H. Notice that the second integral only goes up to $1/2$ because of (5.35).

Since the initial condition Z_0 is admissible by assumption, the $\|\cdot\|_\alpha$-norm of the first term is bounded by 1. The $\|\cdot\|_\alpha$-norm of the second term is also bounded by a constant, using (5.35) and the assumption $\alpha < H$.

Deriving (5.36) with respect to t, we see that there exists a constant K such that

$$\begin{aligned} \left\|\frac{d\tilde{g}_w(t)}{dt}\right\| \le \frac{K}{t+1}\biggl(\int_{-\infty}^{-1} \frac{|t+1|^{1/2-H}|s+1|^{H-1/2}}{t-s} \|g_w(s)\|\,ds \\ + \int_{-1}^{-1/2} (t-s)^{-H-1/2} \|g_b(s)\|\,ds\biggr), \end{aligned} \tag{5.37}$$

and the bound on the derivative follows as previously. $\square$

The definition of our coupling function will be based on the following lemma:

LEMMA 5.13. *Let $\mathcal{N}$ be the normal distribution on $\mathbf{R}$, choose $a \in \mathbf{R}$, $b \ge |a|$, and define $M = \max\{4b, 2\log(8/b)\}$. Then, there exists a measure $\mathcal{N}^2_{a,b}$ on $\mathbf{R}^2$ satisfying the following properties*:

1. *Both marginals of* $\mathcal{N}^2_{a,b}$ *are equal to* $\mathcal{N}$.
2. *If* $|b| \le 1$, *one has*

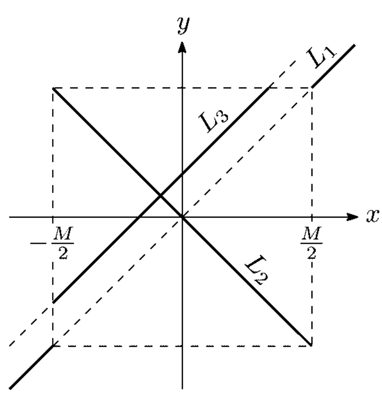
$$\mathcal{N}^2_{a,b}\big(\{(x, y)|y = x + a\}\big) > 1 - b.$$

Furthermore, the above quantity is always positive.

3. *One has*

$$\mathcal{N}^2_{a,b}\big(\{(x, y)||y - x| \le M\}\big) = 1.$$

PROOF. Consider the following figure:



$L_1 : y = x$,
$L_2 : y = -x$,
$L_3 : y = x + a$.

Denote by $\mathcal{N}_x$ the normal distribution on the set $L_x = \{(x, y)|y = 0\}$ and by $\mathcal{N}_y$ the normal distribution on the set $L_y = \{(x, y)|x = 0\}$. We also define the maps $\pi_{i,x}$ (resp. $\pi_{i,y}$) from L_x (resp. L_y) to L_i, obtained by only modifying the y (resp. x) coordinate. Notice that these maps are invertible and denote their inverses by $\tilde{\pi}_{i,x}$ (resp. $\tilde{\pi}_{i,y}$). We also denote by $\mathcal{N}_x|_M$ (resp. $\mathcal{N}_y|_M$) the restriction of $\mathcal{N}_x$ (resp. $\mathcal{N}_y$) to the square $[-\frac{M}{2}, \frac{M}{2}]^2$.

With these notation, we define the measure $\mathcal{N}_3$ on L_3 as

$$\mathcal{N}_3 = \pi^*_{3,x}(\mathcal{N}_x|_M) \wedge \pi^*_{3,y}(\mathcal{N}_y|_M).$$

The measure $\mathcal{N}^2_{a,b}$ is then defined as

$$\mathcal{N}^2_{a,b} = \mathcal{N}_3 + \pi^*_{2,x}\big((\mathcal{N}_x|_M) - \tilde{\pi}^*_{3,x}\mathcal{N}_3\big) + \pi^*_{1,x}\big(\mathcal{N}_x - (\mathcal{N}_x|_M)\big).$$

