IMA Journal of Numerical Analysis (2024) **00**, 1–95
https://doi.org/10.1093/imanum/drae043

Long-term accuracy of numerical approximations of SPDEs with the stochastic Navier–Stokes equations as a paradigm

Nathan E. Glatt-Holtz
Department of Statistics, Indiana University, Bloomington, IN, 47405, USA

and

Cecilia F. Mondaini*
Department of Mathematics, Drexel University, Philadelphia, PA, 19104, USA
*Corresponding author: cf823@drexel.edu

[Received on 5 February 2023; revised on 3 May 2024]

This work introduces a general framework for establishing the long time accuracy for approximations of Markovian dynamical systems on separable Banach spaces. Our results illuminate the role that a certain uniformity in Wasserstein contraction rates for the approximating dynamics bears on long time accuracy estimates. In particular, our approach yields weak consistency bounds on $\mathbb{R}^+$ while providing a means to sidestepping a commonly occurring situation where certain higher order moment bounds are unavailable for the approximating dynamics. Additionally, to facilitate the analytical core of our approach, we develop a refinement of certain 'weak Harris theorems'. This extension expands the scope of applicability of such Wasserstein contraction estimates to a variety of interesting stochastic partial differential equation examples involving weaker dissipation or stronger nonlinearity than would be covered by the existing literature. As a guiding and paradigmatic example, we apply our formalism to the stochastic 2D Navier–Stokes equations and to a semi-implicit in time and spectral Galerkin in space numerical approximation of this system. In the case of a numerical approximation, we establish quantitative estimates on the approximation of invariant measures as well as prove weak consistency on $\mathbb{R}^+$. To develop these numerical analysis results, we provide a refinement of L^2_x accuracy bounds in comparison to the existing literature, which are results of independent interest.

Keywords: long time accuracy; weak Harris theorems; contraction in Wasserstein distance; numerical analysis of stochastic partial differential equations; stochastic Navier–Stokes equations.

1. Introduction

Questions concerning long time accuracy under approximations for dynamical systems exhibiting chaotic behavior are notoriously difficult. This is nevertheless a topic of wide interest particularly given that statistical theories of turbulence in fluid dynamics can be framed in terms of observables against invariant measures. According to this widely used paradigm such measures are connected to the fundamental governing equations through (putative) ergodic averages and thus may be regarded as containers for statistically robust properties of turbulent flows. Thus, from this point of view, it is natural to ask if the essential features of these invariant measures are maintained under suitable numerical approximations or in a variety of physically interesting singular parameter limits.

Unfortunately, the robustness of statistical properties, i.e. the verification of an ergodic hypothesis, for solutions of deterministic models such as the Navier–Stokes equations and its many variations are typically far from the reach of rigorous analysis. On the other hand, certain stochastic versions of these equations are more tractable to analyze in this regard. Moreover, such stochastic models often retain physical relevance while providing an important motivation and a set of unique challenges that have been driving a flurry of developments in the ergodic and mixing theory of infinite dimensional Markov processes in recent decades. In this stochastic setting, the question of the stability of long time statistical properties as a function of model parameters is therefore of broad interest for a diverse variety of nonlinear, infinite dimensional, randomly stirred systems.

This work develops a novel framework for addressing such long time stability and accuracy questions for parameter dependent Markov processes on a Polish space. Our approach leverages a certain uniform Wasserstein contraction condition (a strong form of exponential mixing), which, as we will illustrate on several paradigmatic examples, has a rather broad scope of applicability for finite and infinite dimensional stochastic systems. Our results demonstrate that appropriately leveraging uniform contraction provides an important twist on an existing vein of research concerning infinite time stability under parameter perturbation for certain stochastic systems (Shardlow & Stuart, 2000; Kuksin & Shirikyan, 2003; Hairer & Mattingly, 2008; Hairer & Majda, 2010; Mattingly *et al.*, 2010; Hairer *et al.*, 2011; Foldes *et al.*, 2017; Johndrow & Mattingly, 2017; Földes *et al.*, 2019; Cerrai & Glatt-Holtz, 2020). Here we also note that the framework in Wang (2010); Gottlieb *et al.* (2012) for deterministic dynamical systems anticipate some of the developments here, including the invocation of a uniform dissipativity condition. However, the scope of Wang (2010); Gottlieb *et al.* (2012) is fundamentally limited by its inability to rule out nonuniqueness (let alone address ergodic and mixing properties) for the long term statistics of the infinite dimensional deterministic models considered therein.

