The Annals of Probability
2020, Vol. 48, No. 6, 3041–3076
https://doi.org/10.1214/20-AOP1449

WELL-POSEDNESS, STABILITY AND SENSITIVITIES FOR STOCHASTIC DELAY EQUATIONS: A GENERALIZED COUPLING APPROACH

BY ALEXEI KULIK[1] AND MICHAEL SCHEUTZOW[2]

[1]*Faculty of Pure and Applied Mathematics, Wroclaw University of Science and Technology, kulik.alex.m@gmail.com*
[2]*Institut für Mathematik, Technische Universität Berlin, ms@math.tu-berlin.de*

We develop a new generalized coupling approach to the study of stochastic delay equations with Hölder continuous coefficients, for which analytical PDE-based methods are not available. We prove that such equations possess unique weak solutions, and establish weak ergodic rates for the corresponding segment processes. We also prove, under additional smoothness assumptions on the coefficients, stabilization rates for the sensitivities in the initial value of the corresponding semigroups.

1. Introduction. In this paper, we introduce a new technique which makes it possible to study stochastic equations whose coefficients are assumed to be only *Hölder continuous*, and which does not rely on analytical results from the PDE theory. The analytic approach to the study of diffusion processes dates back to Kolmogorov, and nowadays is a common tool for the analysis of SDEs with low regularity of coefficients; for example, [26]. For stochastic systems of more complicated structure, for example, those described by stochastic equations with delay, this approach is not realistic because of the necessity to study PDEs in (infinite-dimensional) functional spaces. For such systems, the Itô–Lévy stochastic approach is typically used which requires (one-sided local) Lipschitz continuity of the coefficients; for example, [21] or [29]. The current paper shows that the range of application of the standard stochastic analysis tools can be substantially extended, including delay equations with low regularity of the coefficients.

Our approach is based on the concept of *generalized coupling*, which extends the classical notion of *coupling* in the following way. By definition, a coupling is a probability measure on a product space with prescribed marginal distributions. For a *generalized* coupling, the marginals satisfy instead milder deviation bounds from the prescribed distributions. The class of generalized couplings is much wider than of classical couplings, and it is typically much easier to construct for a given system a generalized coupling with desired properties than a true one; for more details, see Section 3 below. This makes generalized couplings quite an efficient tool in the ergodic theory of Markov processes; see the recent paper [6] where they were used as a key ingredient in the construction of contracting/nonexpanding distance-like functions for complicated SPDE models.

In [6], generalized couplings were first constructed using stochastic control arguments, and then used for the construction of true couplings; in this last step, the change of the marginal laws caused by the control terms was in a sense reimbursed. We call this type of argument a *Control-and-Reimburse (C-n-R)* strategy. The same general idea—to apply a stochastic control in order to improve the system, and then to take into account the impact of the control—is scattered in the literature; for example, it is used in [17], Section 5.2, in a construction of of contracting/nonexpanding distance-like function $d(x, y)$ for delay equations, in [15] in an approach to the study of weak ergodicity of SPDEs, in [1] in the proof of ergodicity in total

Received September 2018; revised February 2020.
MSC2020 subject classifications. 34K50, 60H25, 60J25.
Key words and phrases. Stochastic delay equation, generalized coupling, weak solution, weak ergodic rate, sensitivity.

variation for degenerate diffusions, and in [4] in the proof of ergodicity in total variation for solutions to Lévy driven SDEs. Related ideas were used to establish the Harnack inequality for SDEs and SFDEs [13, 30].

We further develop this general idea in the following two directions. First, we show that the C-n-R trategy is well applicable under just Hölder continuity assumptions on the coefficients (actually, one-sided Hölder continuity for the drift). This makes it possible to establish ergodic rates for delay equations with non-Lipschitz coefficients; moreover, essentially the same generalized coupling construction allows one to prove well-posedness of the system, that is, that the weak solution to the equation is uniquely defined and the corresponding *segment process* is a time-homogeneous Markov process with the Feller property. Second, we establish stabilization rates for *sensitivities* for the model; that is, for the derivatives of the semigroup rather than for the semigroup itself. The natural and commonly adopted way to get such rates in a finite-dimensional setting is based on the Bismut–Elworthy–Li-type formulae ([3, 12]) which give integral representations of sensitivities based on the integration-by-parts formulae. Such a *regularization* effect in an infinite-dimensional setting becomes much more structure demanding, since the random noise (which is the source of the integration-by-parts formula) needs to be nondegenerate in the entire space; for one result of such type and a detailed discussion we refer to [8], where reaction-diffusion equations with a cylindrical noise are considered. In the delay case, the noise is finite-dimensional, and thus is strongly degenerate; hence the Bismut–Elworthy–Li-type formula for the (Fréchet) derivatives of the semigroup is hardly available. Nevertheless, employing the C-n-R strategy we are able to derive a family of representation formulae for these derivatives, which can be understood as "poor man's Bismut–Elworthy–Li-type formulae"; see (2.21) and (6.27). Namely, these formulae are not completely free from gradient terms like ∇f, but the weights in the corresponding integral expressions can be forced to decay exponentially fast at an arbitrarily large rate. Using these representation formulae, we manage to establish stabilization rates for sensitivities (derivatives) of arbitrary order; note that the (full) regularization effect now has no reason to appear, and thus for these results we have to assume certain smoothness of the coefficients.

The structure of the paper is the following. In Section 2, the main results are formulated and briefly discussed. To make the exposition transparent, we explain in a separate Section 3 the essence of the method used in all the proofs. The detailed proofs of the three main groups of results are given in Section 4, Section 5 and Section 6, respectively.

2. Main results.

2.1. *Weak solution*: *Existence and uniqueness*. Let $n \in \mathbb{N}$ and $r > 0$. Denote by $\mathbb{C} = C([-r, 0], \mathbb{R}^n)$ the space of continuous functions with the supremum norm $\|\cdot\|$. For a stochastic process $X = \{X(t), t \geq -r\}$ in $\mathbb{R}^n$ define the corresponding *segment process* $\mathbf{X} = \{\mathbf{X}_t, t \geq 0\}$ in $\mathbb{C}$ by

$$\mathbf{X}_t = \{X(t+s), s \in [-r, 0]\} \in \mathbb{C}, \quad t \geq 0.$$

Consider the stochastic delay differential equation (SDDE)

$$\mathrm{d}X(t) = a(\mathbf{X}_t)\,\mathrm{d}t + \sigma(\mathbf{X}_t)\,\mathrm{d}W(t), \quad t \geq 0, \tag{2.1}$$

with the initial condition $\mathbf{X}_0 = x \in \mathbb{C}$. Here, W is a Brownian motion in $\mathbb{R}^m$, $m \geq 1$, and $a : \mathbb{C} \to \mathbb{R}^n$ and $\sigma : \mathbb{C} \to \mathbb{R}^{n\times m}$ are given functions. We will focus on *weak* solutions; that is, processes X with continuous trajectories such that (2.1) holds true with *some* Wiener process W.

Our main assumptions are listed below.

$\mathbf{H}_1$. The function a is continuous, bounded on bounded subsets of $\mathbb{C}$ and satisfies the following *finite range one-sided Hölder conditio*n with index $\alpha > 0$: there exists C such that

$$(2.2) \qquad \big(a(x) - a(y), x(0) - y(0)\big) \le C\|x - y\|^{\alpha+1}, \quad \|x - y\| \le 1.$$

Here and below, we denote the scalar product in $\mathbb{R}^n$ by $(\cdot, \cdot)$. We can and will assume without loss of generality that $\alpha \le 1$.

$\mathbf{H}_2$. The function σ satisfies the following *finite range Hölder condition* with index $\beta > 1/2$:

$$(2.3) \qquad |||\sigma(x) - \sigma(y)||| \le C\|x - y\|^{\beta}, \quad \|x - y\| \le 1.$$

Here and below, $|||\cdot|||$ denotes the Frobenius norm of a matrix, $|||M||| := \sqrt{\sum M_{ij}^2}$. We can and will assume without loss of generality that $\beta \le 1$.

$\mathbf{H}_3$. For each $x \in \mathbb{C}$, there exists a right inverse $\sigma(x)^{-1}$ of the matrix $\sigma(x)$, and

$$(2.4) \qquad \sup_{x\in\mathbb{C}} |||\sigma(x)^{-1}||| < \infty.$$

$\mathbf{H}_4$. The following one-sided linear growth bound for a holds:

$$(2.5) \qquad \big(a(x), x(0)\big) \le C\big(1 + \|x\|^2\big), \quad x \in \mathbb{C}.$$

Note that a similar linear growth bound for σ holds true by (2.3):

$$(2.6) \qquad |||\sigma(x)||| \le C\big(1 + \|x\|\big), \quad x \in \mathbb{C}.$$

THEOREM 2.1. *Assume* $\mathbf{H}_1$–$\mathbf{H}_4$. *Then the following statements hold*:

1. *For any* $x \in \mathbb{C}$, *there exists a weak solution* X *to* (2.1) *with* $\mathbf{X}_0 = x$.
2. *The weak solution to* (2.1) *is unique in law*; *that is*, *any two such solutions with the same initial segment* $x \in \mathbb{C}$ *have the same law in* $C([-r, \infty), \mathbb{R}^n)$.
3. *The segment process* $\mathbf{X}$, *which corresponds to the weak solution to* (2.1), *is a time-homogeneous Markov process in* $\mathbb{C}$, *which has the Feller property.*

The main difficulty in this theorem is the uniqueness statement 2. We note that by a slight modification of the proof one can get the same result assuming a being just continuous and bounded on bounded subsets (i.e., allowing $\alpha = 0$ in $\mathbf{H}_1$). This minor improvement however does not apply to Theorem 2.2 below, and in order to keep the exposition reasonably short we thoroughly explain the one generalized coupling construction which suites well for both these results, and requires $\alpha > 0$.

2.2. *Ergodic rates for the segment process.* Let $d(\cdot, \cdot)$ be a metric on $\mathbb{C}$. The corresponding *coupling* (or *minimal*) distance on the set $\mathcal{P}(\mathbb{C})$ of probability distributions on $\mathbb{C}$ is given by

$$(2.7) \qquad d(\mu, \nu) = \inf_{\lambda\in\mathcal{C}(\mu,\nu)} \int_{\mathbb{C}\times\mathbb{C}} d(x, y)\lambda(\mathrm{d}x, \mathrm{d}y), \quad \mu, \nu \in \mathcal{P}(\mathbb{C}).$$

Here, $\mathcal{C}(\mu, \nu)$ denotes the set of all *couplings* between μ and ν, that is, probability measures on $\mathbb{C} \times \mathbb{C}$ with marginals μ and ν. In what follows, we will consider $d(\cdot, \cdot)$ on $\mathbb{C}$ which generates the same topology as the usual distance $\|\cdot - \cdot\|$ and is bounded. In this case, the corresponding coupling distance is a metric, and convergence in this metric is equivalent to weak convergence in $\mathcal{P}(\mathbb{C})$. The famous Kantorovich–Rubinshtein theorem provides an alternative expression for $d(\mu, \nu)$: denote for $f : \mathbb{C} \to \mathbb{R}$,

$$(2.8) \qquad \|f\|_{\mathrm{Lip}_d} = \sup_{x\neq y} \frac{|f(x) - f(y)|}{d(x, y)},$$

then

$$d(\mu,\nu)=\sup_{f:\|f\|_{\mathrm{Lip}_d}=1}\left|\int f\,\mathrm{d}\mu-\int f\,\mathrm{d}\nu\right|. \tag{2.9}$$

In the literature, $d(\mu,\nu)$ is frequently called the *1-Wasserstein* distance, though the name *Kantorovich* distance is historically more appropriate.

In this section, we will establish weak ergodic rates for the segment process $\mathbf{X}_t, t\ge 0$ with respect to a properly chosen coupling distance $d(\cdot,\cdot)$. That is, we will give sufficient conditions for X to have a unique invariant probability measure (IPM) π and quantitative bounds for the convergence

$$d(P_x^t,\pi)\to 0,\quad t\to\infty;$$

here and below we denote by

$$P_x^t(A)=\mathbb{P}_x(\mathbf{X}_t\in A),\quad A\in\mathcal{B}(\mathbb{C}), x\in\mathbb{C}, t\ge 0$$

the transition probability for the segment process. We adopt the method introduced in [17] and further developed in [5, 11, 19], Chapter 4. The method is based on a proper combination of *contraction*, *nonexpansion*, and *recurrence* properties, which we briefly explain here. Fix a time discretization step $h>0$ and consider the *skeleton chain* $\mathbf{X}^h=\{\mathbf{X}_{kh}, k\in\mathbb{Z}_+\}$ for the segment process $\mathbf{X}$. The distance $d(x,y)$ is called *contracting for* $\mathbf{X}^h$ *on a set* $B\subset\mathbb{C}\times\mathbb{C}$, if there exists $\theta\in(0,1)$ such that

$$d(P_x^h,P_y^h)\le\theta d(x,y),\quad (x,y)\in B. \tag{2.10}$$

The distance $d(x,y)$ is called *nonexpanding for* $\mathbf{X}^h$, if

$$d(P_x^h,P_y^h)\le d(x,y),\quad x,y\in\mathbb{C}. \tag{2.11}$$

With a slight abuse of terminology, we will say that a set $K\subset\mathbb{C}$ is d*-small for* $\mathbf{X}^h$ if d is nonexpanding for $\mathbf{X}^h$ and is contracting on $K\times K$ (this definition differs from the original one [17], Definition 4.4, but has essentially the same scope and is technically more convenient).

The crucial question in the entire approach is how to construct a nonexpanding metric d, which in addition is contracting on a sufficiently large class of sets. The following theorem, which is the main result of this section, resolves this question for the SDDE (2.1). Denote for $x,y\in\mathbb{C}$,

$$d_{N,\gamma}(x,y)=(N\|x-y\|^\gamma)\wedge 1,\quad N\ge 1,\gamma\in(0,1].$$

Clearly, each $d_{N,\gamma}$ is a metric on $\mathbb{C}$.

THEOREM 2.2. I. *Assume* $\mathbf{H}_1$–$\mathbf{H}_4$. *Then for any* $h>r$ *and* $\gamma<\min(\alpha,2\beta-1)$ *there exists* $N_{h,\gamma}$ *such that for any* $N\ge N_{h,\gamma}$ *any bounded set* $K\subset\mathbb{C}$ *is* $d_{N,\gamma}$*-small for* $\mathbf{X}^h$.

II. *Assume in addition that the following stronger version of* $\mathbf{H}_4$ *holds true*:

$$(a(x),x(0))\le C(1+|x(0)|^2),\quad x\in\mathbb{C}, \tag{2.12}$$

$$|||\sigma(x)|||\le C(1+|x(0)|),\quad x\in\mathbb{C}. \tag{2.13}$$

Then for any $h>r$ *and positive* $\gamma<\alpha\wedge(2\beta-1)$ *there exists* $N_{h,\gamma}$ *such that for any* $N\ge N_{h,\gamma}$ *each set*

$$H_c=\{x\in\mathbb{C}:|x(0)|\le c\},\quad c\ge 0$$

is $d_{N,\gamma}$*-small for* $\mathbf{X}^h$.

REMARK 2.1. Conditions (2.12), (2.13) may seem strong because the quantities on the left- hand side depend on the entire trajectory x, while the right-hand side bounds depend on $x(0)$, only. The typical situation where such conditions are satisfied is

$$a(x) = a_0\big(x(0)\big) + a_1(x), \qquad \sigma(x) = \sigma_0\big(x(0)\big) + \sigma_1(x),$$

where a_1, σ_1 are bounded and a_0, σ_0 satisfy linear growth conditions.

Once a proper nonexpanding metric d is constructed, the general theory can be applied which allows one to obtain (weak) ergodic rates, taking into account recurrence properties of the process and measuring how quickly the system visits a d-small set; for example, [19], Section 4.5. Namely, we have the following statement. Denote

$$d_\gamma(x, y) = d_{1,\gamma}(x, y) = \|x - y\|^\gamma \wedge 1, \tag{2.14}$$

and for a given measurable function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ define the functions

$$\Phi(v) = \int_1^v \frac{dw}{\phi(w)}, \qquad r(t) = \phi\big(\Phi^{-1}(t)\big).$$

THEOREM 2.3. *Assume* $\mathbf{H}_1$–$\mathbf{H}_4$. *Assume also that, for some* $h > r$, *the following* Lyapunov-type condition *holds*:

$$\mathbb{E}_x V(\mathbf{X}_h) - V(x) \le -\phi\big(V(x)\big) + C_V, \quad x \in \mathbb{C}. \tag{2.15}$$

Here, $V : \mathbb{C} \to [1, +\infty)$ *is a measurable* Lyapunov function, C_V *is a constant, and the function* $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ *with* $\phi(\infty) = \infty$ *is concave and strictly increasing. Assume that either*

$$V(x) \to \infty, \quad \|x\| \to \infty, \tag{2.16}$$

or

$$V(x) \to \infty, \quad \big|x(0)\big| \to \infty \tag{2.17}$$

and in addition (2.12), (2.13) *hold true.*

Then there exists a unique IPM π *for the segment process* $\mathbf{X}$, *and for any* $\gamma \in (0, 1]$, $\delta \in (0, 1)$ *there exist* $\zeta, C > 0$ *such that*

$$d_\gamma\big(P_x^t, \pi\big) \le \frac{C}{r(\zeta t)^\delta} \phi\big(V(x)\big)^\delta, \quad x \in \mathbb{C}. \tag{2.18}$$

REMARK 2.2. Theorem 2.3 gives a wide set of convergence rates, depending on the function ϕ in the Lyapunov-type condition (2.15). The natural cases include:

(i) (exponential): if $\phi(v) = cv$ with $c > 0$, then $\Phi(v) = \frac{1}{c}\log v$, $r(t) = ce^{ct}$;

(ii) (subexponential): if $\phi(v) = c(v+b)\log^{-\varsigma}(v+b)$ with $c > 0$, $\varsigma > 0$, and $b \ge e^{1+\varsigma} - 1$, then $\phi(v)$ is increasing and concave on $[0, \infty)$ and

$$\Phi(v) = \frac{1}{c(1+\varsigma)}\big[\log^{1+\varsigma}(v+b) - \log^{1+\varsigma}(1+b)\big],$$

$$\Phi^{-1}(t) = \exp\big[\big(\log^{1+\varsigma}(1+b) + c(1+\varsigma)t\big)^{1/(1+\varsigma)}\big] - b,$$

and for any $c_1 < c(1+\varsigma)$ there exists $c_0 > 0$ such that

$$r(t) \ge c_0 e^{c_1 t^{1/(1+\varsigma)}};$$

(iii) (polynomial): if $\phi(v) = ct^{\varsigma}$ with $c > 0$, $\varsigma \in (0, 1)$, then

$$\Phi(v) = \frac{1}{c(1-\varsigma)}\big(v^{1-\varsigma} - 1\big), \qquad \Phi^{-1}(t) = \big(1 + c(1-\varsigma)t\big)^{1/(1-\varsigma)},$$
$$r(t) = c\big(1 + c(1-\varsigma)t\big)^{\varsigma/(1-\varsigma)}.$$

REMARK 2.3. The bound (2.18) can be alternatively considered as a convergence rate for the semigroup

$$P_t f(x) = \mathbb{E}_x f(\mathbf{X}_t), \quad x \in \mathbb{C}, t \ge 0.$$

Namely, denote for $\gamma \in (0, 1]$

$$\|f\|_{H_\gamma} = \sup_{0<\|x-y\|\le 1} \frac{|f(x) - f(y)|}{\|x-y\|^\gamma} + \sup_{\|x-y\|>1} \big|f(x) - f(y)\big|,$$

which is just the Lipschitz constant of f w.r.t. d_γ; see (2.8) and (2.14). Let also H_γ denote the class of functions $f : \mathbb{C} \to \mathbb{R}$ with $\|f\|_{H_\gamma} < \infty$. Then by (2.9) inequality (2.18) is equivalent to the following:

$$\left|P_t f(x) - \int_{\mathbb{C}} f \, d\pi\right| \le \frac{C\phi(V(x))^\delta}{r(\zeta t)^\delta} \|f\|_{H_\gamma}, \quad x \in \mathbb{C}, t \ge 0, f \in H_\gamma. \tag{2.19}$$

In general, it is a separate nontrivial question how to verify the Lyapunov condition (2.15) for delay equations. We do not address this question here, referring to [7] and references therein. Note, however, that there are simple models, where this condition can be checked essentially in the same way as in the (nondelayed) diffusion setting.