A straightforward calculation, using the symmetries of the problem, shows that property 1 is indeed satisfied. Property 3 follows immediately from the construction, so it remains to check that property 2 holds, that is, that

$$\mathcal{N}_3(L_3) \ge 1 - b$$

for $|b| < 1$, and $\mathcal{N}_3(L_3) > 0$ otherwise. It follows from the definition of the total variation distance $\|\cdot\|_{\mathrm{TV}}$ that

$$\mathcal{N}_3(L_3) = 1 - \tfrac{1}{2}\|(\mathcal{N}_x|_M) - \tau_a^*(\mathcal{N}_x|_M)\|_{\mathrm{TV}},$$

where $\tau_a(x) = x - a$. Since $M \geq 4b \geq 4a$, it is clear from the figure and from the fact that the density of the normal distribution is everywhere positive, that $\mathcal{N}_3(L_3) > 0$ for every $a \in \mathbf{R}$. It therefore suffices to consider the case $|b| \leq 1$. Since $\int_M^\infty e^{-x^2/2}\,dx < b/8$, one has $\|\mathcal{N}_x|_M - \mathcal{N}_x\|_{\mathrm{TV}} \leq b/4$, which implies

$$\mathcal{N}_3(L_3) \geq 1 - \frac{b}{4} - \frac{1}{2}\|\mathcal{N}_x - \tau_a^*\mathcal{N}_x\|_{\mathrm{TV}}.$$

A straightforward computation shows that, for $|a| \leq 1$,

$$\|\mathcal{N}_x - \tau_a^*\mathcal{N}_x\|_{\mathrm{TV}} \leq \sqrt{e^{a^2} - 1} \leq \sqrt{2}a,$$

and the claim follows. □

We will use the following corollary:

COROLLARY 5.14. Let W be the Wiener measure on $\mathcal{W}_+$, let $g \in \mathsf{L}^2(\mathbf{R}_+)$ with $\|g\| \leq b$, let $M = \max\{4b, 2\log(8/b)\}$, and define the map $\Psi_g : \mathcal{W}_+ \to \mathcal{W}_+$ by

$$(\Psi_g w)(t) = w(t) + \int_0^t g(s)\,ds.$$

Then, there exists a measure $\mathsf{W}^2_{g,b}$ on $\mathcal{W}^2_+$ such that the following properties hold:

1. Both marginals of $\mathsf{W}^2_{g,b}$ are equal to the Wiener measure W.
2. If $b \leq 1$, one has the bound

$$\mathsf{W}^2_{g,b}\big(\{(\tilde{w}_x, \tilde{w}_y) | \tilde{w}_y = \Psi_g(\tilde{w}_x)\}\big) \geq 1 - b. \tag{5.38}$$

   Furthermore, at fixed $b > 0$, the above quantity is always positive and a decreasing function of $\|g\|$.
3. The set

$$\left\{(\tilde{w}_x, \tilde{w}_y) | \exists \kappa : \tilde{w}_y(t) = \tilde{w}_x(t) + \kappa \int_0^t g(s)\,ds, |\kappa|\|g\| \leq M\right\}$$

   has full $\mathsf{W}^2_{g,b}$-measure.

PROOF. This is an immediate consequence of the L^2-expansion of white noise, using g as one of the basis functions and applying Lemma 5.13 on that component. □

Given this result (and using the same notation as above), we turn to the construction of the coupling function $\mathscr{C}$ for step 2. Given an initial condition $Z = (x_0, y_0, w_x, w_y)$, remember that g_w is defined by (5.7). We furthermore define the function $\tilde{g}_w : \mathbf{R}_+ \to \mathbf{R}^n$ by

$$\tilde{g}_w(t) = C \int_{-\infty}^0 \frac{t^{1/2-H}(-s)^{H-1/2}}{t-s} g_w(s)\, ds, \tag{5.39}$$

with C the constant appearing in (4.13). By (4.13), $\tilde{g}_w$ is the only function that ensures that (5.30) holds if $\tilde{w}_x$ and $\tilde{w}_y$ are related by (5.31). [Notice that, although (5.36) seems to differ substantially from (5.39), they do actually define the same function.] Given Z as above and $a \in \mathcal{A}$, denote by $g_{a,Z}$ the restriction of $\tilde{g}_w$ to the interval $[0, 2^N]$ (prolonged by 0 outside). It follows from Lemma 5.12 that there exists a constant K such that if the coupled process was in an admissible state at the beginning of step 1, then the a priori estimate

$$\|g_{a,Z}\|^2 \equiv \int_0^{2^N} \|g_{a,Z}(s)\|^2\, ds \le C 2^{-2\alpha N} \equiv b_N^2 \tag{5.40}$$

holds for some constant C. We thus define $b = \max\{b_N, \|g_{a,Z}\|\}$ and denote by $\mathsf{W}^2_{Z,a}$ the restriction of $\mathsf{W}^2_{g_{a,Z},b}$ to the "good" set (5.38) and by $\widetilde{\mathsf{W}}^2_{Z,a}$ its restriction to the complementary set.