As an important technical foundation to carry out our broad program, we develop a refinement of the so-called 'weak Harris approach' to exponential mixing. This portion of our contribution builds on the seminal works (Hairer & Mattingly, 2008; Hairer *et al.*, 2011), which lay out a powerful framework for addressing Wasserstein contraction. These earlier works make use of delicate norm constructions that sidestep the need to Byzantine explicit coupling constructions. On the other hand, the representative and natural selection of examples presented in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020) demonstrates the need to refine the approach for the typical situation where models lack certain higher order moment estimates or possess a weaker form of smoothing at small scales or both.

A primary domain of application for our framework regards the error analysis for numerical approximations of certain stochastic partial differential equations (SPDEs). This is an area of applied analysis that has undergone some rapid development in the past decade; see for example Mattingly *et al.* (2002); Jentzen & Kloeden (2009); Carelli & Prohl (2012); Brzeźniak *et al.* (2013); Bessaih & Millet (2019, 2021) and containing references. Thus, to illustrate the scope of our approach on a paradigmatic example, we carry out a case study of the space-time numerical approximation of the stochastic 2D Navier–Stokes equations given by a spectral Galerkin discretization in space and a semi-implicit Euler time discretization. In the course of our analysis, we provide some novel approximation bounds and some significant refinements of existing finite time error bounds in comparison to the existing literature (Carelli & Prohl, 2012; Bessaih & Millet, 2019, 2021), which are of independent interest. Note that our general framework has also been useful for several concurrent projects. In a recent work by the first author, Glatt-Holtz *et al.* (2022b), we make use of uniform contractivity to address certain singular limit problems concerning SPDEs with diffusive memory terms. Elsewhere in Glatt-Holtz *et al.* (2022a), we address application in bias estimation for statistical sampling algorithms.

1.1 *The Uniform Contraction Framework for Long Time Stability*

Let us now give an overview of the abstract foundation of our approach. We provide an idealized version here so that the reader can observe the underlying simplicity of our framework. Of course, to carry out our program in practice we will need to impose a number of technical assumptions; we refer the reader to Theorem 2.5, Theorem 2.8 and Corollary 2.11 below for these more involved formulations.

Our departure point is to observe that, for certain stochastic Markovian systems, long time accuracy estimates can be developed in the presence of a strong type of mixing taking the form of a contraction estimate in a suitable Wasserstein distance. Note that such contraction estimates have been previously exploited in a variety of specific contexts for SDEs and SPDEs and other Markovian processes (Hairer & Mattingly, 2008; Hairer & Majda, 2010; Hairer *et al.*, 2011; Foldes *et al.*, 2017; Johndrow & Mattingly, 2017; Földes *et al.*, 2019; Cerrai & Glatt-Holtz, 2020) to provide rigorous bounds on parameter dependent invariant measures. The crucial new element here centers on suitably exploiting parameter independent uniformity in the contraction rates.

Suppose that $\{P_t^\theta\}_{t\geq 0}$ is a collection of Markov transition operators defined on a Polish space (X,ρ) parameterized by $\theta\in\Theta$. These operators act on Borel probability measures ν and observables φ as

$$\nu P_t^\theta(\mathrm{d}u) := \int P_t^\theta(v,\mathrm{d}u)\nu(\mathrm{d}v), \qquad P_t^\theta\varphi(u) := \int \varphi(v)P_t^\theta(u,\mathrm{d}v),$$

respectively. Let us suppose that for some θ_0, corresponding to the 'true' or 'limiting' dynamics of interest, we have Wasserstein contraction. Namely, for any $t\geq 0$

$$\mathscr{W}\left(\mu P_t^{\theta_0}, \tilde{\mu} P_t^{\theta_0}\right) \leq C_0\, e^{-\kappa t}\mathscr{W}(\mu,\tilde{\mu}), \tag{1.1}$$

for any Borel probability measure $\mu,\tilde{\mu}$, where $C_0,\kappa>0$ are constants independent of $\mu,\tilde{\mu}$ and $t\geq 0$. Here, as in e.g. Villani (2008), $\mathscr{W}$ is the Wasserstein distance corresponding to ρ, i.e.