PROPOSITION 2.1. *Let the coefficient $\sigma(\cdot)$ be bounded, and the coefficient $a(\cdot)$ satisfy*

$$\big(a(x), x(0)\big) \le -A_\kappa \big|x(0)\big|^{\kappa+1}, \quad \big|x(0)\big| \ge R$$

for some $\kappa \ge -1$, $A_\kappa > 0$, and $R > 0$. Assume also $\mathbf{H}_1$–$\mathbf{H}_3$. Then:

(i) *If $\kappa \ge 0$, then there exist $\alpha > 0$, $c > 0$ such that (2.15) holds true with*

$$V(x) = e^{\alpha|x(0)|}, \qquad \phi(v) = cv.$$

In this case, (2.18) holds true with $r(t) = ce^{ct}$.

(ii) *If $\kappa \in (-1, 0)$, then for any $b > 0$ there exist $\alpha > 0$, $c > 0$ such that (2.15) holds true with*

$$V(x) = e^{\alpha|x(0)|^{\kappa+1}}, \qquad \phi(v) = c(v+b)\log^{2\kappa/(\kappa+1)}(v+b).$$

In this case, (2.18) holds true with $r(t) = c_0 e^{c_1 t^{(1-\kappa)/(1+\kappa)}}$ for any $c_1 < c\frac{1-\kappa}{1+\kappa}$ and some $c_0 > 0$.

(iii) *If $\kappa = -1$ and in addition $2A_{-1} > \Lambda := \sup_x \|\!|\sigma(x)|\!\|^2$, then for $p > 2$,*

$$p < 2 + (2A_{-1} - \Lambda)\Big(\sup_x \|\sigma(x)\|^2\Big)^{-1}$$

there exists $c > 0$ such that condition (2.15) holds true with

$$V(x) = \big|x(0)\big|^p, \qquad \phi(v) = cv^{1-2/p}.$$

In this case, (2.18) holds true with $r(t) = c(1 + \frac{2c}{p}t)^{p/2-1}$.

The proof is analogous to the one of [5], Theorem 3.3; see also [19], Proposition 4.6.1.

Finally, let us mention that, without assuming a Lyapunov function to exist, we still have the following stabilization property: if $\mathbf{H}_1$–$\mathbf{H}_4$ hold and there exists *some* IPM π for the segment process $\mathbf{X}$, then this IPM is *unique* and $P_x^t \to \pi$ weakly as $t \to \infty$ for every x. This follows from [17], Theorem 2.4, and a slightly rearranged argument from the proof of Theorem 2.2; see Remark 5.2 below. Alternatively, one can refer to the approach developed in [18] and [2], based on the notion of *e-processes*; it is easy to see that Remark 5.2 yields the e-process property for $\mathbf{X}_t, t \ge 0$.

2.3. *Sensitivities w.r.t. The initial condition*: *Integral representation and stabilization*. Denote by $C^k(\mathbb{C})$ the class of k times Fréchet differentiable functions $f : \mathbb{C} \to \mathbb{R}$ with continuous derivatives. The kth derivative and directionwise derivatives for $f \in C^k(\mathbb{C})$ will be denoted by $\nabla^k f$ and

$$\nabla^k_{z_1,\dots,z_k} f = \langle \nabla \dots \langle \nabla f, z_1 \rangle, \dots z_k \rangle, \quad z_1, \dots, z_k \in \mathbb{C},$$

respectively. Similarly, the classes $C^k(\mathbb{C}, \mathbb{R}^n)$ and $C^k(\mathbb{C}, \mathbb{R}^{n\times m})$ of the functions valued in $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$ are defined, and the notation for the derivatives is the same. By $C^k_b(\mathbb{C})$, we denote the class of $C^k(\mathbb{C})$ functions, bounded with their derivatives up to order k. For a fixed $k \ge 1$, assume the following.

ASSUMPTION $\mathbf{C}^{(k)}$. $a \in C^k(\mathbb{C}, \mathbb{R}^n)$, $\sigma \in C^k(\mathbb{C}, \mathbb{R}^{n\times m})$, and their derivatives of the orders $1, \dots, k$ are bounded and uniformly continuous on $\mathbb{C}$.

We first consider the case $k = 1$. Define for $\lambda \ge 0, z \in \mathbb{C}$ the process $U^{\lambda,z}$ as the solution to the SDDE

$$\mathrm{d}U^{\lambda,z}(t) = \langle \nabla a(\mathbf{X}_t), \mathbf{U}_t^{\lambda,z} \rangle \mathrm{d}t + \langle \nabla \sigma(\mathbf{X}_t), \mathbf{U}_t^{\lambda,z} \rangle \mathrm{d}W(t) - \lambda U^{\lambda,z}(t)\,\mathrm{d}t \tag{2.20}$$

with the initial condition $\mathbf{U}_0^{\lambda,z} = z$.

THEOREM 2.4. *Let $\mathbf{C}^{(1)}$ and $\mathbf{H}_3$ hold true. Then for any $f \in C^1_b(\mathbb{C})$ the functions $P_t f, t \ge 0$ belong to $C^1_b(\mathbb{C})$. For any $\lambda \ge 0, z \in \mathbb{C}$, the following representation formula holds*:

$$\begin{aligned} \nabla_z \mathbb{E}_x f(\mathbf{X}_t) &= \mathbb{E}_x \langle \nabla f(\mathbf{X}_t), \mathbf{U}_t^{\lambda,z} \rangle \\ &\quad + \lambda \mathbb{E}_x \biggl(f(\mathbf{X}_t) \int_0^t \sigma(\mathbf{X}_s)^{-1} U^{\lambda,z}(s)\,\mathrm{d}W(s) \biggr). \end{aligned} \tag{2.21}$$

Combining the representation formula (2.21) and Theorem 2.3, we get the following stabilization bound for $\nabla P_t f$ as $t \to \infty$. In what follows, we assume that $\phi(0) > 1$, which yields $r(t) > 1, t \ge 0$; this assumption does not restrict generality because one can simultaneously increase ϕ and C_V in (2.15) by 1.

THEOREM 2.5. *Let $\mathbf{C}^{(1)}$ and the assumptions of Theorem 2.3 hold true. Then for any $\gamma \in (0, 1]$, $\delta \in (0, 1)$ and $Q > 0$ there exists $\zeta > 0$ such that the following holds*: *for any $Q > 0$ there exists a constant $C = C_Q > 0$ such that for any $f \in C^1_b(\mathbb{C})$ and $x \in \mathbb{C}, t \ge 0$,*

$$\|\nabla P_t f(x)\| \le \frac{C(\log r(\zeta t) + \phi(V(x)))^\delta}{r(\zeta t)^\delta} \|f\|_{H_\gamma} + Ce^{-Qt} \sup_{y\in\mathbb{C}} \|\nabla f(y)\|. \tag{2.22}$$

REMARK 2.4. The bound (2.22) looks similar to the well-known sufficient condition for the asymptotic strong Feller property; see [16], Proposition 3.12. We do not have a special interest in proving the asymptotic strong Feller property, since it is typically used as a tool for proving unique ergodicity, and we have seen in Section 2.2 that Theorem 2.2 gives an efficient alternative tool for such a proof and, moreover, for getting explicit ergodic rates. In fact, the way Theorem 2.5 is used is quite different from that of [16], Proposition 3.12: instead of using (2.22) to prove ergodicity, we use the ergodic rate (2.19) for the semigroup itself to derive the rate (2.22) for its sensitivities. Our main motivation here comes from the diffusion approximation/homogenization theory for fully coupled systems, where the sensitivity rates appear naturally, for example, [23] for such a theory for diffusions. In [23], such rates were derived using analytic PDE methods, which are not available in the current setting, that is, for delay equations. In further research, we plan to use the sensitivity rates from Theorem 2.5 and Theorem 2.6 below to study the diffusion approximation/homogenization for fully coupled systems with delay.

REMARK 2.5. Note that the first term on the right-hand side of (2.22) coincides with the bound (2.19) up to an extra logarithmic term, which does not affect the structure of the estimate. The derivative ∇f is involved in the second term only, and this term is decaying very rapidly: at exponential rate, and the index Q in this rate can be made arbitrarily large.

Next, let $k > 1$ be arbitrary. For $f \in C^k(\mathbb{C})$ and $j = 1, \dots .k$, for any $x \in C$ one can naturally treat $\nabla^j f(x)$ as a j-linear form on $\mathbb{C}$. We endow the space of such forms by the usual norm

$$\|L\|_j = \sup_{\|z_1\|=\cdots=\|z_j\|=1} |L(z_1, \dots, z_j)|,$$

and denote for $f \in C_b^k(\mathbb{C})$

$$\|f\|_{(k)} = \sup_{x\in\mathbb{C}} \sum_{j=1}^{k} \|\nabla^j f(x)\|_j;$$

note that $\|\cdot\|_{(k)}$ is actually a seminorm because the values of f itself are not involved in it.

THEOREM 2.6. *Let $\mathbf{C}^{(k)}$ hold for some $k > 1$ and let the assumptions of Theorem 2.3 hold true. Then for any $\gamma \in (0, 1]$, $\delta \in (0, 1)$ there exists $\zeta > 0$ such that the following holds: for any $Q > 0$ there exists a constant $C = C_Q > 0$ such that for any $f \in C_b^k(\mathbb{C})$ and $x \in \mathbb{C}, t \geq 0$,*

$$\|\nabla^k P_t f(x)\|_k \leq \frac{C(\log r(\zeta t) + \phi(V(x)))^\delta}{r(\zeta t)^\delta} \|f\|_{H_\gamma} + Ce^{-Qt} \|f\|_{(k)}. \tag{2.23}$$

Note that the structure of the estimate for the higher order derivatives remains exactly the same as for the first-order one: the first term essentially coincides with (2.19) and contains the H_γ-seminorm of f, only, while the second term, which contains the $\|f\|_{(k)}$-seminorm, decays exponentially fast. We mention that there exists an integral representation for the higher order derivatives, analogous to (2.21), see (6.27); actually, the proof of Theorem 2.6 is based on this representation. However, this representation is now less explicit and more cumbersome; that is why we do not formulate it separately here.

3. Outline of the method: Generalized couplings and the control-and-reimburse strategy. Within the classical *coupling* approach to the study of ergodic properties of Markov systems, one has to construct, on a common probability space, a pair of stochastic processes with prescribed law, such that the distance between the components of the pair obeys certain bounds. For instance, inequality (2.10) means that for any $(x, y) \in B$ there exists a pair of segment processes $\mathbf{X}, \mathbf{Y}$ with $\mathrm{Law}(\mathbf{X}) = P_x, \mathrm{Law}(\mathbf{Y}) = P_y$ such that

$$\mathbb{E}d(\mathbf{X}_h, \mathbf{Y}_h) \le \theta d(x, y), \tag{3.1}$$

with some $\theta < 1$. The key question is how to construct a pair $(\mathbf{X}, \mathbf{Y})$ with such a *contraction* property. One natural way is to take $d(x, y) = \|x - y\|$ and to consider the coupling which consists of two solutions to equation (2.1) with the same noise W and given initial conditions x, y. This *synchronous* (or *marching*) coupling is often not a good choice. Namely, assume for the moment that the coefficients a, σ are Lipschitz continuous, then such a pair is well defined, but the contraction property (3.1) in general has no reason to hold true. Namely, to get (3.1) by Itô's formula one has to assume a much stronger version of (2.2):

$$\big(a(x) - a(y), x(0) - y(0)\big) \le -C\|x - y\|^2$$

with a positive constant C sufficiently large when compared with the Lipschitz constant for σ; for example, [19], Section 4.2. Such a *dissipativity* assumption is used quite often for infinite-dimensional SDEs, for example, [9], Chapter 11.5, and [24], Chapter 16.2. However, this is a strong structural limitation which we aim to exclude from the list of assumptions.

Similar obstacles appear if one tries to apply the Itô stochastic calculus tools to get weak uniqueness of solution to (2.1). A natural guess here is that the weak solution to (2.1) should be identifiable as the weak limit as $\varepsilon \to 0$ of the *strong* solutions to the equations

$$\mathrm{d}X^\varepsilon(t) = a^\varepsilon\big(\mathbf{X}_t^\varepsilon\big)\,\mathrm{d}t + \sigma^\varepsilon\big(\mathbf{X}_t^\varepsilon\big)\,\mathrm{d}W(t), \quad t \ge 0, \tag{3.2}$$

where $a^\varepsilon, \sigma^\varepsilon, \varepsilon > 0$ are the families of Lipschitz continuous functions approximating, in a proper sense, the coefficients a, σ. However, for the synchronous coupling of X, X^ε (i.e., the pair of solutions to (2.1), (3.2) with the same W) the estimate for the $\|\cdot\|$-norm of the difference can hardly be derived using Itô's formula unless a, σ are Lipschitz continuous, which is the another assumption we aim to avoid.

In order to overcome these difficulties, we propose a modification of the synchronous coupling construction, which we now explain in detail. In what follows, let $x \in \mathbb{C}$ be given, and X be a weak solution to (2.1) with $\mathbf{X}_0 = x$. Next, let $y \in \mathbb{C}$ and the process Y satisfies the following equation: $\mathbf{Y}_0 = y$,

$$\mathrm{d}Y(t) = \widetilde{a}(\mathbf{Y}_t)\,\mathrm{d}t + \widetilde{\sigma}(\mathbf{Y}_t)\,\mathrm{d}W(t) + \lambda\big(X(t) - Y(t)\big)1_{t \le \tau}\,\mathrm{d}t, \quad t \ge 0, \tag{3.3}$$

where $\widetilde{a}, \widetilde{\sigma}$ are some coefficients and the constant $\lambda > 0$ and the stopping time τ will be determined later. Equation (3.3) should be understood as a *controlled* version of

$$\mathrm{d}\widetilde{Y}(t) = \widetilde{a}(\widetilde{\mathbf{Y}}_t)\,\mathrm{d}t + \widetilde{\sigma}(\widetilde{\mathbf{Y}}_t)\,\mathrm{d}W(t), \quad t \ge 0, \widetilde{\mathbf{Y}}_0 = y. \tag{3.4}$$

The pair $\mathbf{X}, \widetilde{\mathbf{Y}}$ is the synchronous coupling discussed above. The main idea is that, while the distance $\|\mathbf{X}_t - \widetilde{\mathbf{Y}}_t\|$ can hardly be estimated for the synchronous coupling, such an estimate is available for its controlled version under a proper choice of λ, τ in the "control term" $\lambda(X(t) - Y(t))1_{t \le \tau}$. This estimate is the key point in the entire approach, thus we formulate it here. Let $K \subset \mathbb{C}$ be a closed set, and denote

$$\begin{aligned} &\theta_K = \inf\{t : \mathbf{X}_t \notin K\},\\ &\Delta_{a,K} = \sup_{z \in K}\big|a(z) - \widetilde{a}(z)\big|, \qquad \Delta_{\sigma,K} = \sup_{z \in K}\big\|\sigma(z) - \widetilde{\sigma}(z)\big\|,\\ &\upsilon_{x,y,K} = \max\big(\|x - y\|, \Delta_{a,K}^{1/\alpha}, \Delta_{\sigma,K}^{1/\beta}\big) \in [0, \infty]. \end{aligned} \tag{3.5}$$

Fix a positive $\gamma < \min(\alpha, 2\beta - 1)$ and define

$$\lambda_{x,y,K} = \upsilon_{x,y,K}^{\gamma-1}, \tau_{x,y,K} = \inf\{t : |X(t) - Y(t)| > 2\upsilon_{x,y,K}\},$$

with the convention $\infty^{\gamma-1} = 0, 0^{\gamma-1} = 1$.

The following proposition gives a deviation bound between the processes X, Y given by (2.1), (3.3); to make the overall presentation more transparent we postpone its proof to Appendix C.

PROPOSITION 3.1. *Let the coefficients a, σ and $\widetilde{a}, \widetilde{\sigma}$ satisfy* $\mathbf{H}_1$, $\mathbf{H}_2$. *Then for every $T > 0$ there exist $\chi > 0, \upsilon_0 > 0, C_1, C_2 > 0$, depending only on T, γ and the constants in conditions* $\mathbf{H}_1$, $\mathbf{H}_2$, *such that, for an arbitrary closed set $K \subset \mathbb{C}$ and $x \in K, y \in \mathbb{C}$ such that $\upsilon_{x,y,K} \in (0, \upsilon_0]$, for any pair of processes X, Y which satisfy* (2.1) *and* (3.3) *with $\lambda = \lambda_{x,y,K}, \tau = \tau_{x,y,K}$,*

$$\begin{aligned} &\mathbb{P}\Big(\sup_{t \le \theta_K \wedge T} \big(|X(t) - Y(t)|^2 - e^{-\upsilon_{x,y,K}^{\gamma-1} t}\|x - y\|^2\big) \ge \upsilon_{x,y,K}^{2+\chi}\Big) \\ &\quad \le C_1 e^{-C_2 \upsilon_{x,y,K}^{-2\chi}}. \end{aligned} \tag{3.6}$$

We will use Proposition 3.1 in two ways: with $y = x, \widetilde{a} = a^\varepsilon, \widetilde{\sigma} = \sigma^\varepsilon$ to prove weak uniqueness (Theorem 2.1) and $\widetilde{a} = a, \widetilde{\sigma} = \sigma$ to prove the contraction property of the metric $d_{N,\gamma}$ (statement I of Theorem 2.2). Roughly speaking, by adding a control term to the second component of a coupling we become able either to apply the (sort of) Gronwall inequality without the Lipschitz continuity or to guarantee a contraction property without the dissipativity condition.