We choose furthermore an arbitrary exponent β satisfying the condition

$$\beta > \frac{1}{1-2\alpha}. \tag{5.41}$$

With these notation at hand, we define the coupling function for step 2:

$$T(Z,a) = 2^N, \qquad \mathsf{W}_2(Z,a) = \mathsf{W}^2_{Z,a} \times \delta_{a'} + \widetilde{\mathsf{W}}^2_{Z,a} \times \delta_{a''},$$

where

$$a' = (2, N+1, \widetilde{N}, 0), \qquad a'' = \big(3, 0, \widetilde{N}+1, \tilde{t}_* 2^{\beta N} \widetilde{N}^{4/(1-2\alpha)}\big), \tag{5.42}$$

for some constant $\tilde{t}_*$ to be determined in the remainder of this section. The waiting time in (5.42) has been chosen in such a way that the following holds.

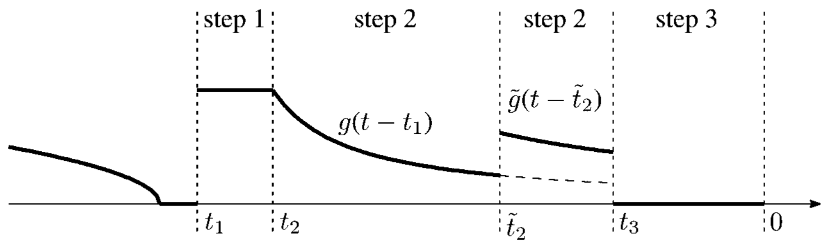
LEMMA 5.15. *Let $(Z_0, a_0) \in \mathcal{X}^2 \times \mathcal{W}^2 \times \mathcal{A}$ with Z_0 admissible and denote by $\mathcal{T}$ the measure on $\mathcal{X}^2 \times \mathcal{W}^2$ obtained by evolving it according to the successful realization of step* 1, *followed by N successful realizations of step* 2, *one failed realization of step* 2 *and one waiting period* 3. *There exists a constant $\tilde{t}_*$ such that $\mathcal{T}$-almost every $Z = (x, y, w_x, w_y)$ satisfies*

$$\|x - y\| \le 1 + \frac{1 + C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}}, \qquad \mathcal{K}_\alpha(Z) \le \mathcal{K}_\alpha(Z_0) + \frac{1}{2\widetilde{N}^2},$$

where $\widetilde{N}$ denotes the value of the corresponding component of a_0.

PROOF. We first show the bound on the cost function. Given Z distributed according to $\mathcal{T}$ as in the statement, we define g_w by (5.33) as usual. The bounds we get on the function g_w are schematically depicted in the following figure, where the time interval $[\tilde{t}_2, t_3]$ corresponds to the failed realization of step 2:

(5.43)



Notice that, except for the contribution coming from times smaller than t_1, we are exactly in the situation of (4.12). Since the cost of a function is decreasing under time shifts, the contribution to $\mathcal{K}_\alpha(Z)$ coming from $(-\infty, t_1]$ is bounded by $\mathcal{K}_\alpha(Z_0)$. Denote by g the function defined by

$$g(t) = \begin{cases} g_w(t+t_1), & \text{for } t \in [0, t_3 - t_1], \\ 0, & \text{otherwise.} \end{cases}$$

Using the definition of the cost function together with Proposition 4.4 and the Cauchy–Schwarz inequality, we obtain for some constants C_1 and C_2 the bound

$$\mathcal{K}_\alpha(Z) \le \mathcal{K}_\alpha(Z_0) + C_1\sqrt{|t_3|^{2H-2} - |t_1|^{2H-2}}\|g\| + C_2\left|\frac{t_1}{t_3 - t_1}\right|^{\alpha-1/2}\|g\|_\alpha,$$

where $\|\cdot\|$ denotes the L^2-norm. Since step 1 has length 1 and the Nth occurrence of step 2 has length 2^{N-1}, we have

$$|t_3 - t_1| = 2^{N+1}, \qquad |t_3| = \tilde{t}_* 2^{\beta N}\widetilde{N}^{4/(1-2\alpha)}.$$