$$\mathscr{W}(\nu_1,\nu_2) = \inf_{\Gamma\in\mathscr{C}(\nu_1,\nu_2)}\int \rho(u,\tilde{u})\Gamma(\mathrm{d}u,\mathrm{d}\tilde{u}), \tag{1.2}$$

with $\mathscr{C}(\nu_1,\nu_2)$ denoting all of the couplings of ν_1 and ν_2. Note that such Wasserstein contraction estimates can be obtained using the so-called 'weak Harris approach' developed in Hairer & Mattingly (2008); Hairer *et al.* (2011), which we refine for our purposes here in Theorem 2.1 below.

Suppose now that for every $\theta\in\Theta$ we have a corresponding measure μ_θ, which is invariant under $\{P_t^\theta\}_{t\geq 0}$, namely $\mu_\theta P_t^\theta = \mu_\theta$ for any $t\geq 0$. We then make the following simple observation. Exploiting invariance, the triangle inequality and the contraction estimate (1.1), we have

$$\begin{aligned}\mathscr{W}(\mu_{\theta_0},\mu_\theta) = \mathscr{W}\left(\mu_{\theta_0}P_t^{\theta_0},\mu_\theta P_t^\theta\right) &\leq \mathscr{W}\left(\mu_{\theta_0}P_t^{\theta_0},\mu_\theta P_t^{\theta_0}\right) + \mathscr{W}\left(\mu_\theta P_t^{\theta_0},\mu_\theta P_t^\theta\right)\\ &\leq C_0\, e^{-\kappa t}\mathscr{W}(\mu_{\theta_0},\mu_\theta) + \mathscr{W}\left(\mu_\theta P_t^{\theta_0},\mu_\theta P_t^\theta\right),\end{aligned} \tag{1.3}$$

which holds for any $\theta \in \Theta$ and any $t \geq 0$. Thus, by selecting t_* such that, say, $C_0 e^{-\kappa t_*} \leq 1/2$, we can rearrange the above expression and obtain

$$\mathscr{W}(\mu_{\theta_0}, \mu_\theta) \leq 2\mathscr{W}\left(\mu_\theta P^{\theta_0}_{t*}, \mu_\theta P^{\theta}_{t*}\right). \tag{1.4}$$

Thus, we obtain a bound that reduces the question of long time accuracy in the sense of invariant statistics to a certain finite time error estimate and alongside suitable θ-uniform moment bound on μ_θ.

To make this significance of (1.4) a bit more concrete, we recall that $\mathscr{W}$ possesses a desirable Lipschitz structure. For example, if for each $u \in X$, we can find a coupling, $u_{\theta_0}(t,u), u_\theta(t,u)$ of $P^{\theta_0}_t(u,\cdot), P^\theta_t(u,\cdot)$ such that

$$\mathbb{E}\rho(u_{\theta_0}(t,u), u_\theta(t,u)) \leq e^{\tilde{C}_0 t} f(u) g_{\theta_0}(\theta), \tag{1.5}$$

where g_{θ_0} is a (bounded) function on Θ, then basic properties of $\mathscr{W}$ lead to

$$\mathscr{W}\left(\mu_\theta P^{\theta_0}_{t*}, \mu_\theta P^{\theta}_{t*}\right) \leq e^{\tilde{C}_0 t_*} g_{\theta_0}(\theta) \int f(u) \mu_\theta(\mathrm{d}u). \tag{1.6}$$