Clearly, the pair $\mathbf{X}, \mathbf{Y}$ is not a (true) coupling: since equation (3.3) contains an extra control term, the law of $\mathbf{Y}$ has no reason to coincide with that of $\mathbf{Y}$. However, there is still a link between these laws, which is the reason for us to call the pair $\mathbf{X}, \mathbf{Y}$ a *generalized coupling*. In what follows, we denote by d_{TV} the total variation distance between the probability measures, defined by

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|.$$

PROPOSITION 3.2. *Assume that, in addition to the conditions of Proposition* 3.1, *the coefficients $\widetilde{a}, \widetilde{\sigma}$ are Lipschitz continuous and the nondegeneracy assumption* $\mathbf{H}_3$ *for the coefficient $\widetilde{\sigma}$ holds. Then for any T there exists a constant C, depending only on the constant from the nondegeneracy assumption* $\mathbf{H}_3$, *such that*

$$d_{\mathrm{TV}}\big(\mathrm{Law}(Y|_{[0,T]}), \mathrm{Law}(\widetilde{Y}|_{[0,T]})\big) \le C T^{1/2} \upsilon_{x,y,K}^{\gamma}. \tag{3.7}$$

PROOF. We can and will assume that $\upsilon_{x,y,K} < \infty$. Equation (3.3) can be written in the form (3.4) with $\mathrm{d}W(t)$ changed to

$$\mathrm{d}\widetilde{W}(t) = \mathrm{d}W(t) + \eta(t)\,\mathrm{d}t, \quad \eta(t) = \widetilde{\sigma}(\mathbf{Y}_t)^{-1}\lambda_{x,y,K}\big(X(t) - Y(t)\big)1_{t \le \tau_{x,y,K} \wedge T}\,\mathrm{d}t.$$

Note that, by the choice of τ and $\mathbf{H}_3$,

$$\mathbb{E}\int_0^\infty |\eta(t)|^2\,\mathrm{d}t \le C T \upsilon_{x,y,K}^{2\gamma}.$$

Then the law of $\widetilde{W}$ on $C([0,\infty), \mathbb{R}^m)$ is absolutely continuous w.r.t. the law of W and, moreover, the following bound for the total variation distance holds (see Theorem A.1 and (A.1)):

$$d_{\mathrm{TV}}\big(\mathrm{Law}(\widetilde{W}|_{[0,T]}), \mathrm{Law}(W|_{[0,T]})\big) \le C^{1/2} T^{1/2} \upsilon_{x,y,K}^{\gamma}.$$

Since the coefficients $\widetilde{a}, \widetilde{\sigma}$ are Lipschitz continuous, $\mathbf{Y}, \widetilde{\mathbf{Y}}$ are the strong solutions to (3.3), (3.4), respectively, and thus can be understood as images of $\widetilde{W}$, W under a measurable mapping Φ, which gives

$$d_{\mathrm{TV}}\big(\mathrm{Law}(Y|_{[0,T]}), \mathrm{Law}(\widetilde{Y}|_{[0,T]})\big) \leq d_{\mathrm{TV}}\big(\mathrm{Law}(\widetilde{W}|_{[0,T]}), \mathrm{Law}(W|_{[0,T]})\big),$$

and completes the proof. □

Proposition 3.2 shows that the change of the law of the solution, caused by the additional stochastic control term, becomes smaller for smaller deviations $\|x - y\|$, $\Delta_{a,K}$, $\Delta_{\sigma,K}$. This observation enables us to construct a new (true) coupling from the generalized one $\mathbf{X}, \mathbf{Y}$ with the required properties; see [6], Theorem 2.4, and Proposition 5.1 below.

Let us summarize: because the direct construction of a coupling with the required properties may be difficult, we first construct a generalized one. At this stage, using additional control-type terms, the properties of the system can be improved, for example, a contraction-type bound (3.1) can be provided for a nondissipative system. Then we, in a sense, reimburse the changes to the marginal laws, generated by the control-type terms using, for example, the bound (3.7) and constructing a true coupling from the generalized one. This is the essence of the two-stage C-n-R strategy mentioned in the Introduction.

The C-n-R strategy appears to be quite flexible; now we explain how it can be applied to the study of sensitivities. Under the condition $\mathbf{C}^{(1)}$ the solution to (2.1) is *L_p-Fréchet differentiable* w.r.t. $x \in \mathbb{C}$; see Section 6.1 for the corresponding definition and proofs. The respective derivative in the direction $z \in \mathbb{C}$ equals just $U^{0,z}$, which clearly yields (2.21) with $\lambda = 0$. However, in order for the latter identity to provide the stabilization of the sensitivity as $t \to \infty$, it is required that $\mathbf{U}_t^{0,z} \to 0, t \to \infty$. This can be guaranteed under an additional *confluence* assumption, which is an analogue of the dissipativity assumption for the gradient process; see [22] for a systematic treatment of confluent SDEs. Using generalized couplings, we avoid using this strong additional assumption. Namely, together with the *true* derivative $\mathbf{U}_t^{0,z}$ in the direction z, we construct a family of *controlled* derivatives $\mathbf{U}_t^{\lambda,z}$, which are the limits of

$$\frac{\mathbf{Y}_t^{\lambda,x+\varepsilon z} - \mathbf{X}_t^x}{\varepsilon},$$

where $\mathbf{Y}^{\lambda,x+\varepsilon z}$ is defined by a modification of (3.3) with slightly changed control term and the initial value $x + \varepsilon z$. We have for $\lambda > 0$ large enough $\mathbf{U}_t^{\lambda,z} \to 0, t \to \infty$ exponentially fast; that is, using the control-type argument we actually transform a nonconfluent system to a (sort of) confluent one. The "reimbursement" for such a control is represented by the additional integral term on the right-hand side of (2.21), which appears due to the Girsanov formula.

4. Proof of Theorem 2.1.

4.1. *Existence of a weak solution.* Existence of a weak solution can be established in a quite standard way, based on a compactness argument. Both for this purpose and for the subsequent proof of weak uniqueness, we fix families $\{a^\varepsilon\}$, $\{\sigma^\varepsilon\}$ such that:

(i) $a^\varepsilon \to a, \sigma^\varepsilon \to \sigma, \varepsilon \to 0$ uniformly on each compact subset of $\mathbb{C}$;

(ii) conditions $\mathbf{H}_1$–$\mathbf{H}_4$ hold true for $a^\varepsilon, \sigma^\varepsilon$ uniformly in ε; that is, with constants which do not depend on ε.

(iii) the functions $a^\varepsilon, \sigma^\varepsilon$ are Lipschitz continuous on each bounded subset of $\mathbb{C}$.

Note that such a family is easy to construct. Namely, one can consider a family P^ε of finite-dimensional projectors in $\mathbb{C}$ which strongly converge to the identity and such that $P^\varepsilon x(0) = x(0)$, $x \in \mathbb{C}$. Then $\widetilde{a}^\varepsilon(x) = a(P^\varepsilon x)$, $\widetilde{\sigma}^\varepsilon(x) = \sigma(P^\varepsilon x)$ satisfy (i), (ii) and now $\widetilde{a}^\varepsilon, \widetilde{\sigma}^\varepsilon$ are essentially finite-dimensional. Taking convolutions with finite-dimensional approximate δ-functions one obtains the required families $\{a^\varepsilon\}, \{\sigma^\varepsilon\}$.

By property (iii) and (2.5), (2.6) equation (3.2) with the initial condition $\mathbf{X}_0^\varepsilon = x$ has a unique strong solution. By Itô's formula and (2.5), (2.6), for any $p \geq 2$ there exists some C_p such that

$$\mathrm{d}\left|X^\varepsilon(t)\right|^p = \xi^{\varepsilon,p}(t)\,\mathrm{d}t + \eta^{\varepsilon,p}(t)\,\mathrm{d}W(t)$$

with

$$\xi^{\varepsilon,p}(t) \leq C_p\left(1 + \|\mathbf{X}_t^\varepsilon\|^p\right), \qquad \left|\eta^{\varepsilon,p}(t)\right| \leq C_p\left(1 + \|\mathbf{X}_t^\varepsilon\|^p\right).$$

We have

$$\left|X^\varepsilon(t)\right|^p \leq \left|x(0)\right|^p + \int_0^t \left(\xi^{\varepsilon,p}(s)\right)_+ \mathrm{d}s + \int_0^t \eta^{\varepsilon,p}(s)\,\mathrm{d}W(s),$$

and thus

$$\sup_{\tau\in[0,t]} \left|X^\varepsilon(\tau)\right|^p \leq \left|x(0)\right|^p + \int_0^t \left(\xi^{\varepsilon,p}(s)\right)_+ \mathrm{d}s + \sup_{\tau\in[0,t]} \left|\int_0^\tau \eta^{\varepsilon,p}(s)\,\mathrm{d}W(s)\right|.$$

Then by Cauchy's inequality and Doob's inequality,

$$\mathbb{E} \sup_{\tau\in[0,t]} \left|X^\varepsilon(\tau)\right|^{2p} \leq 3\left|x(0)\right|^{2p} + 3C_p^2 \mathbb{E}\left(t + \int_0^t \|\mathbf{X}_s^\varepsilon\|^p\,\mathrm{d}s\right)^2$$
$$+ 12C_p^2 \mathbb{E}\left(t + \int_0^t \|\mathbf{X}_s^\varepsilon\|^{2p}\,\mathrm{d}s\right).$$

Note that

$$\|\mathbf{X}_s^\varepsilon\| \leq \sup_{\tau\in[0,s]} \left|X^\varepsilon(\tau)\right| + \|x\|.$$

Hence, by the Gronwall inequality, we get the bound

$$\sup_{t\leq T,\varepsilon>0} \mathbb{E} \sup_{\tau\in[0,t]} \left|X^\varepsilon(t)\right|^{2p} < \infty, \qquad T > 0, p \geq 2. \tag{4.1}$$

Denote

$$\tau_R^\varepsilon = \inf\left\{t : \left|X^\varepsilon(t)\right| \geq R\right\},$$

then it follows from (4.1) that for any T,

$$\sup_\varepsilon \mathbb{P}\left(\tau_R^\varepsilon < T\right) \to 0, \qquad R \to \infty.$$

Recall that the coefficients $a^\varepsilon, \sigma^\varepsilon$ are bounded (uniformly in ε) on each bounded subset in $\mathbb{C}$. Then it is a standard routine based on the Kolmogorov continuity theorem to show that, for any $\nu < 1/2, q > 0$, and T there exists Q such that

$$\sup_\varepsilon \mathbb{P}\left(X^\varepsilon|_{[0,T]} \notin B_\nu^Q(0,T)\right) \leq q,$$

where, for $-r \leq u \leq v < \infty$,

$$B_\nu^Q(u,v) = \left\{z : \left|z(t)\right| \leq Q, \left|z(t) - z(s)\right| \leq Q|t-s|^\nu, s,t \in [u,v]\right\}$$

is a ball in the space of ν-Hölder continuous functions on $[u,v]$. This yields that the family of laws of $X^\varepsilon, \varepsilon > 0$ in $C(\mathbb{R}^+, \mathbb{R}^n)$ is weakly compact. Since $a^\varepsilon \to a, \sigma^\varepsilon \to \sigma$ uniformly on compacts in $\mathbb{C}$ and a, σ are continuous, any weak limit point for $X^\varepsilon, \varepsilon \to 0$ is a weak solution to (2.1); this argument is again quite standard, and thus we omit the details. This completes the proof of statement 1 of Theorem 2.1.

4.2. *Weak uniqueness*. We will specify the law of an arbitrary weak solution X to (2.1) with $\mathbf{X}_0 = x^0 \in \mathbb{C}$ as the weak limit of the laws of solutions to (3.2). For future reference, we will show that the corresponding weak convergence is uniform w.r.t. x^0 taken from a compact set $K_0 \subset \mathbb{C}$. We take $Q > 0$ (a free parameter, whose value will be specified later) and denote by K_Q the set of $x \in \mathbb{C}$ such that, for some $s \in [0, r]$ and $x^0 \in K_0$,

$$x(t) = x^0(t), \quad t \in [-r, -r+s], x|_{[-r+s,0]} \in B^Q_{1/3}(-r+s, 0).$$

Then K_Q is a compact subset of $\mathbb{C}$, and the calculation from the previous section yields that for any $T > 0$,

$$\mathbb{P}(\mathbf{X}_t \in K_Q, t \in [0, T]) \to 1, \quad Q \to \infty \tag{4.2}$$

uniformly w.r.t. the set of weak solutions to (2.1) with $x_0 \in K_0$.

For the given weak solution X with $\mathbf{X}_0 = x^0 \in \mathbb{C}$, let W be the corresponding Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$. Let $a^\varepsilon, \sigma^\varepsilon$ be the approximating sequence for the coefficients a, σ defined in the previous section. We define Y^ε as the solution to (3.3) with $y = x, \widetilde{a} = a^\varepsilon, \widetilde{\sigma} = \sigma^\varepsilon$. Then, for any $Q > 0$, we have

$$\upsilon^\varepsilon_Q := \max\Big(\sup_{x \in K_Q} \big|a(x) - a^\varepsilon(x)\big|^{1/\alpha}, \sup_{x \in K_Q} \big|\sigma(x) - \sigma^\varepsilon(x)\big|^{1/\beta}\Big) \to 0, \quad \varepsilon \to 0.$$

The case $\upsilon^\varepsilon_Q = 0$ is exceptional, and we have then $X(t) = X^\varepsilon(t), t \le \theta_{K_Q}$ because $a^\varepsilon, \sigma^\varepsilon$ are Lipschitz continuous and coincide with a, σ on K_Q. When $\upsilon^\varepsilon_Q > 0$, we can apply Proposition 3.1 with $y = x = x^0$, $\widetilde{a} = a^\varepsilon, \widetilde{\sigma} = \sigma^\varepsilon$, and since $\upsilon_{x,y,K_Q} = \upsilon^\varepsilon_Q \to 0$ get that for every $\kappa > 0$,

$$\mathbb{P}\Big(\sup_{t \le \theta_{K_Q} \wedge T} \big|X(t) - Y^\varepsilon(t)\big| > \kappa\Big) \to 0, \quad \varepsilon \to 0. \tag{4.3}$$

Moreover, this convergence is uniform w.r.t. any family of weak solutions X to (2.1) with $\mathbf{X}_0 = x^0 \in K_0$.

Now we can complete the proof. Denote by X^ε the solution to (3.4) with $\widetilde{a} = a^\varepsilon, \widetilde{\sigma} = \sigma^\varepsilon$, and $y = x^0$. Let $T > 0$ be fixed and F be a bounded continuous function on $C([0, T], \mathbb{R}^n)$. By Proposition 3.2, we have

$$d_{\mathrm{TV}}\big(\mathrm{Law}(Y^\varepsilon|_{[0,T]}), \mathrm{Law}(X^\varepsilon|_{[0,T]})\big) \to 0, \quad \varepsilon \to 0.$$

Since F is bounded, this gives

$$\mathbb{E}F\big(Y^\varepsilon|_{[0,T]}\big) - \mathbb{E}F\big(X^\varepsilon|_{[0,T]}\big) \to 0, \quad \varepsilon \to 0.$$

On the other hand, it follows from (4.3) that, on the set $\{\theta_Q \ge T\}$,

$$Y^\varepsilon|_{[0,T]} \to X|_{[0,T]}, \quad \varepsilon \to 0$$

in probability in $C([0, T], \mathbb{R}^n)$. Then

$$\limsup_{\varepsilon \to 0} \big|\mathbb{E}F\big(Y^\varepsilon|_{[0,T]}\big) - \mathbb{E}F\big(X|_{[0,T]}\big)\big| \le 2 \sup_x \big|F(x)\big| \mathbb{P}(\theta_Q \ge T).$$

Combining these two inequalities, we get

$$\limsup_{\varepsilon \to 0} \big|\mathbb{E}F\big(X^\varepsilon|_{[0,T]}\big) - \mathbb{E}F\big(X|_{[0,T]}\big)\big| \le 2 \sup_x \big|F(x)\big| \mathbb{P}(\theta_Q \ge T). \tag{4.4}$$

Recall that the choice of Q determines further details in the construction of the generalized coupling, such as the choice of υ_{x,y,K_Q} and subsequent choice of λ, τ. However, (4.4) does not involve Y^ε, and $Q > 0$ therein is just a free parameter. Taking $Q \to \infty$ and using (4.2), we finally deduce that

$$\mathbb{E}F\bigl(X^\varepsilon|_{[0,T]}\bigr) \to \mathbb{E}F\bigl(X|_{[0,T]}\bigr), \quad \varepsilon \to 0. \tag{4.5}$$

This completes the proof of weak uniqueness, since an arbitrary weak solution X^ε to (2.1) is now uniquely specified on any finite time interval $[0, T]$ as the weak limit of the solutions to (3.2). We remark also that the convergence (4.5) is uniform w.r.t. the family of weak solutions X to (2.1) with $\mathbf{X}_0 = x^0 \in K_0$.

4.3. *Continuity and the Markov property.* Denote by $P_{t,x}, t \geq 0, x \in \mathbb{C}$ the law of $\mathbf{X}_t$, where X is the (unique in law) solution to (2.1) with $\mathbf{X}_0 = x$. Denote by $\{P^\varepsilon_{t,x}\}$ the corresponding laws for the approximating sequence X^ε defined by (3.2), and consider the respective families of integral operators

$$T_t f(x) = \int_{\mathbb{C}} f(y) P_{t,x}(\mathrm{d}y), \qquad T^\varepsilon_t f(x) = \int_{\mathbb{C}} f(y) P^\varepsilon_{t,x}(\mathrm{d}y), \quad f \in C_b(\mathbb{C}).$$

We have just proved (see (4.5)) that, for a given $f \in C_b(\mathbb{C})$,

$$T^\varepsilon_t f(x) \to T_t f(x), \quad \varepsilon \to 0$$

uniformly on each compact subset K_0 of $\mathbb{C}$. Then the functions $T_t f, t \geq 0, f \in C_b(\mathbb{C})$ are continuous and bounded.

Now the Markov property for $\mathbf{X}$ is obtained from the same property for $\mathbf{X}^\varepsilon$ by the usual approximation argument: for arbitrary $t > s > s_1, \dots s_k$, $f \in C_b(\mathbb{C})$ and $G \in C_b(\mathbb{C}^{k+1})$, we have

$$\begin{aligned}\mathbb{E}f(\mathbf{X}_t)G(\mathbf{X}_s, \dots, \mathbf{X}_{s_k}) &= \lim_{\varepsilon\to 0} \mathbb{E}f\bigl(\mathbf{X}^\varepsilon_t\bigr)G\bigl(\mathbf{X}^\varepsilon_s, \dots, \mathbf{X}^\varepsilon_{s_k}\bigr) \\ &= \lim_{\varepsilon\to 0} \mathbb{E}T^\varepsilon_{t-s} f\bigl(\mathbf{X}^\varepsilon_s\bigr)G\bigl(\mathbf{X}^\varepsilon_s, \dots, \mathbf{X}^\varepsilon_{s_k}\bigr) \\ &= \mathbb{E}T_{t-s} f(\mathbf{X}_s)G(\mathbf{X}_s, \dots, \mathbf{X}_{s_k});\end{aligned}$$

in the last identity we use that $X^\varepsilon \to X$ weakly and $T^\varepsilon_{t-s} f \to T_{t-s} f$ uniformly on compacts. This proves that $\mathbf{X}$ is a time-homogeneous Markov process with the transition function $\{P_{t,x}(\mathrm{d}y)\}$. The Feller property has already been proved: for $f \in C_b(\mathbb{C})$, the functions $T_t f, t \geq 0$ also belong to $C_b(\mathbb{C})$.