In particular, one has $|t_3| > |t_3 - t_1|$ if $\tilde{t}_*$ is larger than 1. Since

$$\sqrt{|t_3|^{2H-2} - |t_1|^{2H-2}} \le |t_3|^{H-3/2}|t_3 - t_1|^{1/2} \le \left|\frac{t_3}{t_3 - t_1}\right|^{-1/2},$$

this yields (for a different constant C_1) the bound

$$\mathcal{K}_\alpha(Z) \le \mathcal{K}_\alpha(Z_0) + C_1\left|\frac{t_3}{t_3 - t_1}\right|^{\alpha-1/2}\|g\|_\alpha \le \mathcal{K}_\alpha(Z_0) + C_1\frac{\tilde{t}_*^{\alpha-1/2}2^{-\gamma N}}{\widetilde{N}^2}\|g\|_\alpha,$$

where we defined $\gamma = (\beta - 1)(\frac{1}{2} - \alpha)$. Notice that (5.41) guarantees that $\gamma > \alpha$.

We now bound the $\|\cdot\|_\alpha$-norm of g. We know from Lemma 5.12 that the contribution coming from the time interval $[t_1, \tilde{t}_2]$ is bounded by some constant K. Furthermore, by (5.40), we have for the contribution coming from the interval $[\tilde{t}_2, t_3]$ a bound of the type

$$\int_{\tilde{t}_2}^{t_3} \|\tilde{g}(s)\|^2\, ds \le C(N+1)^2,$$

for some positive constant C. This yields for g the bound

$$\|g\|_\alpha \le C(N+1)2^{\alpha N},$$

for some other constant C. Since $\gamma > \alpha$, there exists a constant C such that

$$\mathcal{K}_\alpha(Z) \le \mathcal{K}_\alpha(Z_0) + C\frac{\tilde{t}_*^{\alpha-1/2}}{\tilde{N}^2}.$$

By choosing $\tilde{t}_*$ sufficiently large, this proves the claim concerning the increase of the total cost.

It remains to show that, at the end of step 3, the two realizations of (SDE) did not drift too far apart. Define g_b by (5.34) as usual and notice that, by construction, $x_t = y_t$ for $t = \tilde{t}_2$. Writing as before $\rho_t = y_t - x_t$, one has for $t > \tilde{t}_2$ the estimate

$$\|\rho_t\| \le \frac{C_4^{(\mathrm{A1})}}{C_2^{(\mathrm{A1})}} + \int_{\tilde{t}_2}^t e^{-C_2^{(\mathrm{A1})}(t-s)}\|g_b(s)\|\,ds. \tag{5.44}$$

We first estimate the contribution coming from the time interval $[\tilde{t}_2, t_3]$. Denote by $\tilde{g} : [\tilde{t}_2, t_3] \to \mathbf{R}^n$ the value g_w would have taken, had the last occurrence of step 2 succeeded and not failed [this corresponds to the dashed curve in (5.43)]. Defining $\hat{g} = g_w - \tilde{g}$, we have by (4.11e) that, on the interval $t \in [\tilde{t}_2, t_3]$,

$$g_b(t) = C_1\frac{\hat{g}(\tilde{t}_2)}{(t-\tilde{t}_2)^{1/2-H}} + C_2\int_{\tilde{t}_2}^t \frac{(d\hat{g}/ds)(s)}{(t-s)^{1/2-H}}\,ds. \tag{5.45}$$

By Corollary 5.14 and the construction of the coupling function, $\hat{g}$ is proportional to g_w and, by (5.40), we also have for $\hat{g}$ a bound of the type $\|\hat{g}\| \le C(N+1)$ (the norm is the L^2-norm over the time interval $[\tilde{t}_2, t_3]$). Furthermore, (5.37) yields $\|\frac{d\hat{g}}{ds}\| \le C(N+1)2^{-N}$. Recall that every differentiable function defined on an interval of length L satisfies

$$|f(t)| \le \frac{\|f\|}{\sqrt{L}} + \left\|\frac{df}{dt}\right\|\sqrt{L}.$$

(The norms are L^2-norms.) Using this to bound the first term in (5.45) and the Cauchy–Schwarz inequality for the second term, we get a constant C such that g_b is bounded by

$$\|g_b(t)\| \le C(N+1)\bigl(1 + 2^{-N/2}(t-\tilde{t}_2)^{H-1/2}\bigr).$$

From this and (5.44), we get another constant C such that $\|\rho_t\| \le C(N+1)$ at the time $t = t_3$. We finally turn to the interval $[t_3, 0]$. It follows from (4.11c) that, for some constant C, we have

$$\|g_b(t)\| \le \tfrac{1}{2} + C|t-t_3|^{H-1}\|g\|,$$

where the term $\frac{1}{2}$ is the contribution from the times smaller than t_1. Since we know by (5.40) and Corollary 5.14 that the L^2-norm of g is bounded by $C(N+1)$ for some constant C, we obtain the required estimate by choosing $\tilde{t}_*$ sufficiently large. $\square$