Hence, we obtain from (1.4) that

$$\mathscr{W}(\mu_{\theta_0}, \mu_\theta) \leq 2 e^{\tilde{C}_0 t_*} g_{\theta_0}(\theta) \int f(u) \mu_\theta(\mathrm{d}u). \tag{1.7}$$

Of course obtaining (1.1) and then providing suitable qualitative estimates for $\mathscr{W}(\mu_\theta P^{\theta_0}_{t*}, \mu_\theta P^{\theta}_{t*})$ to leverage via (1.4) as in (1.5)–(1.7) represents a bespoke and nontrivial mathematical challenge for each of specific works mentioned previously, Hairer & Mattingly (2008); Hairer *et al.* (2011); Foldes *et al.* (2017); Johndrow & Mattingly (2017); Földes *et al.* (2019); Cerrai & Glatt-Holtz (2020). Furthermore, we emphasize for what follows that in order to exploit (1.7) we must obtain a uniform bound on $\int f(u)\mu_\theta(\mathrm{d}u)$ as a function of θ.

This work develops a different and seemingly novel variation on the reduction in (1.1), (1.4). Suppose that, instead of (1.1), we impose the stronger uniform contraction assumption

$$\mathscr{W}\left(\mu P^\theta_t, \tilde{\mu} P^\theta_t\right) \leq C e^{-\kappa t} \mathscr{W}(\mu, \tilde{\mu}), \tag{1.8}$$

where, to emphasize, the constants $C, \kappa > 0$ are now supposed to be independent of the parameter $\theta \in \Theta$. In comparison to (1.3), we now proceed as

$$\mathscr{W}(\mu_{\theta_0}, \mu_\theta) \leq \mathscr{W}\left(\mu_{\theta_0} P^{\theta_0}_t, \mu_{\theta_0} P^\theta_t\right) + \mathscr{W}\left(\mu_{\theta_0} P^\theta_t, \mu_\theta P^\theta_t\right) \leq \mathscr{W}\left(\mu_{\theta_0} P^{\theta_0}_t, \mu_{\theta_0} P^\theta_t\right) + C_0 e^{-\kappa t} \mathscr{W}(\mu_{\theta_0}, \mu_\theta), \tag{1.9}$$

so that, by again choosing t_* such that

$$C_0 e^{-\kappa t_*} \leq 1/2, \tag{1.10}$$

we now find

$$\mathscr{W}(\mu_{\theta_0}, \mu_\theta) \leq 2\mathscr{W}\left(\mu_{\theta_0} P^{\theta_0}_{t*}, \mu_{\theta_0} P^{\theta}_{t*}\right). \tag{1.11}$$

This seemingly innocent difference in comparison to (1.4) trades uniformity in the contraction rate for a single moment bound on the limit system. Indeed, under (1.5), we obtain

$$\mathscr{W}(\mu_{\theta_0}, \mu_\theta) \leq 2e^{\tilde{C}_0 t_*} g_{\theta_0}(\theta) \int f(u)\mu_{\theta_0}(\mathrm{d}u), \tag{1.12}$$

so that we trade the requirement (1.8) for the uniform bound $\sup_{\theta\in\Theta}\int f(u)\mu_\theta(\mathrm{d}u)$.

This difference between (1.12) and (1.7) turns out to sometimes be an indispensable trade off. We exploit it for the questions of numerical accuracy we consider here as well as other situations of interest as in the concurrent work (Glatt-Holtz *et al.*, 2022b). Specifically, as we will describe in further detail immediately below, for our applications here $\theta \neq \theta_0$ represents a numerical approximation parameter for the stochastic Navier–Stokes equations. These numerical approximations destroy (or complicate) certain crucial Lyapunov structures, namely we lack the availability of moments for μ_θ when $\theta \neq \theta_0$ as would be needed in (1.7). In any case we refer to Theorem 2.5, which is framed in a context applicable to a slightly weaker form of the uniform contraction estimates (1.8) required for our applications.