5. Proofs of Theorem 2.2 and Theorem 2.3. Let us give a short outline. We will prove Theorem 2.2 in two steps. First, we will show that for N large enough, $d_{N,\gamma}$ is contracting for $\mathbf{X}^h$ on the set

$$D_{N,\gamma} = \bigl\{(x, y) : d_{N,\gamma}(x, y) < 1\bigr\}. \tag{5.1}$$

Since $d_{N,\gamma}(x, y) \leq 1$ everywhere, this will immediately yield that $d_{N,\gamma}$ is nonexpanding. Then we will prove the following support-type statement: for any given $\delta > 0$ and $h > r$:

$$\inf_{x\in K} \mathbb{P}_x\bigl(\|\mathbf{X}_h\| \leq \delta\bigr) > 0, \tag{5.2}$$

where either K is a bounded closed set, or $K = H_c$ and (2.12), (2.13) hold true. For convenience of the reader, we prove these two principal statements separately in Sections 5.1 and 5.2 below. It follows immediately from (5.2) that $d_{N,\gamma}$ is contracting for $\mathbf{X}^h$ on the set $(K \times K) \setminus D_{N,\gamma}$. Indeed, for any $(x, y) \notin D_{N,\gamma}$, we have $d_{N,\gamma}(x, y) = 1$; on the other hand, taking the independent coupling of X, Y with $\mathbf{X}_0 = x, \mathbf{Y}_0 = y$, we get by (5.2),

$$\begin{aligned}d_{N,\gamma}\bigl(P^h_x, P^h_y\bigr) &\leq \mathbb{E}d_{N,\gamma}(\mathbf{X}_h, \mathbf{Y}_h) \\ &\leq 1 - \frac{1}{2}\biggl(\inf_{z\in K} \mathbb{P}_z\biggl(\|X_h\| \leq \frac{1}{2^{1+1/\gamma}N^{1/\gamma}}\biggr)\biggr)^2 < 1, \quad x, y \in K.\end{aligned}$$

That is, we can get the statement of Theorem 2.2 by combining the contraction property of $d_{N,\gamma}$ on the set $D_{N,\gamma}$ and the support-type statement (5.2). In Section 5.3, we prove Theorem 2.3 as a corollary of Theorem 2.2 and the general theory.

5.1. *Contraction property of $d_{N,\gamma}$ on $D_{N,\gamma}$.* The proof is based on the generalized coupling construction, introduced in Section 4.2. One technical difficulty here is that the coefficients of (2.1) are not Lipschitz continuous. To overcome this minor difficulty, we will systematically use the following trick: first, we make the construction for Lipschitz continuous coefficients; then we provide estimates for the generalized coupling, which involve the constants from the conditions $\mathbf{H}_1$–$\mathbf{H}_4$ only; finally, we remove the additional assumption for the coefficients to be Lipschitz continuous by an approximation argument. With this plan in mind, we assume first a, σ to be Lipschitz continuous and consider the solution X to (2.1) together with the solution Y to (3.3) with same W, coefficients $\widetilde{a} = a, \widetilde{\sigma} = \sigma$, and some $y \in \mathbb{C}$. We take $K = \mathbb{C}$, so that $\theta_K = +\infty$. On the other hand, since $\widetilde{a}, \widetilde{\sigma}$ coincide with a, σ, we have

$$\Delta_{a,K} = \Delta_{\sigma,K} = 0, \qquad \upsilon_{x,y,K} = \|x - y\|,$$

see the notation prior to Proposition 3.1. Then by Proposition 3.1 and Proposition 3.2, for any $T > 0$ there exist $C_1, C_2, C_3 > 0$ and $\chi > 0$, $\upsilon_0 > 0$ such that, for $x \neq y$,

$$\begin{aligned}(5.3)\qquad &\mathbb{P}\Big(\sup_{t\in[0,T]} \big(\big|X(t) - Y(t)\big|^2 - e^{-\|x-y\|^{\gamma-1}t}\|x-y\|^2\big) \geq \|x-y\|^{2+\chi}\Big)\\ &\quad \leq C_1 e^{-C_2\|x-y\|^{-2\chi}}, \quad \|x-y\| \in (0, \upsilon_0],\end{aligned}$$

and

$$(5.4)\qquad d_{\mathrm{TV}}\big(\mathrm{Law}(Y|_{[0,T]}), \mathrm{Law}(\widetilde{Y}|_{[0,T]})\big) \leq C_3\|x-y\|^{\gamma}, \quad x \neq y,$$

where $\widetilde{Y}$ denotes the solution to (2.1) with $\widetilde{\mathbf{Y}}_0 = y$.

Now, let $h > r$ be fixed. The inequality

$$\sup_{t\leq h}\big(\big|Y(t) - X(t)\big|^2 - e^{-\|x-y\|^{\gamma-1}t}\|x-y\|^2\big) \leq \|x-y\|^{2+\chi}$$

yields the bound

$$\|\mathbf{X}_h - \mathbf{Y}_h\| \leq \big(e^{-\|x-y\|^{\gamma-1}(h-r)} + \|x-y\|^{\chi}\big)^{1/2}\|x-y\|.$$

Clearly,

$$e^{-\upsilon^{\gamma-1}(h-r)} + \upsilon^{\chi} \to 0, \quad \upsilon \to 0,$$

and from (5.3) we finally obtain that there exists $\upsilon_1 > 0$ such that

$$(5.5)\qquad \mathbb{P}\bigg(\|\mathbf{X}_h - \mathbf{Y}_h\| \geq \frac{1}{2}\|x-y\|\bigg) \leq C_1\exp\big(-C_2\|x-y\|^{-2\chi}\big), \quad \|x-y\| \leq \upsilon_1.$$

In addition, it follows from (5.4) that

$$(5.6)\qquad d_{\mathrm{TV}}\big(\mathrm{Law}(\mathbf{Y}_h), P_y^h\big) \leq C_3\|x-y\|^{\gamma}, \quad x \neq y.$$

Now it is easy to perform the "reimbursement" step; that is, to derive the required bound for $\mathbb{E}d_{N,\gamma}(\mathbf{X}_h, \widehat{\mathbf{Y}}_h)$, where $X, \widehat{Y}$ is a properly constructed (true) coupling. For the reader's convenience, we formulate this step in a separate proposition, which is a modification of statement (i) of [6], Theorem 2.4.

PROPOSITION 5.1. *Let, for a family $\{\mu^x, x \in \mathbb{C}\} \subset \mathcal{P}(\mathbb{C})$, the families of $\mathbb{C}$-valued random elements $\{\xi^{x,y}, x, y \in \mathbb{C}\}$, $\{\eta^{x,y}, x, y \in \mathbb{C}\}$ be given such that*:

(i) $\mathrm{Law}(\xi^{x,y}) = \mu^x$, $x, y \in \mathbb{C}$, *and for some* $\gamma \in (0, 1]$, $\upsilon > 0$, $C > 0$,

$$d_{\mathrm{TV}}\big(\mathrm{Law}(\eta^{x,y}), \mu^y\big) \le C\|x - y\|^\gamma, \quad \|x - y\| \in (0, \upsilon];$$

(ii) *for some* $\theta \in (0, 1)$, *and a function* $p(s) = o(s^\gamma)$, $s \to 0$, $p(s) \le 1$,

$$\mathbb{P}\big(\|\xi^{x,y} - \eta^{x,y}\| > \theta\|x - y\|\big) \le p\big(\|x - y\|\big), \quad \|x - y\| \in (0, \upsilon].$$

Then for any $\theta_1 \in (\theta^\gamma, 1)$ *there exists* $N_0 = N_0(\gamma, \upsilon, C, \theta, \theta_1, p(\cdot))$ *such that for* $N \ge N_0$

$$d_{N,\gamma}\big(\mu^x, \mu^y\big) \le \theta_1 d_{N,\gamma}(x, y)$$

on the set $\{(x, y) : d_{N,\gamma}(x, y) < 1\}$.

PROOF. Since

$$d_{N,\gamma}\big(\mu^x, \mu^x\big) = d_{N,\gamma}(x, x) = 0,$$

the required bound is trivial for $x = y$. Take $N_1 = \upsilon^{-\gamma}$, then for $N \ge N_1$

$$d_{N,\gamma}(x, y) < 1 \quad \Leftrightarrow \quad N\|x - y\|^\gamma < 1 \quad \Rightarrow \quad \|x - y\| < \upsilon.$$

In what follows, we take $N \ge N_1$ and $x \ne y$ such that $d_{N,\gamma}(x, y) < 1$; then (i) and (ii) hold true.

The following useful fact is well known ([10], Problem 11.8.8, see also [19], Lemma 4.3.2): if (ξ, η) and (ξ', η') are two pairs of random elements valued in a Borel measurable space, such that η and ξ' have the same distribution, then on a properly chosen probability space there exists a triple of random elements ζ_1, ζ_2, ζ_3 such that the law of (ζ_1, ζ_2) coincides with the law of (ξ, η) and the law of (ζ_2, ζ_3) coincides with the law of (ξ', η'). On the other hand, by the assumption (i) and the Coupling lemma (e.g., [27], Section 1.4, or [19], Theorem 2.2.2), on a properly chosen probability space there exists a pair of random elements ξ', η' such that $\mathrm{Law}(\xi') = \mathrm{Law}(\eta^{x,y})$, $\mathrm{Law}(\eta') = \mu^y$, and

$$\mathbb{P}(\xi' \ne \eta') = d_{\mathrm{TV}}\big(\mathrm{Law}(\eta^{x,y}), \mu^y\big) \le C\|x - y\|^\gamma.$$

Take $\xi = \xi^{x,y}$, $\eta = \eta^{x,y}$ and consider the corresponding triple $\zeta_1, \zeta_2, \zeta_3$. Then ζ_1, ζ_3 is a (true) coupling for μ^x, μ^y, and

$$\mathbb{P}\big(\|\zeta_1 - \zeta_2\| \ge \theta\|x - y\|\big) \le p\big(\|x - y\|\big), \qquad \mathbb{P}(\zeta_2 \ne \zeta_3) \le C\|x - y\|^\gamma.$$

Recall that $d_{N,\gamma} \le 1$, hence

$$\begin{aligned} \mathbb{E}d_{N,\gamma}(\zeta_1, \zeta_3) &\le \mathbb{E}d_{N,\gamma}(\zeta_1, \zeta_2) + \mathbb{P}(\zeta_2 \ne \zeta_3) \\ &\le \mathbb{E}d_{N,\gamma}(\zeta_1, \zeta_2)1_{\|\zeta_1 - \zeta_2\| \le \theta\|x-y\|} \\ &\quad + \mathbb{P}\big(\|\zeta_1 - \zeta_2\| \ge \theta\|x - y\|\big) + \mathbb{P}(\zeta_2 \ne \zeta_3). \end{aligned}$$

Recall that $d_{N,\gamma}(x, y) < 1$, hence

$$\mathbb{E}d_{N,\gamma}(\zeta_1, \zeta_2)1_{\|\zeta_1 - \zeta_2\| \le \theta\|x-y\|} \le N\theta^\gamma\|x - y\|^\gamma,$$

and

$$N\|x - y\|^\gamma = d_{N,\gamma}(x, y).$$

Then

$$\begin{aligned} d_{N,\gamma}\big(\mu^x, \mu^y\big) \le \mathbb{E}d_{N,\gamma}(\zeta_1, \zeta_3) &\le N\theta^\gamma\|x - y\|^\gamma + p\big(\|x - y\|\big) + C\|x - y\|^\gamma \\ &\le \left(\theta^\gamma + \frac{1}{N}\|x - y\|^{-\gamma} p\big(\|x - y\|\big) + \frac{C}{N}\right) d_{N,\gamma}(x, y). \end{aligned}$$

Since $p(s) \le 1$ and $s^{-\gamma} p(s) \to 0, s \to 0$, we have that

$$C_p := \sup_{s>0} s^{-\gamma} p(s) < \infty.$$

Define N_2 by the identity

$$\frac{C_p + C}{N_2} = \theta_1 - \theta^\gamma.$$

Then the required statement holds true for $N_0 = \max(N_1, N_2)$. □

Now, we can complete the proof of the contraction property of $d_{N,\gamma}$ for $\mathbf{X}^h$ on $D_{N,\gamma}$. Let $\gamma, \chi, \upsilon_0, C_1, C_2, C_3$ be the same as in (5.5) and (5.6). We apply Proposition 5.1 with $\upsilon = \upsilon_0$, $C = C_3$, $p(s) = (C_1 \exp(-C_2 s^{-2\chi})) \wedge 1$, $\theta = 2^{-1}$, $\theta_1 = 2^{-1}(1 + 2^{-\gamma})$, and obtain that there exists N_0 such that

$$d_{N,\gamma}(P_x^h, P_y^h) \le (2^{-1} + 2^{-1-\gamma}) d_{N,\gamma}(x, y), \quad (x, y) \in D_{N,\gamma}, N \ge N_0, \tag{5.7}$$

which provides the required contraction property under the additional assumption that a, σ are Lipschitz continuous.

The last step in the proof is to remove this limitation; for that, we use an approximation procedure. The choice of the index γ and the constant N_0 in (5.7) is determined only by the assumptions $\mathbf{H}_1$–$\mathbf{H}_3$. Let a family of processes $\{X^\varepsilon\}$ be defined by (3.2) with $a^\varepsilon, \sigma^\varepsilon$ same as in Section 4.1, then the corresponding transition probabilities satisfy a uniform analogue of (5.7): for all $N \ge N_0, \varepsilon > 0$,

$$d_{N,\gamma}(P_x^{h,\varepsilon}, P_y^{h,\varepsilon}) \le (2^{-1} + 2^{-1-\gamma}) d_{N,\gamma}(x, y), \quad (x, y) \in D_{N,\gamma}. \tag{5.8}$$

We have already proved in Section 4.2 that $P_x^{h,\varepsilon} \to P_x^h$ weakly as $\varepsilon \to 0$. Note that $d_{N,\gamma}(x, y)$ is a bounded metric on $\mathbb{C}$, and the convergence in this metric is the same as in the standard one. Hence weak convergence in $\mathcal{P}(\mathbb{C})$ is equivalent to convergence w.r.t. the coupling distance $d_{N,\gamma}$. In particular,

$$d_{N,\gamma}(P_x^{h,\varepsilon}, P_y^{h,\varepsilon}) \to d_{N,\gamma}(P_x^h, P_y^h), \quad \varepsilon \to 0,$$

and thus (5.7) follows from (5.8).

REMARK 5.1. The generalized coupling construction used in the proof above can be also used for a study of the continuous time family $P_x^t, t \ge 0$. Namely, using (3.6) in a similar way as in the proof of Proposition 5.1, we get that there exists a constant C_h such that

$$d_{\gamma,N}(P_x^t, P_y^t) \le C_h d_{\gamma,N}(x, y), \quad x, y \in \mathbb{C}, t \in [0, h]. \tag{5.9}$$

REMARK 5.2. There is another possibility, not used in the previous proof: instead of making the "reimbursement step" at the time segment $[0, h]$, one can iterate the "control" step on the segments $[h, 2h], [2h, 3h], \ldots$. By (5.5), the corresponding pair of processes $\mathbf{X}_t, \mathbf{Y}_t \ge 0$ will satisfy then

$$\mathbb{P}\left(\|\mathbf{X}_{lh} - \mathbf{Y}_{lh}\| \ge \frac{1}{2^l}\|x - y\|\right) \le C_1 \exp(-C_2 2^{2\kappa l} \|x - y\|^{-2\kappa}) \tag{5.10}$$

for all $\|x - y\| \le \upsilon_1, l \ge 1$. This bound combined with the Markov property and (5.6) will give for $\|x - y\| \le \upsilon_1$,

$$\begin{aligned} d_{\mathrm{TV}}(\mathrm{Law}(\mathbf{Y}), P_y) &\le C_3 \|x - y\|^\gamma \\ &\quad + \sum_{l=1}^\infty (C_3 2^{-\gamma l} \|x - y\|^\gamma + C_1 \exp(-C_2 2^{2\kappa l} \|x - y\|^{-2\kappa})), \end{aligned} \tag{5.11}$$

where P_y denotes the law of $\mathbf{X}_t, t \geq 0$ with $\mathbf{X}_0 = y$ in the path space $C([0,\infty), \mathbb{C})$, and $\text{Law}(\mathbf{Y})$ is understood in the same sense. That is, essentially the same construction as in the above proof gives a generalized coupling for the entire path of the segment process. Making now the "reimbursement step" similar to (and simpler than) Proposition 5.1, one can construct a (true) coupling $\mathbf{X}_t, \mathbf{Y}_t, t \geq 0$ for P_x, P_y such that

$$\mathbb{P}\big(\|\mathbf{X}_t - \mathbf{Y}_t\| \not\to 0, t \to \infty\big) \leq C_4\|x-y\|^\gamma.$$

5.2. *Proof of* (5.2). We prove the support-type assertion (5.2) using a stochastic control argument, which is similar to, and simpler than, the one from Section 4.2 and Section 5.1. We consider a family of processes $X^{\lambda,x}$, $\lambda > 0$ defined by

$$\text{(5.12)} \qquad \mathrm{d}X^{\lambda,x}(t) = a\big(\mathbf{X}_t^{\lambda,x}\big)\,\mathrm{d}t + \sigma\big(\mathbf{X}_t^{\lambda,x}\big)\,\mathrm{d}W(t) - \lambda X^{\lambda,x}(t)\,\mathrm{d}t, \qquad \mathbf{X}_0^{\lambda,x} = x.$$

Since we need these processes to be well defined, we assume for a while that a, σ are Lipschitz continuous. We state the following.

PROPOSITION 5.2. *Let $\delta > 0$ and either K be a bounded set, or $K = H_c$ and* (2.12), (2.13) *hold true. Then there exists λ large enough such that*

$$\text{(5.13)} \qquad \inf_{x\in K} \mathbb{P}\big(\big\|\mathbf{X}_h^{\lambda,x}\big\| \leq \delta\big) \geq \frac{1}{2}.$$

Moreover, the choice of λ depends only on δ, the set K, and the constants in conditions (2.5), (2.6) *or* (2.12), (2.13).

The proof of Proposition 5.2 is similar to that of Proposition 3.1; we give both of them in Appendix C.