REMARK 5.16. To summarize this section, we have shown the following, assuming that the coupled system was in an admissible state before performing step 1 and that step 1 succeeded:

1. There exist constants $\delta' \in (0,1)$ and $K > 0$ such that the Nth consecutive occurrence of step 2 succeeds with probability larger than $\max\{\delta', 1 - K2^{-\alpha N}\}$. This occurrence has length 2^{N-1}.
2. If the Nth occurrence of step 2 fails and the waiting time for step 3 is chosen longer than $\tilde t_* 2^{\beta N}\tilde N^{4/(1-2\alpha)}$, then the state of the coupled system is again admissible after the end of step 3, provided that the cost $\mathcal{K}_\alpha(Z)$ at the beginning of step 1 was smaller than $1 - \frac{1}{2\tilde N^2}$.
3. The increase in the cost given between the beginning of step 1 and the end of step 3 is smaller than $\frac{1}{2\tilde N^2}$.

Now that the construction of the coupling function $\mathscr{C}$ is completed, we can finally turn to the proof of the results announced in the Introduction.

6. Proof of the main result. Let us first reformulate Theorems 1.2 and 1.3 in a more precise way, using the notation developed in this paper.

THEOREM 6.1. *Let $H \in (0,1) \setminus \{\frac{1}{2}\}$, let f and σ satisfy* (A1)–(A3) *if $H < \frac{1}{2}$ and* (A1), (A2′), (A3) *if $H > \frac{1}{2}$, and let $\gamma < \max_{\alpha<H} \alpha(1-2\alpha)$. Then, the SDS defined in Proposition* 3.11 *has a unique invariant measure μ_*. Furthermore, there exist positive constants C and δ such that, for every generalized initial condition μ, one has*

$$\|\mathscr{Q}\mathcal{Q}_t\mu - \mathscr{Q}\mu_*\|_{\mathrm{TV}} \le 2\mu(\{\|x_0\| > e^{\delta t}\}) + Ct^{-\gamma}. \tag{6.1}$$

PROOF. The existence of μ_* follows from Proposition 3.12 and Lemma 2.20. Furthermore, the assumptions of Proposition 2.18 hold by the invertibility of σ, so the uniqueness of μ_* will follow from (6.1).

Denote by φ the SDS constructed in Proposition 3.11, and consider the self-coupling $\mathscr{Q}(\mu,\mu_*)$ for φ constructed in Section 5. We denote by (x_t, y_t) the canonical process associated to $\mathscr{Q}(\mu,\mu_*)$ and we define a random time $\tilde\tau_\infty$ by

$$\tilde\tau_\infty = \inf\{t > 0 | x_s = y_s \ \forall s \ge t\}.$$

It then follows immediately from (4.2) that

$$\|\mathscr{Q}\mathcal{Q}_t\mu - \mathscr{Q}\mu_*\|_{\mathrm{TV}} \le 2\mathbf{P}(\tilde\tau_\infty > t).$$

Remember that $\mathscr{Q}(\mu,\mu_*)$ was constructed as the marginal of the law of a Markov process with continuous time, living on an augmented phase space $\mathscr{X}$. Since we are only interested in bounds on the random time $\tilde\tau_\infty$ and since we know that

$x_s = y_s$ as long as the coupled system is in the state 2, it suffices to consider the Markov chain (Z_n, τ_n) constructed in (4.8). It is clear that $\tilde{\tau}_\infty$ is then dominated by the random time τ_∞ defined as

$$\tau_\infty = \inf\{\tau_n | S_m = 2 \; \forall m \geq n\},$$

where S_n is the component of Z_n indicating the type of the corresponding step. Our interest therefore only goes to the dynamic of τ_n and S_n. We define the sequence of times $t(n)$ by

$$t(0) = 1, \qquad t(n+1) = \inf\{m > t(n) | S_m = 1\}, \tag{6.2}$$

and the sequence of durations $\Delta\tau_n$ by

$$\Delta\tau_n = \tau_{t(n+1)} - \tau_{t(n)},$$

with the convention $\Delta\tau_n = +\infty$ if $t(n)$ is infinite [i.e., if the set in (6.2) is empty]. Notice that we set $t(0) = 1$ and not 0 because we will treat step 0 of the coupled process separately. The duration $\Delta\tau_n$ therefore measures the time needed by the coupled system starting in step 1 to come back again to step 1. We define the sequence ξ_n by

$$\xi_0 = 0, \qquad \xi_{n+1} = \begin{cases} -\infty, & \text{if } \Delta\tau_n = +\infty, \\ \xi_n + \Delta\tau_n, & \text{otherwise.} \end{cases}$$

By construction, one has

$$\tau_\infty = \tau_1 + \sup_{n \geq 0} \xi_n, \tag{6.3}$$

so we study the tail distribution of the $\Delta\tau_n$.