Leaving this consideration aside, the uniform contraction assumption (1.8) combined with finite time error estimate bounds as in (1.5) leads to other desirable long time approximation estimates. Indeed with (1.8) and invoking invariance we obtain the bound

$$\begin{aligned}\mathscr{W}\left(\mu P^\theta_t, \mu P^{\theta_0}_t\right) &\leq \mathscr{W}\left(\mu P^\theta_t, \mu_\theta P^\theta_t\right) + \mathscr{W}(\mu_\theta, \mu_{\theta_0}) + \mathscr{W}\left(\mu_{\theta_0} P^{\theta_0}_t, \mu P^{\theta_0}_t\right)\\ &\leq C_0\, e^{-\kappa t}\Big(\mathscr{W}(\mu, \mu_\theta) + \mathscr{W}(\mu, \mu_{\theta_0})\Big) + \mathscr{W}\left(\mu_\theta, \mu_{\theta_0}\right)\\ &\leq C_0\, e^{-\kappa t}\mathscr{W}\left(\mu, \mu_{\theta_0}\right) + (1 + C_0)\mathscr{W}\left(\mu_\theta, \mu_{\theta_0}\right),\end{aligned}$$

for any 'initial' distribution μ. Hence, with this bound and (1.12), valid under (1.4) and (1.5), we obtain

$$\begin{aligned}&\mathscr{W}\left(\mu P^\theta_t, \mu P^{\theta_0}_t\right)\\ &\quad\leq \min\left\{C_0\, e^{-\kappa t}\mathscr{W}\left(\mu, \mu_{\theta_0}\right) + 2(1 + C_0)\, e^{\tilde{C}_0 t_*}\int f(u)\mu_{\theta_0}(\mathrm{d}u)\, g_{\theta_0}(\theta), e^{\tilde{C}_0 t}\int f(u)\mu(\mathrm{d}u)\, g_{\theta_0}(\theta)\right\},\end{aligned}$$

valid for any $t \geq 0$, where we recall that $t^* > 0$ is given as in (1.10). Hence, optimizing appropriately over $t \geq 0$ in this bound we conclude

$$\sup_{t\geq 0}\mathscr{W}\left(\mu P^\theta_t, \mu P^{\theta_0}_t\right) \leq C\left(\mathscr{W}(\mu, \mu_{\theta_0}) + \int f(u)\mu_{\theta_0}(\mathrm{d}u) + \int f(u)\mu(\mathrm{d}u)\right) g_{\theta_0}(\theta)^\eta, \tag{1.13}$$

where $C, \eta > 0$ are independent of μ and θ.

Note that, in the numerical analysis context of interest here, this bound, (1.13), immediately yields a weak order approximation estimate valid on the entire time interval $[0, \infty)$. The operational version of (1.13) formulated in order to address our nonlinear SPDE applications is formulated in Theorem 2.8 and in Corollary 2.11, which make explicit the connection with weak order convergence in stochastic numerical analysis.

1.2 *Contributions to the weak Harris approach for Wasserstein contraction*

Of course the elegant simplicity of the above discussion obscures a number of bedeviling technical challenges that one must address in order to carry out this program in practice. One challenge is to establish (uniform) Wasserstein contraction bounds as in (1.1) and in (1.8). For this purpose that we develop Theorem 2.1 below, which provides general criteria for such contraction estimates. This is a result that has independent interest for a variety of infinite dimensional contexts as highlighted by the recent contributions (Glatt-Holtz *et al.*, 2017; Butkovsky *et al.*, 2020; Glatt-Holtz *et al.*, 2021).

As previously mentioned, Theorem 2.1 provides a new variation on the so-called 'weak Harris approach' developed in Hairer & Mattingly (2008); Hairer *et al.* (2011). This weak Harris approach builds on a wide and well developed literature on mixing rates for Markov chains; see e.g. Da Prato & Zabczyk (1996); Meyn & Tweedie (2009); Douc *et al.* (2018); Kulik (2018) for a systematic presentation. The classical Harris theorems date back to the 1950s by building on Doeblin's coupling approach to address mixing in unbounded phase spaces. The key is to appropriately incorporate the role of Lyapunov structure to facilitate coupling at 'large scales'. Typically, in this literature mixing occurs in a total variation (TV) topology or other related 'strong topologies' on probability measures; see Hairer & Mattingly (2011b) for a recent treatment close to our setting. This use of a total variation topology highlights a limitation of the classical Harris approach: it turns out to be ill-adapted to infinite dimensional contexts where measures tend to be mutually singular as exemplified by the Feldman–Hajek theorem (see e.g. Da Prato & Zabczyk, 2014, Theorem 2.25).