In addition, by the usual argument based on the Itô formula, the Burkholder–Davis–Gundy inequality and the Gronwall inequality, one has

$$\text{(5.14)} \qquad C_1(\lambda) := \sup_{x\in B} \mathbb{E} \sup_{t\in[0,h]} \big|X^{\lambda,x}(t)\big|^2 < \infty.$$

With these preliminaries, we can proceed with the proof of (5.2). Note that $X^{\lambda,x}$ solves (2.1) with $\mathrm{d}W(t)$ changed to

$$\mathrm{d}W^{\lambda,x}(t) = \mathrm{d}W(t) - \lambda\sigma\big(\mathbf{X}_t^{\lambda,x}\big)^{-1}X^{\lambda,x}(t)\,\mathrm{d}t.$$

Since we have assumed a, σ to be Lipschitz continuous, for each $x \in \mathbb{C}$ there exists a measurable mapping $\Phi_x : C([0,h], \mathbb{R}^m) \to C([0,h], \mathbb{R}^m)$ which resolves (2.1) with the initial condition $\mathbf{X}_0 = x$ up to the time moment h; that is,

$$X|_{[0,h]} = \Phi_x(W|_{[0,h]}).$$

At the same time, we have

$$X^{\lambda,x}|_{[0,h]} = \Phi_x\big(W^{\lambda,x}|_{[0,h]}\big).$$

By Theorem A.1 and (5.14),

$$\text{(5.15)} \qquad \sup_{x\in B} D_{\text{KL}}\big(\text{Law}\,W^{\lambda,x}|_{[0,h]}\|\text{Law}\,W|_{[0,h]}\big) \leq \frac{\lambda h}{2} C_1(\lambda) \sup_{y\in\mathbb{C}} \big\|\sigma(x)^{-1}\big\| =: C_2.$$

Now we apply (A.2) with $\mu = \text{Law}\,W^{\lambda,x}|_{[0,h]}$, $\nu = \text{Law}\,W|_{[0,h]}$, and

$$A = \Big\{w \in C\big([0,h], \mathbb{R}^m\big) : \sup_{t\in[h-r,h]} \big|\Phi_x(w)(t)\big| \leq \delta\Big\}$$

to get for any $N > 1$,

$$\inf_{x\in B} \mathbb{P}_x\big(\|\mathbf{X}_h\| \le \delta\big) \ge \frac{1}{N} \inf_{x\in B} \mathbb{P}\big(\|\mathbf{X}_h^{\lambda,x}\| \le \delta\big) - \frac{C_2 + \log 2}{N \log N}$$
$$\ge \frac{1}{2N} - \frac{C_2 + \log 2}{N \log N}.$$

Taking $N = N_1 = \exp(4C_2 + 4\log 2)$, we get

$$\inf_{x\in K} \mathbb{P}_x\big(\|\mathbf{X}_h\| \le \delta\big) \ge \frac{1}{4N_1}. \tag{5.16}$$

In the above construction λ, C_1, C_2, N_1 depend only on the constants from the assumptions $\mathbf{H}_3, \mathbf{H}_4$ and inequalities (2.5), (2.6), the set K, and the time step h. That is, using the same approximation argument as in the previous section, we can get rid of the additional assumption that a, σ are Lipschitz continuous. This gives (5.16) without any extra assumptions, and completes the proof of (5.2).

5.3. *Proof of Theorem* 2.3. Since $d_{N,\gamma}(x, y)$ decreases as a function of γ, without loss of generality we further assume that γ satisfies the assumption $\gamma < \min(\alpha, 2\beta - 1)$ from Theorem 2.2. Fix ℓ such that

$$\phi(1 + \ell) > 2C_V,$$

where ϕ, C_V are respectively the function and the constant from the Lyapunov condition (2.15). Take $K_{V,\ell} = \{x : V(x) \le \ell\}$, this set is either bounded under (2.16), or is contained in H_c for c large enough if (2.17) holds. Hence by Theorem 2.2 there exists N such that $K_{V,\ell}$ is a $d_{N,\gamma}$-small set; that is, the condition **I** of [19], Theorem 4.5.2, holds true for

$$d = d_{N,\gamma}, \qquad B = K_{V,\ell} \times K_{V,\ell}.$$

On the other hand, by [19], Theorem 2.8.6, the recurrence condition **R** (i), (ii) of [19], Theorem 4.5.2, holds true with $W(x, y) = V(x) + V(y)$, and $\lambda(t) = \Phi^{-1}(t)$. We define

$$\widehat{W}(x, y) = \frac{\phi(V(x)) + \phi(V(y))}{\phi(1)}, \qquad \widehat{\lambda}(t) = \frac{\phi(\Phi^{-1}(t))}{\phi(1)} = \frac{r(t)}{\phi(1)}$$

and observe that the recurrence condition **R** (i), (ii) of [19], Theorem 4.5.2, with these functions and the same B holds true, as well; see the proof of [19], Theorem 2.8.8. In addition, by [19], Proposition 2.8.5, applied to $U = V$, $V = \phi(V)$, we have that

$$\frac{1}{n}\mathbb{E}_x \sum_{k=1}^{n} \widehat{W}(x, \mathbf{X}_{kh}) \le \phi\big(V(x)\big) + C_V + \frac{1}{n}V(x), \quad n \ge 1. \tag{5.17}$$

Now we can obtain the required statement as a direct corollary of [19], Theorem 4.5.2. Namely, take $q = \delta^{-1}$, $p = (1-\delta)^{-1}$, and denote

$$d_{N,\gamma,p}(x, y) = d_{N,\gamma}(x, y)^{1/p}.$$

This function is bounded by 1, hence (5.17) implies the additional assumption (4.5.8) in [19], Theorem 4.5.2:

$$\begin{aligned}\frac{1}{n}\mathbb{E}_x \sum_{k=1}^{n} d_{N,\gamma,p}(x, \mathbf{X}_{kh})\widehat{W}(x, \mathbf{X}_{kh})^{1/q} &\le \left(\frac{1}{n}\mathbb{E}_x \sum_{k=1}^{n} \widehat{W}(x, \mathbf{X}_{kh})\right)^{1/q} \\ &\le \left(\phi\big(V(x)\big) + C_V + \frac{1}{n}V(x)\right)^{1/q}.\end{aligned} \tag{5.18}$$

Then [19], Theorem 4.5.2, yields that there exists unique IPM for $\mathbf{X}$ and for some $c, C > 0$,

$$d_{N,\gamma,p}(P_x^{nh}, \pi) \le \frac{C}{r(cnh)^{1/q}} \widehat{U}(x), \quad x \in \mathbb{C}, n \ge 1 \tag{5.19}$$

with

$$\widehat{U}(x) = \int_{\mathbb{C}} d_{N,\gamma,p}(x, y) \widehat{W}(x, y)^{1/q} \pi(\mathrm{d}y).$$

By Fatou's lemma and (5.18),

$$\widehat{U}(x) \le (\phi(V(x)) + C_V)^\delta \le \phi(V(x))^\delta + C_V^\delta \le C\phi(V(x))^\delta,$$

in the last inequality we have used that $\inf_x \phi(V(x)) > 0$. By the Markov property and (5.9), we have

$$d_{N,\gamma,p}(P_x^{t+nh}, \pi) \le C_h^{1/p} d_{N,\gamma,p}(P_x^{nh}, \pi),$$

hence (5.19) implies

$$d_{N,\gamma,p}(P_x^t, \pi) \le \frac{C}{r(\zeta t)^\delta} \phi(V(x))^\delta, \quad x \in \mathbb{C}, t \ge 0. \tag{5.20}$$

Since

$$d_\gamma(x, y) \le d_{N,\gamma,p}(x, y),$$

this completes the proof of (2.18).

6. Proofs of Theorems 2.4–2.6. The proofs of all these three theorems will be based on the following auxiliary construction. Denote

$$\psi(v) = \arctan(|v|) \frac{v}{|v|}, v \in \mathbb{R}^n.$$

Let $x \in \mathbb{C}$ be arbitrary but fixed. Denote by X^x the (strong) solution to the SDDE (2.1) with $\mathbf{X}_0^x = x$, and consider a family of processes $Y^{\lambda,y}$, $\lambda \ge 0$, $y \in \mathbb{C}$ defined as the solutions to

$$\mathrm{d}Y^{\lambda,y}(t) = a(\mathbf{Y}_t^{\lambda,y})\,\mathrm{d}t + \sigma(\mathbf{Y}_t^{\lambda,y})\,\mathrm{d}W(t) - \lambda\psi(Y^{\lambda,y}(t) - X^x(t))\,\mathrm{d}t, \quad \mathbf{Y}_0^{\lambda,y} = y. \tag{6.1}$$

Note that each $Y^{\lambda,y}$ depends also on x, but we do not indicate this explicitly in order to keep the notation easy to read. Denote

$$\beta^{\lambda,y}(t) = \lambda\sigma(\mathbf{Y}_t^{\lambda,y})^{-1}\psi(Y^{\lambda,y}(t) - X^x(t)),$$

and observe that $|\beta^{\lambda,y}(t)| \le C$ (C may depend on λ). Then the classical Girsanov theorem applies; for example, [20], Chapter 7. Namely, the family

$$\mathcal{E}^{\lambda,y}(t) = \exp\left(\int_0^t \beta^{\lambda,y}(s)\,\mathrm{d}W_s - \frac{1}{2}\int_0^t |\beta^{\lambda,y}(s)|^2\,\mathrm{d}s\right), \quad y \in \mathbb{C}$$

satisfies $\mathbb{E}\mathcal{E}^{\lambda,y}(t) = 1$, and the process

$$W^{\lambda,y}(\tau) = W(\tau) - \int_0^\tau \beta^{\lambda,y}(s) ds, \quad \tau \in [0, t]$$

is a Wiener process on $[0, t]$ w.r.t. the probability measure $\mathcal{E}^{\lambda,y}(t) d\mathbb{P}$. Equation (6.1) is just (2.1) with $\mathrm{d}W$ changed to $\mathrm{d}W^{\lambda,y}$; thus for any bounded measurable f

$$\mathbb{E} f(\mathbf{Y}_t^{\lambda,y}) \mathcal{E}^{\lambda,y}(t) = \mathbb{E} f(\mathbf{X}_t^y). \tag{6.2}$$

In addition, since $\beta^{\lambda,y}$ is bounded by a constant, we have for each $p \ge 1$, $t > 0$,

$$\sup_y \mathbb{E}(\mathcal{E}_t^{\lambda,y})^p < \infty, \qquad \sup_y \mathbb{E}(\mathcal{E}_t^{\lambda,y})^{-p} < \infty. \tag{6.3}$$

6.1. *Proof of Theorem* 2.4. Let $\mathbb{B}$ be a separable Banach space. We will say that a family of $\mathbb{B}$-valued processes $Z^y(t), t \in [0, T], y \in \mathbb{C}$ is *L_p-Fréchet differentiable* at a given point $y \in \mathbb{C}$, if for any $z \in \mathbb{C}$ there exists a family $\nabla_z Z^y(t) \in \mathbb{B}, t \in [0, T], z \in \mathbb{C}$ such that

$$\sup_{t\in[0,T],\|z\|\le 1} \mathbb{E}\left\| \frac{Z^{y+\varepsilon z}(t) - Z^y(t)}{\varepsilon} - \nabla_z Z^y(t) \right\|_{\mathbb{B}}^p \to 0, \quad \varepsilon \to 0, \tag{6.4}$$

the mappings

$$\mathbb{C} \ni z \mapsto \nabla_z Z^y(t) \in L_p(\Omega, \mathbb{P}, \mathbb{B}), \quad t \in [0, T] \tag{6.5}$$

are linear, and

$$\sup_{t\in[0,T]} \sup_{\|z\|\le 1} \mathbb{E}\|\nabla_z Z^y(t)\|_{\mathbb{B}}^p < \infty. \tag{6.6}$$

We have the following.

LEMMA 6.1. *Let $\mathbf{C}^{(1)}$ hold true. Then for any $\lambda \ge 0, T > 0, p \ge 1$ the families*

$$\{Y^{\lambda,y}(t), t \in [0, T], y \in \mathbb{C}\}, \qquad \{\mathbf{Y}_t^{\lambda,y}, t \in [0, T], y \in \mathbb{C}\}$$

of processes taking values in $\mathbb{R}^n$ and $\mathbb{C}$ respectively are L_p-Fréchet differentiable at any $y \in \mathbb{C}$. Moreover, the processes $U^{\lambda,y,z}(t) = \nabla_z Y^{\lambda,y}(t), t \in [0, T], z \in \mathbb{C}$, satisfy

$$\begin{aligned} \mathrm{d}U^{\lambda,y,z}(t) &= \langle \nabla a(\mathbf{Y}_t^{\lambda,y}), \mathbf{U}_t^{\lambda,y,z}\rangle \mathrm{d}t + \langle \nabla\sigma(\mathbf{Y}_t^{\lambda,y}), \mathbf{U}_t^{\lambda,y,z}\rangle \mathrm{d}W(t) \\ &\quad - \lambda\langle \nabla\psi(Y^{\lambda,y}(t) - X^x(t)), U^{\lambda,y,z}(t)\rangle \mathrm{d}t, \quad \mathbf{U}_0^{\lambda,y,z} = z, \end{aligned} \tag{6.7}$$

and

$$\nabla_z \mathbf{Y}_t^{\lambda,y} = \mathbf{U}_t^{\lambda,y,z}, \quad t \in [0, T]. \tag{6.8}$$

PROOF. The scheme of the proof actually repeats the one from the classical proof of L_2-differentiability w.r.t. to a parameter of a solution to an SDE; see [14], Section 2.7. Thus we just briefly outline the usual steps, and focus on particular (minor) difficulties which arise because the state space $\mathbb{C}$ for the solution to (6.1) is infinite dimensional.

Step 1. By assumption $\mathbf{C}^{(1)}$, a, σ are (globally) Lipschitz continuous. Thus, applying first Itô's formula, then the Burkholder–Davis–Gundy inequality, and finally the Gronwall lemma, one gets the bound

$$\sup_{\|z\|\le 1} \mathbb{E} \sup_{t\in[0,T]} |Y^{\lambda,y+\varepsilon z}(t) - Y^{\lambda,y}(t)|^p \le C\varepsilon^p \tag{6.9}$$

for each fixed $p \ge 1, \lambda \ge 0$. The argument here is the same as in Section 4.1, thus we omit the details.

Step 2. Denote $D^{\lambda,y,\varepsilon z}(t) = Y^{\lambda,y+\varepsilon z}(t) - Y^{\lambda,y}(t), t \ge 0$, and let $\mathbf{D}_t^{\lambda,y,\varepsilon z}, t \ge 0$ be the corresponding segment process. Since a, σ are Fréchet differentiable and have bounded and uniformly continuous derivatives, it follows from (6.9) that

$$\begin{aligned} a(\mathbf{Y}_t^{\lambda,y+\varepsilon z}) - a(\mathbf{Y}_t^{\lambda,y}) &= \langle \nabla a(\mathbf{Y}_t^{\lambda,y}), \mathbf{D}_t^{\lambda,y,\varepsilon z}\rangle + R_a^{\lambda,y,\varepsilon z}(t), \\ \sigma(\mathbf{Y}_t^{\lambda,y+\varepsilon z}) - \sigma(\mathbf{Y}_t^{\lambda,y}) &= \langle \nabla\sigma(\mathbf{Y}_t^{\lambda,y}), \mathbf{D}_t^{\lambda,y,\varepsilon z}\rangle + R_\sigma^{\lambda,y,\varepsilon z}(t) \end{aligned}$$

with

$$\frac{1}{\varepsilon^p} \sup_{t\in[0,T],\|z\|\le 1} (\mathbb{E}|R_a^{\lambda,y,\varepsilon z}(t)|^p + \mathbb{E}|R_\sigma^{\lambda,y,\varepsilon z}(t)|^p) \to 0, \quad \varepsilon \to 0. \tag{6.10}$$

Step 3. Using (6.10), we apply to the pair $\varepsilon^{-1}D^{\lambda,y,\varepsilon z}$, $U^{\lambda,y,z}$ the argument from *Step* 1, and obtain that

$$\sup_{\|z\|\le 1}\mathbb{E}\sup_{t\in[0,T]}\left|\varepsilon^{-1}D^{\lambda,y,\varepsilon z}(t)-U^{\lambda,y,z}(t)\right|^p\to 0,\quad \varepsilon\to 0. \tag{6.11}$$

This is just the relation (6.4) for $Z^y(t)=Y^{\lambda,y}(t)$ with $\nabla_y Z^y(t)=U^{\lambda,y,z}(t)$. The linearity of the mapping (6.5) and the bound (6.6) can be verified straightforwardly, since $U^{\lambda,y,z}$ solves the linear equation (6.7) and $\nabla a, \nabla\sigma, \nabla\psi$, are bounded. This completes the proof of the required differentiability for the $\mathbb{R}^d$-valued family $\{Y^{\lambda,y}(t)\}$.

Step 4. We have already proved (6.4) for $Z^y(t)=Y^{\lambda,y}(t)$ with $\nabla_y Z^y(t)=U^{\lambda,y,z}(t)$. These processes are given by (6.1), (6.7), thus using Doob's maximal inequality we can improve this relation and get

$$\sup_{\|z\|\le 1}\mathbb{E}\sup_{t\in[0,T]}\left\|\frac{Z^{y+\varepsilon z}(t)-Z^y(t)}{\varepsilon}-\nabla_z Z^y(t)\right\|^p_{\mathbb{R}^d}\to 0,\quad \varepsilon\to 0.$$

This yields (6.4) for the $\mathbb{C}$-valued processes $\widetilde{Z}^y(t)=\mathbf{Y}^{\lambda,y}_t$ with $\nabla_z\widetilde{Z}^y(t)=\mathbf{U}^{\lambda,y,z}_t$. The linearity of (6.5) for $\nabla_z\widetilde{Z}^y(t)$ follows trivially from the same property for $\nabla_z\widetilde{Z}^y(t)$, and the bound (6.6) for $\nabla_z\widetilde{Z}^y(t)$ follows from the same bound for $\nabla_z Z^y(t)$ and Doob's maximal inequality. This completes the proof of the required differentiability for the $\mathbb{C}$-valued family $\{\mathbf{Y}^{\lambda,y}_t\}$ and identity (6.8). □

LEMMA 6.2. *Let* $\mathbf{C}^{(1)}$ *hold true. Then for any* $\lambda\ge 0$, $T>0$, $p\ge 1$ *the following hold*:

1. *The family* $\beta^{\lambda,y}(t), t\in[0,T], y\in\mathbb{C}$ *is* L_p*-Fréchet differentiable w.r.t.* y, *and*

$$\Theta^{\lambda,y,z}(t):=\nabla_z\beta^{\lambda,y}(t)=\lambda(\nabla\sigma^{-1})(\mathbf{Y}^{\lambda,y}_t)\mathbf{U}^{\lambda,y,z}_t\psi(Y^{\lambda,y}(t)-X^x(t))$$
$$+\lambda\sigma(\mathbf{Y}^{\lambda,y}_t)^{-1}\langle\nabla\psi(Y^{\lambda,y}(t)-X^x(t)),U^{\lambda,y,z}(t)\rangle.$$

2. *The family* $\ell^{\lambda,y}(t)=\log\mathcal{E}^{\lambda,y}(t), t\in[0,T], y\in\mathbb{C}$ *is* L_p*-Fréchet differentiable w.r.t.* y, *and*

$$\ell^{\lambda,y,z}(t):=\nabla_z\ell^{\lambda,y}(t)=\int_0^t\Theta^{\lambda,y,z}(s)\,\mathrm{d}W_s-\int_0^t\beta^{\lambda,y}(s)\cdot\Theta^{\lambda,y,z}(s)\,\mathrm{d}s.$$

3. *The family* $\mathcal{E}^{\lambda,y}(t), t\in[0,T], y\in\mathbb{C}$ *is* L_p*-Fréchet differentiable w.r.t.* y, *and*

$$\nabla_z\mathcal{E}^{\lambda,y}(t)=\mathcal{E}^{\lambda,y}(t)\ell^{\lambda,y,z}(t).$$

PROOF. Statement *1* follows from Lemma 6.1 by the chain rule, and straightforwardly implies statement *2*. To prove statement *3*, we first observe that by statement *2*,

$$\frac{\mathcal{E}^{\lambda,y+\varepsilon z}(t)-\mathcal{E}^{\lambda,y}(t)}{\varepsilon}\to\mathcal{E}^{\lambda,y}(t)\ell^{\lambda,y,z}(t),\quad \varepsilon\to 0 \tag{6.12}$$

in probability uniformly in $t\in[0,T]$. In addition, by the elementary inequality

$$\left|e^a-1\right|\le C|a|(1+e^a),\quad a\in\mathbb{R},$$

we have

$$\left|\frac{\mathcal{E}^{\lambda,y+\varepsilon z}(t)-\mathcal{E}^{\lambda,y}(t)}{\varepsilon}\right|\le C\left|\frac{\log\mathcal{E}^{\lambda,y+\varepsilon z}(t)-\log\mathcal{E}^{\lambda,y}(t)}{\varepsilon}\right|(\mathcal{E}^{\lambda,y+\varepsilon z}(t)+\mathcal{E}^{\lambda,y}(t)).$$