For the moment, we leave the value α appearing throughout the paper free; we will tune it at the end of the proof. Notice also that, by Remarks 5.11 and 5.16, the cost increases by less than $\frac{1}{2\tilde{N}^2}$ every time the counter $\widetilde{N}$ is increased by 1. Since the initial condition has no cost [by the choice (4.6) of its distribution], this implies that, with probability 1, the system is in an admissible state every time step 1 is performed.

Let us first consider the probability of $\Delta\tau_n$ being infinite. By Remark 5.11, the probability for step 1 to succeed is always greater than δ. After step 1, the Nth occurrence of step 2 has length 2^{N-1}, and a probability greater than $\max\{\delta', 1 - K2^{-\alpha N}\}$ of succeeding. Therefore, one has

$$\mathbf{P}(\Delta\tau_n \geq 2^N) \geq \delta \prod_{k=0}^{N} \max\{\delta', 1 - K2^{-\alpha k}\}.$$

This product always converges, so there exists a constant $p_* > 0$ such that

$$\mathbf{P}(\Delta\tau_n = \infty) \geq p_*,$$

for every $n > 0$. Since our estimates are uniform over all admissible initial conditions and the coupling is chosen in such a way that the system is always in an admissible state at the beginning of step 1, we actually just proved that the conditional probability of $\mathbf{P}(\Delta\tau_n = \infty)$ on any event involving S_m and $\Delta\tau_m$ for $m < n$ is bounded from below by p_*.

For $\Delta\tau_n$ to be finite, there has to be a failure of step 2 at some point [see (4.7)]. Recall that if step 2 succeeds exactly N times, the corresponding value for $\Delta\tau_n$ will be equal to $2^N + \tilde{t}_* 2^{\beta N}(1+n)^{4/(1-2\alpha)}$ for $N > 0$ and to $t_*(1+n)^{4/(1-2\alpha)}$ for $N = 0$. This follows from (5.12b) and (5.42), noticing that $\widetilde{N}$ in those formulae counts the number of times step 1 occurred and is therefore equal to n. We also know that the probability of the Nth occurrence of step 2 to fail is bounded from above by $K2^{-\alpha N}$. Therefore, a very crude estimate yields a constant C such that

$$\mathbf{P}\big((1+n)^{-4/(1-2\alpha)}\Delta\tau_n \geq C2^{\beta N} \text{ and } \Delta\tau_n \neq \infty\big) \leq K \sum_{k>N} 2^{-\alpha k}.$$

This immediately yields for some other constant C

$$\mathbf{P}\big((1+n)^{-4/(1-2\alpha)}\Delta\tau_n \geq T \text{ and } \Delta\tau_n \neq \infty\big) \leq CT^{-\alpha/\beta}. \tag{6.4}$$

As a consequence, the process ξ_n is stochastically dominated by the Markov chain ζ_n defined by

$$\zeta_0 = 0, \qquad \zeta_{n+1} = \begin{cases} -\infty, & \text{with probability } p_*, \\ \zeta_n + (n+1)^{4/(1-2\alpha)} p_n, & \text{with probability } 1 - p_*, \end{cases}$$

where the p_n are positive i.i.d. random variables with tail distribution $CT^{-\alpha/\beta}$, that is,

$$\mathbf{P}(p_n \geq T) = \begin{cases} CT^{-\alpha/\beta}, & \text{if } CT^{-\alpha/\beta} < 1, \\ 1, & \text{otherwise.} \end{cases}$$