More recent variations on this Harris approach, largely developed in an extended body of literature in the SPDE context starting from Bricmont *et al.* (2002); Hairer (2002); Kuksin & Shirikyan (2002); Mattingly (2002), address mixing in Wasserstein (or the closely related dual-Lipschitz) distance. Wasserstein distance reflects a weak topology that sidesteps the issue of mutually singular laws arising in infinite dimensional stochastic systems. A distinguished contribution of the works (Hairer & Mattingly, 2008; Hairer *et al.*, 2011) in this literature is to provide a contraction (or a so-called 'spectral-gap') estimate as in (1.1) as suits our needs here. As in the earlier literature, these results are based on natural conditions leading to couplings that synchronize two point dynamics at large, intermediate and small scales, through Lyapunov structure, irreducibility and smoothing properties, respectively. However, an elegant feature of Hairer & Mattingly (2008); Hairer *et al.* (2011) in this wider mixing literature is the identification of a particular class of metrics (or pseudo-metrics) on the phase space that are carefully tailored to account for the three different mechanisms acting at different scales that drive the coupling. This 'norm approach' thus avoids Byzantine explicit coupling constructions yielding a flexible approach for applications while producing elegant, transparent proofs.

The approach (Hairer & Mattingly, 2008; Hairer *et al.*, 2011) is well adapted to the 2D randomly forced Navier–Stokes on compact domains in the absence of boundaries and as well as several other reaction-diffusion type models of interest. However, a crucial requirement in Hairer & Mattingly (2008); Hairer *et al.* (2011) appears to be stronger than can be expected for a rich variety of interesting SPDE examples involving weaker dissipation and/or stronger nonlinearity as illustrated in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020); Glatt-Holtz *et al.* (2021), e.g. the 2D Navier–Stokes equations (NSE) on a bounded domain, the 2D hydrostatic NSE, the fractionally dissipative Euler model, the damped Euler–Voigt equations, a damped nonlinear wave equation and the damped Korteweg–De Vries equation, under suitable stochastic forcing terms and boundary conditions. Indeed, Hairer & Mattingly (2008) develops their theory around a certain geodesic metric that is adapted to a quadratic exponential Lyapunov structure, namely $V(u) = \exp(\alpha|u|^2)$ for some $\alpha > 0$. It turns out that this geodesic metric approach involves the use of a certain gradient bound on the Markovian dynamics closely related to the so-called 'asymptotic strong Feller' (ASF) condition introduced earlier in Hairer & Mattingly (2006). While the pseudo-metric structures considered later in Hairer *et al.* (2011) are in various ways more flexible, including in terms of its requirement on the Lyapunov structure, Hairer *et al.* (2011) still ultimately relies on these same ASF type gradient bound on the Markov semigroup. To summarize, the existing works (Hairer & Mattingly, 2008; Hairer *et al.*, 2011) require a significant degree of uniformity across the phase space in contraction rates when two point dynamics are in close proximity. This is rather more than can be hoped for in a variety of interesting situations.

Theorem 2.1 provides our new take on the weak Harris approach. Its main advantage over these previous formulations consists in sidestepping the need for a gradient bound. Our approach builds on machinery introduced recently in Butkovsky *et al.* (2020), which provides a powerful and user friendly toolbox for addressing exponential mixing by confronting the representative gallery of SPDE examples introduced in Glatt-Holtz *et al.* (2017). Our result here may be seen to be a sort of intermediate formulation of the topology for contraction, laying between Hairer & Mattingly (2008) and Hairer *et al.* (2011), and focusing specifically on $V(u) = \exp(\alpha|u|^2)$. This intermediate formulation then has the advantage of allowing us to treat the gallery of examples from Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020). Note that our pseudo-metric does not maintain a generalized triangle inequality due to the underlying Lyapunov structure around $V(u) = \exp(\alpha|u|^2)$ as would be strictly required for bounds like (1.3), (1.9). Instead, taking advantage of a stronger 'super-Lyapunov' structure for V, we provide a 'contraction-like' condition (see (2.7), (2.8)), which is strong enough to then follow the general stream of argumentation leading to our reduction bounds (1.12), (1.13).