By (6.4) for the family $\log\mathcal{E}^{\lambda,y}(t), t\in[0,T], y\in\mathbb{C}$ and (6.3), this yields that the left-hand side term in (6.12) has uniformly bounded L_p-norm for every p, and thus is uniformly $L_{p'}$-integrable for any $p'<p$. This yields that (6.12) holds true in L_p for any p. This proves (6.4)

for the family $\mathcal{E}^{\lambda,y}(t), t\in[0,T], y\in\mathbb{C}$; the proofs of linearity for (6.5) and of the bound (6.6) are easy and omitted. □

Now it is easy to complete the proof of Theorem 2.4. By (6.2), Lemma 6.1, and statements 2,3 of Lemma 6.2 applied at the point $y=x$, we get

$$\begin{aligned}&\frac{1}{\varepsilon}\big(\mathbb{E}_{x+\varepsilon z}f(\mathbf{X}_t)-\mathbb{E}_x f(\mathbf{X}_t)\big)\\&\quad=\frac{1}{\varepsilon}\big(\mathbb{E}f\big(\mathbf{Y}_t^{\lambda,x+\varepsilon z}\big)\mathcal{E}^{\lambda,x+\varepsilon z}(t)-\mathbb{E}f\big(\mathbf{Y}_t^{\lambda,x}\big)\mathcal{E}^{\lambda,x}(t)\big)\\&\quad\to\mathbb{E}\langle\nabla f\big(\mathbf{Y}_t^{\lambda,x}\big),\mathbf{U}_t^{\lambda,x,z}\rangle\mathcal{E}^{\lambda,x}(t)\\&\qquad+\mathbb{E}f\big(\mathbf{Y}_t^{\lambda,x}\big)\mathcal{E}^{\lambda,x}(t)\bigg(\int_0^t\Theta^{\lambda,x,z}(s)\,\mathrm{d}W_s+\int_0^t\beta^{\lambda,x}(s)\cdot\Theta^{\lambda,x,z}(s)\,\mathrm{d}s\bigg)\end{aligned}$$

as $\varepsilon\to 0$ uniformly in $t\in[0,T]$, $\|z\|\le 1$. Note that $Y^{\lambda,x}=X^x$, and thus $\beta^{\lambda,x}(t)\equiv 0$, $\mathcal{E}^{\lambda,x}(t)\equiv 1$. In addition, $\psi(0)=0$, $\nabla\psi(0)=I_{\mathbb{R}^n}$, and thus

$$\Theta^{\lambda,x,z}(t)=\lambda\sigma\big(\mathbf{X}_t^x\big)^{-1}U^{\lambda,x,z}(t).$$

Finally, since $\nabla\psi(0)=I_{\mathbb{R}^n}$ equation (6.7) for $U^{\lambda,x,z}$ coincides with equation (2.20). That is, $U^{\lambda,x,z}=U^{\lambda,z}$, which completes the proof of Theorem 2.4.

6.2. *Proof of Theorem* 2.5. First, we fix $h>r$ such that (2.15) holds true; it is an assumption of Theorem 2.3 (and thus of Theorem 2.5) that such h exists. In what follows, the constants may depend on h but we do not indicate this in the notation.

Next, we note that, for any $p\ge 1$ and $Q>0$, one can fix λ large enough such that

$$\mathbb{E}_x\big\|\mathbf{U}_t^{\lambda,z}\big\|^p\le Ce^{-pQt}\|z\|^p,\quad t\ge 0. \tag{6.13}$$

This follows by Itô's formula and Lemma B.2 applied to the family of processes $V^{\lambda,v}(t)=|U^{\lambda,z}(t)|^2$, $\lambda>0$, $v=|z|^2\in\mathbb{C}^+_{\text{real}}$. In the sequel, we use this inequality for the particular value $p=(1-\delta)^{-1}$ and a fixed Q. Note that by (6.13) we have

$$\big|\mathbb{E}_x\langle\nabla f(\mathbf{X}_t),\mathbf{U}_t^{\lambda,z}\rangle\big|\le Ce^{-Qt}\sup_{y\in\mathbb{C}}\big\|\nabla f(y)\big\|\|z\|. \tag{6.14}$$

Next, we note that by (6.13), $\mathbf{H}_3$, and the Burkholder–Davis–Gundy inequality, for any $t_1\le t_2$

$$\mathbb{E}_x\bigg|\int_{t_1}^{t_2}\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\bigg|^p\le Ce^{-pQt_1}\|z\|^p. \tag{6.15}$$

For $t_0\in[0,t]$, we have

$$\begin{aligned}&\mathbb{E}_x\bigg(f(\mathbf{X}_t)\int_{t_0}^t\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\bigg)\\&\quad=\mathbb{E}_x\bigg(\big(f(\mathbf{X}_t)-f(0)\big)\int_{t_0}^t\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\bigg),\end{aligned}$$

hence using that

$$\sup_x\big\|f(x)-f(0)\big\|\le\|f\|_{H_\gamma}, \tag{6.16}$$

we get

$$\bigg|\lambda\mathbb{E}_x\bigg(f(\mathbf{X}_t)\int_{t_0}^t\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\bigg)\bigg|\le Ce^{-Qt_0}\|f\|_{H_\gamma}\|z\|. \tag{6.17}$$

On the other hand, by the Markov property,

$$\mathbb{E}_x\Big(f(\mathbf{X}_t)\int_0^{t_0}\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\Big)$$
$$=\mathbb{E}_x\Big(P_{t-t_0}f(\mathbf{X}_{t_0})\int_0^{t_0}\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\Big).$$

Denote

$$\overline{f}=\int_{\mathbb{C}}f(y)\pi(\mathrm{d}y).$$

By (2.9) and (2.18),

$$\big|P_sf(x)-\overline{f}\big|\le C\frac{1}{r(c_0s)^\delta}\phi\big(V(x)\big)^\delta\|f\|_{H_\gamma}$$

for some $c_0>0$. Clearly,

$$\mathbb{E}_x\Big(\overline{f}\int_0^{t_0}\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\Big)=0,$$

hence by (6.15) with $t_1=0, t_2=t_0$ and Hölder's inequality applied to $p=(1-\delta)^{-1}, q=\delta^{-1}$, we have

$$\begin{aligned}&\Big|\mathbb{E}_x\Big(f(\mathbf{X}_t)\int_0^{t_0}\sigma(\mathbf{X}_s)^{-1}U^{\lambda,z}(s)\,\mathrm{d}W(s)\Big)\Big|\\&\quad\le C\frac{1}{r(c_0(t-t_0))^\delta}\big(\mathbb{E}_x\phi\big(V(\mathbf{X}_{t_0})\big)\big)^\delta\|f\|_{H_\gamma}\|z\|.\end{aligned}\tag{6.18}$$

By Jensen's inequality, $\mathbb{E}_x\phi(V(\mathbf{X}_{t_0}))\le\phi(\mathbb{E}_xV(\mathbf{X}_{t_0}))$. On the other hand, if $t_0=kh, k\in\mathbb{N}\cup\{0\}$, then it follows from (2.15) that

$$\mathbb{E}_xV(\mathbf{X}_{t_0})\le V(x)+kC_V.$$

Since ϕ is concave and nonnegative, we have

$$\phi(u+v)\le\phi(u)+\phi(v)-\phi(0)\le\phi(u)+\phi(v).\tag{6.19}$$

Hence

$$\big(\mathbb{E}_x\phi\big(V(\mathbf{X}_{t_0})\big)\big)^\delta\le C\big(t_0+\phi\big(V(x)\big)\big)^\delta,\quad t_0\in h\mathbb{N}\cup\{0\}.\tag{6.20}$$

Now we can complete the proof. Without loss of generality, we can assume $Q\ge1$. Since the function $\phi(v)$ is sublinear (see (6.19)), the function $r(t)$ is subexponential. In particular, we can fix $c\in(0,c_0/2)$ small enough and $t_*>0$ such that

$$\log r(\zeta t)\le\frac{Qt}{2},\quad t\ge t_*, Q\ge1;$$

recall also that we have assumed $\phi(v)>1$, and thus $\log r(t)>0, t\ge0$. Take for $t\ge t_*$

$$t_0=h\lfloor h^{-1}Q^{-1}\log r(\zeta t)\rfloor\in h\mathbb{N}\cup\{0\},$$

then $t_0\le Q^{-1}\log r(\zeta t), t-t_0\ge t/2$. In addition,

$$r\big(c_0(t-t_0)\big)\ge r(\zeta t)$$

because $r(\cdot)$ is increasing and $c < c_0/2$. Combining the representation (2.21) with (6.14), (6.17), (6.18) and (6.20), we get for $Q \geq 1$,

$$\begin{aligned}\big|\nabla_z \mathbb{E}_x f(\mathbf{X}_t)\big| &\leq Ce^{-Qt} \sup_{y\in\mathbb{C}} \|\nabla f(y)\| \|z\| + \frac{C}{r(\zeta t)} \|f\|_{H_\gamma} \|z\| \\ &\quad + \frac{C}{r(\zeta t)^\delta} \big(Q^{-1} \log r(\zeta t) + \phi\big(V(x)\big)\big)^\delta \|f\|_{H_\gamma} \|z\|, \quad x, y \in \mathbb{C},\end{aligned}$$

which after a simple rearrangement gives (2.22) for $t \geq t_*$. The proof of (2.22) for $t \leq t_*$ is easy and omitted.

6.3. *Proof of Theorem* 2.6. For $k > 1$, the argument remains principally the same as the one developed for $k = 1$ in the two previous sections, with just technical complications which makes the proof more cumbersome. Thus we just outline the main steps of the proof, paying particular attention to one new circumstance; see *Step* 5 below. Everywhere below we assume $\mathbf{C}^{(k)}$ to hold for some $k > 1$.

Step 1. By iteration of the argument in the proof of Lemma 6.1, we obtain that the family $Y^{\lambda,y}(t), t \in [0, T], y \in \mathbb{C}$ is k times L_p-Fréchet differentiable w.r.t. y, and the corresponding direction-wise derivatives $U^{\lambda,y,z_1,\dots,z_k}(t) = \nabla_{z_1} \dots \nabla_{z_k} Y^{\lambda,y}(t)$ satisfy SDDEs of the form

$$\begin{aligned}\mathrm{d}U^{\lambda,y,z_1,\dots,z_k}(t) &= \big\langle \nabla a\big(\mathbf{Y}_t^{\lambda,y}\big), \mathbf{U}_t^{\lambda,y,z_1,\dots,z_k}\big\rangle \mathrm{d}t + \big\langle \nabla\sigma\big(\mathbf{Y}_t^{\lambda,y}\big), \mathbf{U}_t^{\lambda,y,z_1,\dots,z_k}\big\rangle \mathrm{d}W(t) \\ &\quad - \lambda\big\langle \nabla\psi\big(Y^{\lambda,y}(t) - X^x(t)\big), U^{\lambda,y,z_1,\dots,z_k}(t)\big\rangle \mathrm{d}t \\ &\quad + D^{\lambda,y,z_1,\dots,z_k}(t)\,\mathrm{d}t + S^{\lambda,y,z_1,\dots,z_k}(t)\,\mathrm{d}W(t), \quad \mathbf{U}_0^{\lambda,y,z_1,\dots,z_k} = 0.\end{aligned} \tag{6.21}$$

The terms $D^{\lambda,y,z_1,\dots,z_k}(t), S^{\lambda,y,z_1,\dots,z_k}(t)$ can be represented as sums of various l-linear forms, which are expressed in terms of $\nabla^i a(\mathbf{Y}_t^{\lambda,y})$, $\nabla^i \sigma(\mathbf{Y}_t^{\lambda,y})$, $\nabla^i \psi(Y^{\lambda,y}(t) - X^x(t)), i = 1, \dots, k$ (and thus are bounded). The arguments in each of those l-linear forms have the generic form

$$\mathbf{U}_t^{\lambda,y,z_{I_1}}, \dots, \mathbf{U}_t^{\lambda,y,z_{I_l}},$$

where $I_1, \dots, I_l$ is a disjoint partition of $\{1, \dots, k\}$, and we use the notation

$$\mathbf{U}_t^{\lambda,y,z_I} = \mathbf{U}_t^{\lambda,y,z_{i_1},\dots,z_{i_j}}, \qquad I = \{i_1, \dots, i_j\}.$$

Step 2. We have for λ large enough

$$\mathbb{E}\big\|\mathbf{U}_t^{\lambda,x,z_1,\dots,z_k}\big\|^p \leq Ce^{-pQt} \|z_1\|^p \dots \|z_k\|^p, \quad t \geq 0. \tag{6.22}$$

To see this, first recall that $Y^{\lambda,x} = X^x$, $\nabla\psi(0) = I_{\mathbb{R}^n}$. Then (6.21) for $y = x$ transforms to

$$\begin{aligned}&\mathrm{d}U^{\lambda,x,z_1,\dots,z_k}(t) \\ &\quad = \big\langle \nabla a\big(\mathbf{X}_t^x\big), \mathbf{U}_t^{\lambda,x,z_1,\dots,z_k}\big\rangle \mathrm{d}t + \big\langle \nabla\sigma\big(\mathbf{X}_t^x\big), \mathbf{U}_t^{\lambda,x,z_1,\dots,z_k}\big\rangle \mathrm{d}W(t) \\ &\qquad - \lambda U^{\lambda,x,z_1,\dots,z_k}(t)\,\mathrm{d}t + D^{\lambda,x,z_1,\dots,z_k}(t)\,\mathrm{d}t + S^{\lambda,x,z_1,\dots,z_k}(t)\,\mathrm{d}W(t), \\ &\mathbf{U}_0^{\lambda,y,z_1,\dots,z_k} = 0.\end{aligned} \tag{6.23}$$

Now the proof can be made inductively: assuming (6.22) holds true for $1, \dots, k-1$, we have

$$\mathbb{E}\big|D^{\lambda,x,z_1,\dots,z_k}(t)\big|^p + \mathbb{E}\big|S^{\lambda,x,z_1,\dots,z_k}(t)\big|^p \leq Ce^{-pQt} \|z_1\|^p \dots \|z_k\|^p, \quad t \geq 0,$$

which yields (6.22) for k by Lemma B.3.

Step 3. Similar to the proof of Lemma 6.2 we get that the following families are k times L_p-Fréchet differentiable w.r.t. y:

- $\beta^{\lambda,y}(t), t \in [0,T], y \in \mathbb{C}$;
- $\ell^{\lambda,y}(t) := \log \mathcal{E}^{\lambda,y}(t), t \in [0,T], y \in \mathbb{C}$;
- $\mathcal{E}^{\lambda,y}(t_1;t_2) = \mathcal{E}^{\lambda,y}(t_1)^{-1}\mathcal{E}^{\lambda,y}(t_2), t_1,t_2 \in [0,T], y \in \mathbb{C}$.

The corresponding direction-wise derivatives will be defined by

$$\Theta^{\lambda,y,z_1,\dots,z_k}(t) = \nabla_{z_1}\dots\nabla_{z_k}\beta^{\lambda,y}(t), \qquad \ell^{\lambda,y,z_1,\dots,z_k}(t) = \nabla_{z_1}\dots\nabla_{z_k}\ell^{\lambda,y}(t),$$

$$\mathcal{E}^{\lambda,y,z_1,\dots,z_k}(t_1;t_2) = \nabla_{z_1}\dots\nabla_{z_k}\mathcal{E}^{y,\lambda}(t_1;t_2).$$

We have straightforwardly

$$\ell^{\lambda,y,z_1,\dots,z_k}(t) = \int_0^t \Theta^{\lambda,y,z_1,\dots,z_k}(s)\,\mathrm{d}W_s + \sum_{i=1,\dots,k}\int_0^t \Theta^{\lambda,y,z_j}(s)\cdot\Theta^{\lambda,y,z_1,\dots,z_{j-1},z_{j+1},\dots z_k}(s)\,\mathrm{d}s.$$

Step 4. The derivative $\Theta^{\lambda,x,z_1,\dots,z_k}(t)$ can be expressed as a sum of bounded l-linear forms ($l = 1,\dots,k$) applied to

$$\mathbf{U}_t^{\lambda,x,z'_1,\dots,z'_j}, \quad j = 1,\dots,k, z'_1,\dots,z'_j \in \{z_1,\dots,z_k\}.$$

Then by (6.22) we have that for λ large enough

$$\mathbb{E}\big|\Theta^{\lambda,x,z_1,\dots,z_k}(t)\big|^p \le Ce^{-pQt}\|z_1\|^p\dots\|z_k\|^p, \quad t \ge 0. \tag{6.24}$$

This implies

$$\mathbb{E}\big|\ell^{\lambda,x,z_1,\dots,z_k}(t_2) - \ell^{\lambda,x,z_1,\dots,z_k}(t_1)\big|^p \le Ce^{-pQt_1}\|z_1\|^p\dots\|z_k\|^p, \quad t_2 \ge t_1 \ge 0. \tag{6.25}$$

Recall that $\mathcal{E}^{\lambda,x}(t_1;t_2) = 1$, hence the derivative $\mathcal{E}^{\lambda,y,z_1,\dots,z_k}(t_1;t_2)$ is a polynomial of

$$\ell^{\lambda,x,z'_1,\dots,z'_j}(t_2) - \ell^{\lambda,x,z'_1,\dots,z'_j}(t_1), \quad j = 1,\dots,k, z'_1,\dots,z'_j \in \{z_1,\dots,z_k\}.$$

Therefore, for λ large enough

$$\mathbb{E}\big|\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t_1;t_2)\big|^p \le Ce^{-pQt_1}\|z_1\|^p\dots\|z_k\|^p, \quad t_2 \ge t_1 \ge 0. \tag{6.26}$$

Step 5. By (6.2), we have

$$\begin{aligned}\nabla_{z_1}\dots\nabla_{z_k}\mathbb{E}_x f(\mathbf{X}_t) &= \mathbb{E}\nabla_{z_1}\dots\nabla_{z_k}\big(f\big(\mathbf{Y}_t^{\lambda,y}\big)\mathcal{E}^{\lambda,y}(t)\big)\big|_{y=x} \\ &= \sum_{I\cup J=\{1,\dots,k\}, I\cap J=\varnothing}\mathbb{E}\nabla_{z_I}^I\big(f\big(\mathbf{Y}_t^{\lambda,y}\big)\big)\big|_{y=x}\mathcal{E}^{\lambda,x,z_J}(t) \\ &= \sum_{I\cup J=\{1,\dots,k\}, I\cap J=\varnothing, I\neq\varnothing}\mathbb{E}\nabla_{z_I}^I\big(f\big(\mathbf{Y}_t^{\lambda,y}\big)\big)\big|_{y=x}\mathcal{E}^{\lambda,x,z_J}(t) \\ &\quad + \mathbb{E}f\big(\mathbf{X}_t^x\big)\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t),\end{aligned} \tag{6.27}$$

where the following notation is used:

- for $I = \{i_1,\dots,i_m\}$,
$$\nabla_{z_I}^I = \nabla_{z_{i_1}}\dots\nabla_{z_{i_m}};$$
- for $J = \{j_1,\dots,j_r\}$,
$$\mathcal{E}^{\lambda,x,z_J}(t) = \mathcal{E}^{\lambda,x,z_J}(0,t), \qquad \mathcal{E}^{\lambda,x,z_J}(t_1,t_2) = \mathcal{E}^{\lambda,x,z_{j_1},\dots,z_{j_r}}(t_1,t_2).$$

By (6.22) and (6.26) with $t_2 = t, t_1 = 0$, we have

$$\left| \sum_{I\cup J=\{1,\dots,k\}, I\cap J=\varnothing, I\neq\varnothing} \mathbb{E}\nabla^I_{z_I}\bigl(f(\mathbf{Y}^{\lambda,y}_t)\bigr)\big|_{y=x}\mathcal{E}^{\lambda,x,z_J}(t)\right| \\ \le C\|f\|_{(k)}e^{-Qt}\|z_1\|\dots\|z_k\|. \tag{6.28}$$

That is, the first term (the sum) on the right-hand side of (6.27) admits an estimate completely analogous to (6.14).