With these notation and using the representation (6.3), τ_∞ is bounded by

$$\mathbf{P}(\tau_\infty > t) \leq \mathbf{P}(\tau_1 > t/2) + \mathbf{P}\left(\sum_{n=0}^{n_*} (n+1)^{4/(1-2\alpha)} p_n > t/2\right), \tag{6.5}$$

where n_* is a random variable independent of the p_n and such that

$$\mathbf{P}(n_* = k) = p_*(1-p_*)^k. \tag{6.6}$$

In order to bound the second term in (6.5), it thus suffices to estimate terms of the form $\sum_{n=0}^{k} (n+1)^{4/(1-2\alpha)} p_n$ for fixed values of k. Using the Cauchy–Schwarz inequality, one obtains the existence of positive constants C and N such that

$$\mathbf{P}\left(\sum_{n=0}^{k} (n+1)^{4/(1-2\alpha)} p_n > t/2\right) \leq C(k+1)^N t^{-\alpha/\beta}.$$

Combining this with (6.6) and (6.5) yields, for some other constant C,

$$\mathbf{P}(\tau_\infty > t) \le \mathbf{P}(\tau_1 > t/2) + Ct^{-\alpha/\beta}.$$

By the definition of step 0 (5.4), we get for τ_1:

$$\mathbf{P}(\tau_1 > t/2) \le \mu\bigl(\bigl\{\|x_0\| > e^{C_2^{(A1)}t/2}/2\bigr\}\bigr) + \mu_*\bigl(\bigl\{\|y_0\| > e^{C_2^{(A1)}t/2}/2\bigr\}\bigr).$$

Since, by Proposition 3.12, the invariant measure μ_* has bounded moments, the second term decays exponentially fast. Since $\alpha < \min\{\frac{1}{2}, H\}$ and $\beta > (1-2\alpha)^{-1}$ are arbitrary, one can realize $\gamma = \alpha/\beta$ for γ as in the statement.

This concludes the proof of Theorem 6.1. □

We conclude this paper by discussing several possible extensions of our result. The first two extensions are straightforward and can be obtained by simply rereading the paper carefully and (in the second case) combining its results with the ones obtained in the references. The two other extensions are less obvious and merit further investigation.

6.1. *Noise with multiple scalings.* One can consider the case where the equation is driven by several independent FBMs with different values of the Hurst parameter:

$$dx_t = f(x_t)\,dt + \sum_{i=1}^{m} \sigma_i\,dB^i_{H_i}(t).$$

It can be seen that in this case, the invertibility of σ should be replaced by the condition that the linear operator

$$\sigma = \sigma_1 \oplus \sigma_2 \oplus \cdots \oplus \sigma_m : \mathbf{R}^{mn} \to \mathbf{R}^n$$

has rank n. The condition on the convergence exponent γ then becomes

$$\gamma < \min\{\gamma_1, \dots, \gamma_m\},$$

where $\gamma_i = \max_{\alpha < H_i} \alpha(1-2\alpha)$.

6.2. *Infinite-dimensional case.* In the case where the phase space for (SDE) is infinite-dimensional, the question of global existence of solutions is technically more involved and was tackled in [18]. Another technical difficulty arises from the fact that one might want to take for σ an operator which is not boundedly invertible, so (A3) would fail on a formal level. One expects to be able to overcome this difficulty at least in the case where the equation is semilinear and parabolic, that is, of the type

$$dx = Ax\,dt + F(x)\,dt + Q\,dB_H(t),$$

with the domain of F "larger" (in a sense to be quantified) than the domain of A and B_H a cylindrical FBM on some Hilbert space $\mathcal{H}$ on which the solution is defined, provided the eigenvalues of A and of Q satisfy some compatibility condition as in [2, 6, 8].

On the other hand, it is possible in many cases to split the phase space into a finite number of "unstable modes" and an infinite number of "stable modes" that are slaved to the unstable ones. In this situation, it is sufficient to construct step 1 in such a way that the unstable modes meet, since the stable ones will then automatically converge toward each other. A slight drawback of this method is that the convergence toward the stationary state no longer takes place in the total variation distance. We refer to [11, 14, 20] for implementations of this idea in the Markovian case.

6.3. *Multiplicative noise.* In this case, the problem of existence of global solutions can already be hard. In the case $H > 1/2$, the FBM is sufficiently regular, so one obtains pathwise existence of solutions by rewriting (SDE) in integral form and interpreting the stochastic integral pathwise as a Riemann–Stieltjes integral. In the case $H \in (\frac{1}{4}, \frac{1}{2})$, it has been shown [3, 15, 16] that pathwise solutions can also be obtained by realising the FBM as a geometric rough path. More refined probabilistic estimates are required in the analysis of step 1 of our coupling construction. The equivalent of (5.18) then indeed contains a multiplicative noise term, so the deterministic estimate (5.20) fails.