The list of problems in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020) include, notably, the 2D stochastic Navier–Stokes equations on a bounded domain subject to the usual nonslip boundary condition. We provide complete details regarding Wasserstein contraction for this case below in Section 4, which we believe illustrates the full significance for Theorem 2.1 in applications. We refer to Remark 4.6 below, which provides technical level comparison of our Theorem 2.1 to the results in Hairer & Mattingly (2008) and in Hairer *et al.* (2011).

1.3 *Results for the numerical approximation of the stochastic Navier–Stokes equations*

As already alluded to above, our immediate goal is to demonstrate the efficacy of the abstract formalism developed in Section 2 for the numerical analysis of certain classes of nonlinear SPDEs. As a paradigmatic model problem, we carry out a detailed study of a fully discrete numerical scheme to approximate the 2D stochastic Navier–Stokes equations (SNSE) in the presence of spatially smooth, but sufficiently rich (or more precisely 'essentially elliptic' in the terminology of e.g. Mattingly, 2003) stochastic forcing structure under periodic boundary conditions.

Our main result that we preview immediately below as Theorem 1.1 adds to an extensive body of research on the numerical analysis of stochastic dynamical systems. However, the literature on long time numerical approximation for SPDEs is scant, and, to the best of our knowledge, there is no previous literature on the stochastic Navier–Stokes or other such 'strongly nonlinear' equations in this regard. To summarize, our primary contribution in comparison to the existing numerical analysis literature is to provide rigorous approximation bounds on invariant measures and to establish weak convergence estimates, à la (1.18), on an infinite time horizon for the stochastic Navier–Stokes equations. This usage of our abstract framework lays out an approach that would apply to the numerical analysis of a number of other strongly nonlinear infinite dimensional systems seemingly out of reach of the previously existing approaches, one that follow a very different set of methodologies in comparison to the extant literature.

Let us be more concrete. For our numerical application, we consider the 2D Navier–Stokes equations on the torus $\mathbb{T}^2$ so that we can work with the convenient vorticity formulation

$$d\xi + (-\nu \Delta \xi + \mathbf{u} \cdot \nabla \xi)\, dt = \sum_{k=1}^{d} \sigma_k \, dW_k = \sigma \, dW, \quad \mathbf{u} = \mathscr{K} * \xi. \tag{1.14}$$

Here $\mathscr{K}$ is the Biot–Savart operator, which uniquely recovers the divergence free vector field $\mathbf{u}$ from ξ (so that $\xi = \nabla^{\perp} \cdot \mathbf{u}$). The physical parameter $\nu > 0$ represents the kinematic viscosity of the fluid. The system is driven by a white in time and spatially smooth Gaussian process $\sigma \, dW = \sum_{k=1}^{d} \sigma_k \, dW^k$, where $W = (W_1, W_2, \ldots, W_d)$ is a collection of i.i.d. Brownian motions on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and $\sigma_1, \ldots, \sigma_d$ are (spatially) mean free elements of $L^2(\mathbb{T}^2)$. For simplicity, we consider initially spatially mean free fields $\xi_0 = \xi(0)$, a condition maintained by its evolution $\xi(t)$, $t > 0$, in (1.14) so long as the noise itself is mean free.