The second term on the right-hand side of (6.27) is analogous to the second term in (2.21). A slight new difficulty which appears in the case $k > 1$ is that now this term cannot be simply written by means of a stochastic integral: $\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t)$ is actually a mixture of various multiple stochastic integrals. This difficulty can be avoided by the following trick, which makes it possible to estimate this term without a study of its inner structure. We have for arbitrary $t_0 \le t$,

$$\begin{aligned}\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t) &= \nabla_{z_1}\dots\nabla_{z_k}\bigl(\mathcal{E}^{\lambda,y}(0;t_0)\mathcal{E}^{\lambda,y}(t_0;t)\bigr)\big|_{y=x}\\ &= \sum_{I\cup J=\{1,\dots,k\}, I\cap J=\varnothing}\mathcal{E}^{\lambda,x,z_I}(0;t_0)\mathcal{E}^{\lambda,x,z_J}(t_0;t).\end{aligned} \tag{6.29}$$

By (6.16) and (6.26),

$$\left|\mathbb{E}f(\mathbf{X}^x_t)\sum_{I\cup J=\{1,\dots,k\}, I\cap J=\varnothing, J\neq\varnothing}\mathcal{E}^{\lambda,x,z_I}(0;t_0)\mathcal{E}^{\lambda,x,z_J}(t_0;t)\right| \\ \le C\|f\|_{H_\gamma}e^{-Qt_0}\|z_1\|\dots\|z_k\|, \tag{6.30}$$

which is a straightforward analogue to (6.17). The term with $J = \varnothing$ equals

$$\mathbb{E}f(\mathbf{X}^x_t)\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t_0) = \mathbb{E}P_{t-t_0}f(\mathbf{X}_{t_0})\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t_0).$$

Note that

$$\mathbb{E}\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t_0) = \nabla_{z_1}\dots\nabla_{z_k}\mathbb{E}\mathcal{E}^{\lambda,y}(t_0)|_{y=x} = \nabla_{z_1}\dots\nabla_{z_k}1 = 0.$$

Repeating literally the calculations used in the proof of (6.18) and using (6.20), we get

$$\bigl|\mathbb{E}f(\mathbf{X}^x_t)\mathcal{E}^{\lambda,x,z_1,\dots,z_k}(t_0)\bigr| \le C\frac{1}{r(c_0(t-t_0))^\delta}\bigl(t_0+\phi(V(x))\bigr)^\delta\|f\|_{H_\gamma}\|z\| \tag{6.31}$$

for some $c_0 > 0$. Using (6.28), (6.30), (6.31) and repeating the optimization in t_0 procedure from the last part of the proof of Theorem 2.5, we complete the proof of the theorem.

APPENDIX A: THE KULLBACK–LEIBLER DIVERGENCE AND RELATED BOUNDS

For a pair of probability measures $\mu \ll \nu$ on a measurable space $(X, \mathcal{X})$ *the Kullback–Leibler (KL–) divergence of μ from ν* is defined by

$$D_{\mathrm{KL}}(\mu\|\nu) := \int_X \log\frac{d\mu}{d\nu}\,\mathrm{d}\mu = \int_X \frac{d\mu}{d\nu}\log\left(\frac{d\mu}{d\nu}\right)\mathrm{d}\nu.$$

The KL–divergence is known to be a stronger measure of difference between probability distributions than the total variation distance; in particular, the following *Pinsker inequality* holds true, for example, [28], Lemma 2.5(i):

$$d_{\mathrm{TV}}(\mu,\nu) \le \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\mu\|\nu)}. \tag{A.1}$$

In addition, the KL-divergence yields the following lower bound; see [6], Lemma A.1: for any $N > 1$ and any set $A \in \mathcal{X}$,

$$\nu(A) \geq \frac{1}{N}\mu(A) - \frac{D_{\mathrm{KL}}(\mu\|\nu) + \log 2}{N \log N}. \tag{A.2}$$

Next, let ξ be an m-dimensional Itô process with $\xi_0 = 0$ and

$$\mathrm{d}\xi_t = \beta_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0, \tag{A.3}$$

where W is a Wiener process in $\mathbb{R}^m$, and $(\beta_t)_{t\geq 0}$ is a progressively measurable process. The following bound is available for the KL–divergence of the law μ_ξ of the process ξ on $C([0,\infty),\mathbb{R}^m)$ w.r.t. of the law μ_W of W (the Wiener measure).

THEOREM A.1. [6], *Theorem A.2*,

$$D_{\mathrm{KL}}(\mu_\xi\|\mu_W) \leq \frac{1}{2}\mathbb{E}\int_0^\infty |\beta_t|^2 \,\mathrm{d}t.$$

APPENDIX B: AUXILIARY TAIL- AND L_p-ESTIMATES

The following lemma was suggested by R. Schilling.

LEMMA B.1. *Let $V(t) \geq 0$ be an Itô process with*

$$\mathrm{d}V(t) = \eta(t)\,\mathrm{d}t + \mathrm{d}M(t),$$

where M is a continuous local martingale with quadratic variation

$$\langle M\rangle(t) = \int_0^t m(s)\,\mathrm{d}s, \quad t \geq 0.$$

Let for some constants $A \geq 0$, $B > 0$, $\lambda > 0$ and a random variable $\varsigma \geq 0$

$$\eta(t) \leq -\lambda V(t) + A, \quad m(t) \leq B, t \leq \varsigma.$$

Assume also that $\varsigma \leq T$ for some constant $T > 0$.

Then for every $\delta \in (0, 1/2)$ there exist C_1, $C_2 > 0$, which depend only on δ and T, such that

$$\mathbb{P}\Big(\sup_{t\leq\varsigma}\big(V(t) - e^{-\lambda t}V(0)\big) \geq A\lambda^{-1} + B^{1/2}\lambda^{-\delta}R\Big) \leq C_1 e^{-C_2 R^2}, \quad R \geq 0.$$

PROOF. We have

$$V(t) = e^{-\lambda t}V(0) + \int_0^t e^{-\lambda(t-s)}\xi(s)\,\mathrm{d}s + \int_0^t e^{-\lambda(t-s)}\,\mathrm{d}M(s), \tag{B.1}$$

where

$$\xi(t) = \eta(t) + \lambda V(t) \leq A, \quad t \leq \varsigma.$$

Clearly,

$$\int_0^t e^{-\lambda(t-s)}\xi(s)\,\mathrm{d}s \leq A\lambda^{-1}\big(1 - e^{-\lambda t}\big), \quad t \leq \varsigma,$$

and we have to study the third term in (B.1), only. Without loss of generality, we can assume that $M(0) = 0$. By the Dambis–Dubins–Schwarz theorem (see, e.g., [25], Theorem 5.1.6), extending the probability space, if necessary, we can find a standard Brownian motion $\widetilde{W}$ such that

$$\int_0^t e^{\lambda s}\,\mathrm{d}M(s) = B^{1/2}\widetilde{W}\left(\int_0^t e^{2\lambda s}\frac{m(s)}{B}\,\mathrm{d}s\right), \quad t \geq 0.$$

Therefore,

$$\begin{aligned}\sup_{0\leq t\leq \varsigma}\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}M(s) &= B^{1/2}\sup_{0\leq t\leq \varsigma}\left\{e^{-\lambda t}\widetilde{W}\left(\int_0^t e^{2\lambda s}\frac{m(s)}{B}\,\mathrm{d}s\right)\right\}\\ &\leq B^{1/2}\sup_{0\leq t\leq \varsigma}\left\{e^{-\lambda t}\sup_{0\leq u\leq t}\widetilde{W}\left(\int_0^u e^{2\lambda s}\frac{m(s)}{B}\,\mathrm{d}s\right)\right\}\\ &\leq B^{1/2}\sup_{0\leq t\leq \varsigma}\left\{e^{-\lambda t}\sup_{0\leq u\leq t}\widetilde{W}\left(\int_0^u e^{2\lambda s}\,\mathrm{d}s\right)\right\}.\end{aligned}$$

Next, $\widetilde{M}(t) = \widetilde{W}(\int_0^t e^{2\lambda s}\,\mathrm{d}s)$ is a Gaussian martingale with characteristic $\langle\widetilde{M}\rangle(t) = \int_0^t e^{2\lambda s}\,\mathrm{d}s$, hence

$$\widehat{W}(t) = \int_0^t e^{-\lambda s}\,\mathrm{d}\widetilde{M}(s)$$

is a standard Brownian motion such that

$$\widetilde{W}\left(\int_0^t e^{2\lambda s}\,\mathrm{d}s\right) = \int_0^t e^{\lambda s}\,\mathrm{d}\widehat{W}(s), \quad t\geq 0.$$

Using that $t\mapsto e^{-\lambda t}$ is decreasing, we get finally

$$\begin{aligned}\sup_{0\leq t\leq \varsigma}\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}M(s) &\leq B^{1/2}\sup_{0\leq t\leq \varsigma}\left\{e^{-\lambda t}\sup_{0\leq u\leq t}\int_0^u e^{\lambda s}\,\mathrm{d}\widehat{W}(s)\right\}\\ &\leq B^{1/2}\sup_{0\leq t\leq \varsigma}\left\{\sup_{0\leq u\leq t}e^{-\lambda u}\int_0^u e^{\lambda s}\,\mathrm{d}\widehat{W}(s)\right\}\\ &= B^{1/2}\sup_{0\leq t\leq \varsigma}\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}\widehat{W}(s).\end{aligned}$$

For $N > 0$ and $\delta < 1/2$, denote

$$D_N^{\delta,T} = \{|\widehat{W}(t)-\widehat{W}(s)|\leq N|t-s|^{\delta}; s,t\leq T\},$$

and observe that, by Fernique's theorem,

$$\mathbb{P}(\Omega\setminus D_N^{\delta,T})\leq c_1 e^{-c_2N^2}, \quad N>0 \tag{B.2}$$

with constants c_1, c_2 which depend only on δ and T. Observe that

$$\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}\widehat{W}(s) = \int_0^t (\widehat{W}(t)-\widehat{W}(s))\lambda e^{-\lambda(t-s)}\,\mathrm{d}s + e^{-\lambda t}\widehat{W}(t).$$

Therefore, on the set set $D_N^{\delta,T}$, we obtain for $t\leq\varsigma\leq T$

$$\begin{aligned}\int_0^t e^{-\lambda(t-s)}\,\mathrm{d}M(s) &\leq NB^{1/2}\int_0^t (t-s)^{\delta}\lambda e^{-\lambda(t-s)}\,\mathrm{d}s + NB^{1/2}e^{-\lambda t}t^{\delta}\\ &\leq NB^{1/2}\lambda^{-\delta}\left(\Gamma(\delta)+\sup_{x>0}x^{\delta}e^{-x}\right),\end{aligned}$$

where $\Gamma(\delta)$ denotes the Euler Gamma-function. Taking

$$N = \Big(\Gamma(\delta) + \sup_{x>0} x^\delta e^{-x}\Big) R$$

and using (B.2), the proof is complete. □

Denote $\mathbb{C}^+_{\text{real}}$ the set of nonnegative functions in $C([-r,0],\mathbb{R})$.

LEMMA B.2. *For each $\lambda > 0$, $v \in \mathbb{C}^+_{\text{real}}$, let $V^{\lambda,v}(t) \ge 0$, $t \ge -r$ be an adapted process with continuous paths such that*

$$\mathrm{d}V^{\lambda,v}(t) = \eta^{\lambda,v}(t)\,\mathrm{d}t + \mathrm{d}M^{\lambda,v}(t), \quad t \ge 0, \qquad \mathbf{V}_0^{\lambda,v} = v \in \mathbb{C}^+_{\text{real}},$$

where $M^{\lambda,v}$ is a continuous local martingale and for some constant $K \ge 0$,

$$\eta^{\lambda,v}(t) \le K\|\mathbf{V}_t^{\lambda,v}\| - \lambda V^{\lambda,v}(t), \qquad \frac{\mathrm{d}\langle M^{\lambda,v}\rangle(t)}{\mathrm{d}t} \le K^2\|\mathbf{V}_t^{\lambda,v}\|^2.$$

Then for each $p \ge 1$ and $Q > 0$ there exist $\lambda_{p,Q,K} > 0$ and a constant $C_{p,Q,K}$ such that for $\lambda \ge \lambda_{p,Q,K}$,

$$\mathbb{E}\|\mathbf{V}_t^{\lambda,v}\|^p \le C_{p,Q,K} e^{-Qt}\|v\|^p, \quad t \ge 0, v \in \mathbb{C}^+_{\text{real}}.$$

PROOF. First, we note that by the Gronwall lemma and the Burkholder–Davis–Gundy inequality for each $T > 0$, $p \ge 1$ there exists a constant $C_{p,T,K}$ such that

$$\mathbb{E}\sup_{t\le T} V^{\lambda,v}(t)^p \le C_{p,T,K}\|v\|^p; \tag{B.3}$$

the argument here is the same as in Section 4.1. That is, to prove the required statement it is enough to find for given p, Q some $T = T_{p,Q,K} > 0$ and $\lambda_{p,Q,K} > 0$ such that

$$\mathbb{E}\|\mathbf{V}_t^{\lambda,v}\|^p \le C_{p,Q,K} e^{-QT}\|v\|^p, \quad \lambda \ge \lambda_{p,Q,K}, v \in \mathbb{C}^+_{\text{real}}. \tag{B.4}$$

We fix $T = 2r$ and put

$$\tau^{\lambda,v} = \inf\{t : |V^{\lambda,v}(t)| \ge 2\|v\|\} \wedge T.$$

Then the assumption of Lemma B.1 holds true for $V = V^{\lambda,v}$, $\tau = \tau^{\lambda,v}$ with $A = 2K\|v\|$, $B = 4K^2\|v\|^2$. Note that this lemma yields

$$V^{\lambda,v}(t) \le e^{-\lambda t}v(0) + \frac{A}{\lambda} + \frac{B^{1/2}}{\lambda^\delta}\Xi^{\lambda,v}(t), \quad t \le \tau^{\lambda,v}$$

with a fixed $\delta < 1/2$ and

$$\Big\|\sup_{t\le T}\Xi^{\lambda,v}(t)\Big\|_{L_p} \le C'_{p,T}.$$

We will take $\lambda_{p,Q,K} \ge 4K$, then for $\lambda \ge \lambda_{p,Q,K}$,

$$V^{\lambda,v}(t)1_{t\le\tau^{\lambda,v}} \le \Big(e^{-\lambda t} + \frac{2K}{\lambda} + \frac{2K}{\lambda^\delta}\Xi^{\lambda,v}(t)\Big)\|v\| \le \Big(\frac{3}{2} + \frac{2K}{\lambda^\delta}\Xi^{\lambda,v}(t)\Big)\|v\|,$$

which yields

$$\mathbb{P}(\tau^{\lambda,v} < T) \le \mathbb{P}\Big(\Xi^{\lambda,v}(\tau^{\lambda,v}) \ge \frac{\lambda^\delta}{4K}\Big) \le C'_{p,T}(4K)^p\lambda^{-\delta p}.$$

Then by (B.3) and the Cauchy inequality

$$\mathbb{E}\sup_{t\in[0,T]} V^{\lambda,v}(t)^p 1_{\tau^{\lambda,v}<T} \le \Big(\mathbb{E}\sup_{t\le T} V^{\lambda,v}(t)^{2p}\Big)^{1/2}\big(\mathbb{P}(\tau^{\lambda,v}<T)\big)^{1/2}$$
$$\le \frac{(C_{2p,T}C'_{p,T}(4K)^p)^{1/2}}{\lambda^{p\delta/2}}\|v\|^p.$$

Combining these calculations, we get

$$V^{\lambda,v}(t) \le \bigg(e^{-\lambda t} + \frac{2K}{\lambda} + \bigg(\frac{2K}{\lambda^{\delta/2}} \vee \sqrt{\frac{2K}{\lambda^{\delta/2}}}\bigg)\Delta^{\lambda,v,K}\bigg)\|v\|, \quad t\in[0,T]$$

with

$$\|\Delta^{\lambda,v,K}\|_{L_p} \le C'_{p,T}.$$

In particular,

$$\|\mathbf{V}_T^{\lambda,v}\| \le \bigg(e^{-\lambda T/2} + \frac{2K}{\lambda} + \frac{2K}{\lambda^{\delta/2}}\Delta^{\lambda,v,K}\bigg)\|v\|,$$

which easily yields (B.4) for $\lambda \ge \lambda_{p,Q,K}$ and $\lambda_{p,Q,K}$ large enough. □

The following lemma can be proved by essentially the same argument; we leave the details for the reader.

LEMMA B.3. *Let $k>1$, and for each $\lambda>0, v_1,\dots,v_k \in \mathbb{C}^+_{\mathrm{real}}$ let $V^{\lambda,v_1,\dots,v_k}(t)\ge 0$, $t\ge -r$ be an adapted process with continuous paths such that*

$$\mathrm{d}V^{\lambda,v_1,\dots,v_k}(t) = \eta^{\lambda,v_1,\dots,v_k}(t)\,\mathrm{d}t + \mathrm{d}M^{\lambda,v_1,\dots,v_k}(t), \quad t\ge 0, \qquad \mathbf{V}_0^{\lambda,v_1,\dots,v_k}=0,$$

where $M^{\lambda,v_1,\dots,v_k}$ is a continuous local martingale and, for some constant $K\ge 0$,

$$\eta^{\lambda,v_1,\dots,v_k}(t) \le K\|\mathbf{V}_t^{\lambda,v_1,\dots,v_k}\| - \lambda V^{\lambda,v_1,\dots,v_k}(t) + L^{\lambda,v_1,\dots,v_k}(t),$$
$$\frac{\mathrm{d}\langle M^{\lambda,v_1,\dots,v_k}\rangle(t)}{\mathrm{d}t} \le K^2\|\mathbf{V}_t^{\lambda,v_1,\dots,v_k}\|^2 + N^{\lambda,v_1,\dots,v_k}(t).$$

Fix $p\ge 1$ and assume that the nonnegative adapted processes $L^{\lambda,v_1,\dots,v_k}(t)$, $N^{\lambda,v_1,\dots,v_k}(t)$, $t\ge 0$ are such that for each $Q>0$ there exist $\lambda^0_{p,Q}>0$ and a constant $C^0_{p,Q}$ such that

$$\mathbb{E}\|\mathbf{L}_t^{\lambda,v_1,\dots,v_k}\|^p \le C^0_{p,Q}e^{-Qt}\|v_1\|^p\dots\|v_k\|^p,$$
$$\mathbb{E}\|\mathbf{N}_t^{\lambda,v_1,\dots,v_k}\|^{2p} \le C^0_{p,Q}e^{-Qt}\|v_1\|^p\dots\|v_k\|^p, \quad t\ge 0$$

for any $\lambda\ge\lambda^0_{p,Q}$ and $v_1,\dots,v_k\in\mathbb{C}^+_{\mathrm{real}}$.