6.4. *Arbitrary Gaussian noise.* Formally, white noise is a centered Gaussian process ξ with correlation function

$$\mathbf{E}\xi(s)\xi(t) = C_w(t-s) = \delta(t-s).$$

The derivative of the FBM with Hurst parameter H is formally also a centered Gaussian process, but its correlation function is proportional to

$$C_H(t-s) = |t-s|^{2H-2},$$

which should actually be interpreted as the second derivative of $|t-s|^{2H}$ in the sense of distributions.

A natural question is whether the results of the present paper also apply to differential equations driven by Gaussian noise with an arbitrary correlation function $C(t-s)$. There is no conceptual obstruction to the use of the method of proof presented in this paper in that situation, but new estimates are required. It relies on the fact that the driving process is a FBM only to be able to explicitly perform the computations of Section 5.

Acknowledgments. The author wishes to thank Dirk Blömker, David Elworthy, Xue-Mei Li, Neil O'Connell and Roger Tribe for their interest in this work and for many helpful suggestions, stimulating questions and pertinent remarks. He would also like to thank the referee for his careful reading of the manuscript. The author also wishes to thank the Mathematics Research Centre of the University of Warwick for its warm hospitality.

REFERENCES

[1] ARNOLD, L. (1998). *Random Dynamical Systems*. Springer, Berlin.

[2] CERRAI, S. (1999). Smoothing properties of transition semigroups relative to SDEs with values in Banach spaces. *Probab. Theory Related Fields* **113** 85–114.

[3] COUTIN, L. and QIAN, Z. (2002). Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122** 108–140.

[4] CRAUEL, H. (1991). Markov measures for random dynamical systems. *Stochastics Stochastics Rep.* **37** 153–173.

[5] CRAUEL, H. (2002). Invariant measures for random dynamical systems on the circle. *Arch. Math. (Basel)* **78** 145–154.

[6] DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions*. Cambridge Univ. Press.

[7] DUNCAN, T. E., HU, Y. and PASIK-DUNCAN, B. (2000). Stochastic calculus for fractional Brownian motion. I. Theory. *SIAM J. Control Optim.* **38** 582–612 (electronic).

[8] ECKMANN, J.-P. and HAIRER, M. (2001). Uniqueness of the invariant measure for a stochastic PDE driven by degenerate noise. *Comm. Math. Phys.* **219** 523–565.

[9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*. Wiley, New York.

[10] GÄNSSLER, P. and STUTE, W. (1977). *Wahrscheinlichkeitstheorie*. Springer, Berlin.

[11] HAIRER, M. (2002). Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields* **124** 345–380.

[12] JAKŠIĆ, V. and PILLET, C.-A. (1997). Ergodic properties of the non-Markovian Langevin equation. *Lett. Math. Phys.* **41** 49–57.

[13] JAKŠIĆ, V. and PILLET, C.-A. (1998). Ergodic properties of classical dissipative systems. I. *Acta Math.* **181** 245–282.

[14] KUKSIN, S. B. and SHIRIKYAN, A. (2001). A coupling approach to randomly forced nonlinear PDE's. I. *Comm. Math. Phys.* **221** 351–366.

[15] LYONS, T. J. (1994). Differential equations driven by rough signals. I. An extension of an inequality of L. C. Young. *Math. Res. Lett.* **1** 451–464.

[16] LYONS, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14** 215–310.

[17] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

[18] MASLOWSKI, B. and NUALART, D. (2002). Evolution equations driven by a fractional Brownian motion. Preprint.

[19] MASLOWSKI, B. and SCHMALFUSS, B. (2002). Random dynamical systems and stationary solutions of differential equations driven by the fractional Brownian motion. Preprint.

[20] MATTINGLY, J. C. (2002). Exponential convergence for the stochastically forced Navier–Stokes equations and other partially dissipative dynamics. *Comm. Math. Phys.* **230** 421–462.

[21] MEYN, S. P. and TWEEDIE, R. L. (1994). *Markov Chains and Stochastic Stability*. Springer, New York.

[22] ØKSENDAL, B. and HU, Y. (1999). Fractional white noise calculus and applications to finance. Preprint.

[23] RISKEN, H. (1989). *The Fokker–Planck Equation*, 2nd ed. Springer, Berlin.

[24] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, New York.

[25] SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional Integrals and Derivatives*. Gordon and Breach Science, Yverdon.

MATHEMATICS RESEARCH CENTRE
UNIVERSITY OF WARWICK
WARWICK
UNITED KINGDOM
E-MAIL: hairer@maths.warwick.ac.uk
URL: www.maths.warwick.ac.uk/~hairer/