Then for each $Q>0$ there exist $\lambda_{p,Q,K}>0$ and a constant $C_{p,Q,K}$ such that for $\lambda\ge\lambda_{p,Q,K}$

$$\mathbb{E}\|\mathbf{V}_t^{\lambda,v_1,\dots,v_k}\|^p \le C_{p,Q,K}e^{-Qt}\|v_1\|^p\dots\|v_k\|^p, \quad t\ge 0, v_1,\dots,v_k\in\mathbb{C}^+_{\mathrm{real}}.$$

APPENDIX C: DEVIATION BOUNDS FOR THE CONTROLLED PROCESSES

In this section, we prove Proposition 3.1 and Proposition 5.2.

PROOF OF PROPOSITION 3.1. To shorten the notation, we will write υ instead of $\upsilon_{x,y,K}$; recall that υ is assumed to be positive. We have by Itô's formula

$$|X(t)-Y(t)|^2 = |x(0)-y(0)| + \int_0^t A(s)\,ds + \int_0^t \Sigma(s)\,dW(s), \quad t \ge 0, \tag{C.1}$$

where

$$A(s) = 2\big(a(\mathbf{X}_s) - \widetilde{a}(\mathbf{Y}_s), X(s)-Y(s)\big) + \|\sigma(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{Y}_s)\|^2 - 2\upsilon^{\gamma-1}|X(s)-Y(s)|^2 1_{s\le\tau},$$

$$\Sigma(s) = 2\big(X(s)-Y(s)\big)^\top\big(\sigma(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{Y}_s)\big).$$

We have

$$\|\sigma(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{Y}_s)\|^2 \le 2\|\widetilde{\sigma}(\mathbf{X}_s) - \sigma(\mathbf{X}_s)\|^2 + 2\|\widetilde{\sigma}(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{Y}_s)\|^2,$$

and thus $A(s) \le A^{(1)}(s) + A^{(2)}(s)$ with

$$A^{(1)}(s) = 2\big(a(\mathbf{X}_s) - \widetilde{a}(\mathbf{X}_s), X(s)-Y(s)\big) + 2\|\sigma(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{X}_s)\|^2,$$

$$A^{(2)}(s) = 2\big(\widetilde{a}(\mathbf{X}_s) - \widetilde{a}(\mathbf{Y}_s), X(s)-Y(s)\big) + 2\|\widetilde{\sigma}(\mathbf{X}_s) - \widetilde{\sigma}(\mathbf{Y}_s)\|^2 - 2\upsilon^{\gamma-1}|X(s)-Y(s)|^2 1_{s\le\tau}.$$

Denote $\theta = \tau \wedge \theta_K$, then for $t \le \theta$ we have

$$\mathbf{X}_t \in K, \ |X(t)-Y(t)| \le 2\upsilon, \tag{C.2}$$

which simply gives

$$|A^{(1)}(s)| \le 2\big(2\Delta_{a,K}\upsilon + \Delta_{\sigma,K}^2\big) \le 4\big(\upsilon^{\alpha+1} + \upsilon^{2\beta}\big), \quad s \le \theta, \tag{C.3}$$

see (3.5) for the definition of $\upsilon = \upsilon_{x,y,K}$. The second inequality in (C.2) clearly yields

$$\|\mathbf{X}_t - \mathbf{Y}_t\| \le 2\upsilon, \quad t \le \theta.$$

Then by the conditions $\mathbf{H}_1, \mathbf{H}_2$ for $\widetilde{a}, \widetilde{\sigma}$

$$A^{(2)}(s) \le C_a\upsilon^\alpha|X(s)-Y(s)| - 2\upsilon^{\gamma-1}|Y(s)-X(s)|^2 + C_\sigma\upsilon^{2\beta}, \quad s \le \theta. \tag{C.4}$$

By the Young's inequality,

$$C_a\upsilon^\alpha|X(s)-Y(s)| \le \frac{C_a}{2}\big(\upsilon^{\alpha+1} + \upsilon^{\alpha-1}|X(s)-Y(s)|^2\big).$$

On the other hand, since $\gamma < \alpha$, we can choose $\upsilon_0 > 0$ such that

$$\frac{C_a}{2}\upsilon^{\alpha-1} - 2\upsilon^{\gamma-1} \le -\upsilon^{\gamma-1}, \quad \upsilon \in (0, \upsilon_0].$$

In what follows, we consider $\upsilon \in (0, \upsilon_0]$ only. For such υ, we get by (C.3), (C.4)

$$A(s) \le -\upsilon^{\gamma-1}|X(s)-Y(s)|^2 + C\big(\upsilon^{\alpha+1} + \upsilon^{2\beta}\big), \quad s \le \theta. \tag{C.5}$$

Further, we have

$$|\Sigma(s)| \le C\upsilon^{1+\beta}, \quad s \le \theta. \tag{C.6}$$

Now we apply Lemma B.1 with a fixed $T > 0$ and

$$V(t) = \left|X(t) - Y(t)\right|^2, \quad \varsigma = \theta \wedge T.$$

By (C.5), (C.6) the assumptions of Lemma B.1 hold with

$$\lambda = \upsilon^{\gamma-1}, \qquad A = C\left(\upsilon^{\alpha+1} + \upsilon^{2\beta}\right), \qquad B = C\upsilon^{2+2\beta}.$$

Recall that $\gamma < \alpha$ and $\gamma < 2\beta - 1$, hence there exists $\chi > 0$ such that

$$A\lambda^{-1} = C\left(\upsilon^{\alpha+2-\gamma} + \upsilon^{2\beta+1-\gamma}\right) \le \frac{1}{2}\upsilon^{2+\chi}, \quad \upsilon \in (0, \upsilon_0] \tag{C.7}$$

for $\upsilon_0 > 0$ small enough. Next, we have

$$B^{1/2}\lambda^{-\delta} = C\upsilon^{1+\beta+\delta(1-\gamma)},$$

and

$$1 + \beta + \frac{1}{2}(1-\gamma) = 2 + \frac{2\beta - 1 - \gamma}{2} > 2.$$

That is, we can fix $\delta \in (0, 1/2$ close enough to $1/2$ and then choose $\chi > 0$ small enough such that, in addition to (C.7),

$$B^{1/2}\lambda^{-\delta} \le \frac{1}{2}\upsilon^{2+2\chi}, \quad \upsilon \in [0, \upsilon_0]$$

for $\upsilon_0 > 0$ small enough. With A, B specified above and $R = \upsilon^{-\chi}$, denote

$$\Omega_\upsilon = \left\{\omega : \sup_{t \le \theta\wedge T}\left(\left|X(t) - Y(t)\right|^2 - e^{-\lambda t}\left|X(0) - Y(0)\right|^2\right) \le A\lambda^{-1} + B^{1/2}\lambda^{-\delta}R\right\}.$$

Then, by Lemma B.1, we have

$$P(\Omega \setminus \Omega_\upsilon) \le C_1 e^{-C_2\upsilon^{-2\chi}}. \tag{C.8}$$

On the other hand, taking $\upsilon_0 \le 1$ we have for $\upsilon \in (0, \upsilon_0]$

$$\upsilon^\chi \le 1.$$

Then, on the set Ω_υ, we have (for $\upsilon \in (0, \upsilon_0]$ and υ_0 small enough)

$$\left|X(t) - Y(t)\right|^2 \le e^{-\lambda t}\left|X(0) - Y(0)\right|^2 + \frac{1}{2}\upsilon^{2+\chi} + \frac{1}{2}\upsilon^{2+2\chi}\upsilon^{-\chi}, \quad t \le \theta \wedge T.$$

Recall that $|X(0) - Y(0)|^2 \le \upsilon^2$ and $\tau = \inf\{s : |X(s) - Y(s)|^2 > 2\upsilon\}$. Since

$$e^{-\lambda t}\left|X(0) - Y(0)\right|^2 + \frac{1}{2}\upsilon^{2+\chi} + \frac{1}{2}\upsilon^{2+2\chi}\upsilon^{-\chi} \le \upsilon^2\left(1 + \upsilon^\chi\right) < 4\upsilon^2$$

and X_t, Y_t have continuous trajectories, this means that on Ω_υ we have $\tau > \theta \wedge T$ and, therefore,

$$\theta \wedge T = \theta_K \wedge T.$$

Thus

$$\left\{\sup_{t \le \theta_K\wedge T}\left(\left|X(t) - Y(t)\right|^2 - e^{-\upsilon^{\gamma-1}t}\|x - y\|^2\right) \ge \upsilon^{2+\chi}\right\} \subset \Omega \setminus \Omega_\upsilon,$$

which together with (C.8) yields the required bound. □

PROOF OF PROPOSITION 5.2. By Itô's formula,

$$\left|X^{\lambda,x}(t)\right|^2 = |x|^2 + \int_0^t A(s)\,\mathrm{d}s + \int_0^t \Sigma(s)\,\mathrm{d}W(s), \quad t \ge 0, \tag{C.9}$$

with

$$A(s) = 2\big(a\big(\mathbf{X}_s^{\lambda,x}\big), X^{\lambda,x}(s)\big) + \big\|\sigma\big(\mathbf{X}_s^{\lambda,x}\big)\big\|^2 - 2\lambda\big|X^{\lambda,x}(s)\big|^2,$$
$$\Sigma(s) = 2\big(X^{\lambda,x}(s)\big)^\top \sigma\big(\mathbf{X}_s^{\lambda,x}\big).$$

Consider first the case when set K is bounded; then by (2.5), (2.6),

$$A(s) \le C + C\big\|\mathbf{X}_s^{\lambda,x}\big\|^2 - 2\lambda\big|X^{\lambda,x}(s)\big|^2, \qquad \Sigma(s) \le C + C\big\|\mathbf{X}_s^{\lambda,x}\big\|^2. \tag{C.10}$$

Define $S_K = \sup_{x\in K} \|x\|$ and take $\varsigma = \inf\{s : |X^{\lambda,x}(s)| \ge 2S_K\}$. Then

$$A(s) \le C + 4CS_K^2 - 2\lambda\big|X^{\lambda,x}(s)\big|^2, \qquad \Sigma(s) \le C + 4CS_K^2, \quad s \le \varsigma.$$

Denote $C + 4S_K^2 = C_K$, then by Lemma B.1 with $\delta = \frac{1}{4}$, $R = \lambda^{1/8}$, $T = h$ we get

$$\mathbb{P}\Big(\sup_{t\le\varsigma\wedge h}\big(\big|X^{\lambda,x}(t)\big|^2 - e^{-2\lambda t}\big|x(0)\big|^2\big) \ge C_K(2\lambda)^{-1} + C_K^{1/2}(2\lambda)^{-1/8}\Big)$$
$$\le C_1 e^{-C_2\lambda^{1/4}}, \quad \lambda > 0.$$

Taking $\lambda > 0$ large enough, we can guarantee that

$$C_K(2\lambda)^{-1} + C_K^{1/2}(2\lambda)^{-1/8} \le S_K^2,$$

which, similar to the previous proof, yields that $\varsigma \ge h$ on the set

$$\Big\{\sup_{t\le\varsigma\wedge h}\big(\big|X^{\lambda,x}(t)\big|^2 - e^{-2\lambda t}\big|x(0)\big|^2\big) < C_K(2\lambda)^{-1} + C_K^{1/2}(2\lambda)^{-1/8}\Big\},$$

and thus

$$\mathbb{P}\big(\big\|\mathbf{X}_h^{\lambda,x}\big\|^2 \ge e^{-2\lambda(h-r)}S_K^2 + C_K(2\lambda)^{-1} + C_K^{1/2}(2\lambda)^{-1/8}\big)$$
$$\le C_1 e^{-C_2\lambda^{1/4}}, \quad \lambda > 0, x \in K.$$

Taking $\lambda > 0$ large enough, we get the required inequality (5.13).

In the second case $K = H_c$, the proof is similar and actually simpler, because instead of (C.10) we have by (2.12), (2.13) the inequalities

$$A(s) \le C + C\big|X^{\lambda,x}(s)\big|^2 - 2\lambda\big|X^{\lambda,x}(s)\big|^2, \qquad \Sigma(s) \le C + C\big|X^{\lambda,x}(s)\big|^2,$$

which do not involve the segment process $\mathbf{X}^{\lambda,x}$. Taking $S_K = \sup_{x\in K}|x(0)|$ and repeating the above estimates literally, we get the required statement in the second case $K = H_c$. □

Acknowledgments. The authors are grateful to René Schilling for helpful discussions; the first version of Lemma B.1 was actually suggested by him. The work on the project has been finished during the visit of AK to the Technical University of Dresden (Germany); AK is very grateful to the Technical University of Dresden and especially to René Schilling for their support and hospitality.

The first author is supported in part by National Science Centre, Poland, grant no. 2019/33/B/ST1/02923.

The second author is supported in part by DFG Research Unit FOR 2402.

REFERENCES

[1] ABOURASHCHI, N. and VERETENNIKOV, A. Y. (2010). On stochastic averaging and mixing. *Theory Stoch. Process.* **16** 111–129. MR2779833

[2] BESSAIH, H., KAPICA, R. and SZAREK, T. (2014). Criterion on stability for Markov processes applied to a model with jumps. *Semigroup Forum* **88** 76–92. MR3164152 https://doi.org/10.1007/s00233-013-9503-x

[3] BISMUT, J.-M. (1981). Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56** 469–505. MR0621660 https://doi.org/10.1007/BF00531428

[4] BODNARCHUK, S. V. and KULIK, O. M. (2009). Conditions for the existence and smoothness of a density of distribution for Ornstein–Uhlenbeck processes with Lévy noise. *Theory Probab. Math. Statist.* **79** 23–38.

[5] BUTKOVSKY, O. (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** 526–552. MR3178490 https://doi.org/10.1214/13-AAP922

[6] BUTKOVSKY, O., KULIK, A. and SCHEUTZOW, M. (2020). Generalized couplings and ergodic rates for SPDEs and other Markov models. *Ann. Appl. Probab.* **30** 1–39. MR4068305 https://doi.org/10.1214/19-AAP1485

[7] BUTKOVSKY, O. and SCHEUTZOW, M. (2017). Invariant measures for stochastic functional differential equations. *Electron. J. Probab.* **22** Paper No. 98, 23. MR3724566 https://doi.org/10.1214/17-EJP122

[8] CERRAI, S. (1999). Differentiability of Markov semigroups for stochastic reaction-diffusion equations and applications to control. *Stochastic Process. Appl.* **83** 15–37. MR1705598 https://doi.org/10.1016/S0304-4149(99)00014-9

[9] DA PRATO, G. and ZABCZYK, J. (1992). *Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and Its Applications* **44**. Cambridge Univ. Press, Cambridge. MR1207136 https://doi.org/10.1017/CBO9780511666223

[10] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[11] DURMUS, A., FORT, G. and MOULINES, É. (2016). Subgeometric rates of convergence in Wasserstein distance for Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1799–1822. MR3573296 https://doi.org/10.1214/15-AIHP699

[12] ELWORTHY, K. D. and LI, X.-M. (1994). Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** 252–286. MR1297021 https://doi.org/10.1006/jfan.1994.1124

[13] ES-SARHIR, A., VON RENESSE, M.-K. and SCHEUTZOW, M. (2009). Harnack inequality for functional SDEs with bounded memory. *Electron. Commun. Probab.* **14** 560–565. MR2570679 https://doi.org/10.1214/ECP.v14-1513

[14] GIKHMAN, I. and SKOROHOD, A. V. (1972). *Stochastic Differential Equations*. Springer, New York. MR0346904

[15] HAIRER, M. (2002). Exponential mixing properties of stochastic PDEs through asymptotic coupling. *Probab. Theory Related Fields* **124** 345–380. MR1939651 https://doi.org/10.1007/s004400200216

[16] HAIRER, M. and MATTINGLY, J. C. (2006). Ergodicity of the 2D Navier–Stokes equations with degenerate stochastic forcing. *Ann. of Math.* (2) **164** 993–1032. MR2259251 https://doi.org/10.4007/annals.2006.164.993

[17] HAIRER, M., MATTINGLY, J. C. and SCHEUTZOW, M. (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields* **149** 223–259. MR2773030 https://doi.org/10.1007/s00440-009-0250-6

[18] KOMOROWSKI, T., PESZAT, S. and SZAREK, T. (2010). On ergodicity of some Markov processes. *Ann. Probab.* **38** 1401–1443. MR2663632 https://doi.org/10.1214/09-AOP513

[19] KULIK, A. (2018). *Ergodic Behavior of Markov Processes: With Applications to Limit Theorems. De Gruyter Studies in Mathematics* **67**. de Gruyter, Berlin. MR3791835

[20] LIPTSER, R. S. and SHIRYAYEV, A. N. (2013). *Statistics of Random Processes. I*. Springer, Berlin.

[21] MOHAMMED, S. E. A. (1984). *Stochastic Functional Differential Equations. Research Notes in Mathematics* **99**. Pitman (Advanced Publishing Program), Boston, MA. MR0754561

[22] PAGÈS, G. and PANLOUP, F. (2014). A mixed-step algorithm for the approximation of the stationary regime of a diffusion. *Stochastic Process. Appl.* **124** 522–565. MR3131304 https://doi.org/10.1016/j.spa.2013.07.011

[23] PARDOUX, È. and VERETENNIKOV, A. Y. (2003). On Poisson equation and diffusion approximation. II. *Ann. Probab.* **31** 1166–1192. MR1988467 https://doi.org/10.1214/aop/1055425774

[24] PESZAT, S. and ZABCZYK, J. (2007). *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach. Encyclopedia of Mathematics and Its Applications* **113**. Cambridge Univ. Press, Cambridge. MR2356959 https://doi.org/10.1017/CBO9780511721373

[25] REVUZ, D. and YOR, M. (2005). *Continuous Martingales and Brownian Motion*, 3rd ed. Springer, Berlin.
[26] STROOCK, D. W. and VARADHAN, S. R. S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **233**. Springer, Berlin. MR0532498
[27] THORISSON, H. (2000). *Coupling*, *Stationarity*, *and Regeneration*. *Probability and Its Applications* (*New York*). Springer, New York. MR1741181 https://doi.org/10.1007/978-1-4612-1236-2
[28] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. *Springer Series in Statistics*. Springer, New York. MR2724359 https://doi.org/10.1007/b13794
[29] VON RENESSE, M.-K. and SCHEUTZOW, M. (2010). Existence and uniqueness of solutions of stochastic functional differential equations. *Random Oper*. *Stoch*. *Equ*. **18** 267–284. MR2718125 https://doi.org/10.1515/ROSE.2010.015
[30] WANG, F.-Y. (2011). Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on nonconvex manifolds. *Ann*. *Probab*. **39** 1449–1467. MR2857246 https://doi.org/10.1214/10-AOP600