As numerical approximation of (1.14), we consider a spectral Galerkin discretization in space and a semi-implicit Euler time discretization, given by

$$\xi_{N,\delta}^{n} = \xi_{N,\delta}^{n-1} + \delta \Big[\nu \Delta \xi_{N,\delta}^{n} - \Pi_N \Big(\mathbf{u}_{N,\delta}^{n-1} \cdot \nabla \xi_{N,\delta}^{n} \Big) \Big] + \sqrt{\delta} \sum_{k=1}^{d} \Pi_N \sigma_k \eta_n^k, \quad \text{for } n \geq 1, \tag{1.15}$$

where the numerical discretization parameters are the size of the time step $\delta > 0$, and $N \geq 1$ the degree of spectral (spatial) approximation. The operator Π_N denotes a low-Fourier mode projector, i.e. the projection operator onto the space spanned by the first N eigenfunctions of '$-\Delta$' under periodic boundary conditions. As previously, $\mathbf{u}_{N,\delta}^{n-1} = \mathscr{K} * \xi_{N,\delta}^{n-1}$. Here $\sqrt{\delta}\eta_n^k$ have the laws of increments of the Brownian motions W^k, so that η_n^k is generated by a sequence of i.i.d. standard Gaussian random variables.

For both (1.14) and (1.15), we work under the simplifying assumption, the so-called 'essentially-elliptic case', where we suppose a certain nondegeneracy condition that noise excitation acts directly on some number of low Fourier modes, namely

$$\operatorname{span}\{\sigma_1, \ldots, \sigma_d\} \supset \Pi_K L^2(\mathbb{T}^2) \tag{1.16}$$

for $K = K(\nu, \sum_{k=1}^{d} |\sigma_k|_{L^2}^2)$, see (3.29) in Theorem 3.9 below for the precise condition on K. This condition on K is a standard assumption in the SNSE literature, cf. Mattingly (2003); Kuksin & Shirikyan (2012); Glatt-Holtz *et al.* (2017). Indeed, this noise structure is convenient because we are able to establish a discretization uniform contraction à la (1.8) as fits our formalism. We expect our results to still hold under a more spatially degenerate noise setting, where in particular K does not depend on the viscosity ν or the size of σ, but proving this conjecture would require significantly more technical effort, see Section 1.4 below.

Our main numerical result is given here in a heuristic formulation as follows. We note that the orders of convergence with respect to δ, N in (1.17) and (1.18) may not be optimal, as we discuss with more details further below within the literature review section.

THEOREM 1.1 Consider (1.14) and (1.15) under the suitable nondegeneracy condition (1.16) that the stochastic perturbation acts directly on sufficiently many low frequencies (depending only on $\nu > 0$ and $|\sigma|_{L^2}^2 = \sum_k |\sigma_k|_{L^2}^2$). Then (1.14) has a unique statistically invariant state μ_*, and (1.15) has a unique statistically invariant state $\mu_*^{N,\delta}$ for any $N \geq 1$, $\delta > 0$. Moreover, for any sufficiently regular observable $\varphi : L^2(\mathbb{T}^2) \to \mathbb{R}$, we have the bound

$$\left| \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi') - \int \varphi(\xi')\mu_*(\mathrm{d}\xi') \right| \leq C_\varphi(\delta^{r_1} + N^{-r_2}) \tag{1.17}$$

for some $r_1 = r_1(\nu, |\sigma|^2) > 0$, $r_2 = r_2(\nu, |\sigma|^2) > 0$, which do not depend on φ and where C_φ, r_1, r_2 are all δ, N-independent.

Finally, (1.15) is a weakly consistent approximation of (1.14). Namely, for any such observable φ and sufficiently regular ξ_0 it holds that

$$\sup_{n \geq 0} \left| \mathbb{E}\varphi\left(\xi_{N,\delta}^n(\xi_0)\right) - \mathbb{E}\varphi(\xi(n\delta; \xi_0)) \right| \leq C_\varphi(\delta^{\tilde{r}_1} + N^{-\tilde{r}_2}), \tag{1.18}$$

where $\xi(t; \xi_0)$, $t \geq 0$, and $\xi_{N,\delta}^n(\xi_0)$, $n \in \mathbb{N}$, denote the solutions of (1.14) and (1.15), respectively, with initial datum ξ_0. Here again, we have that $\tilde{r}_1 = \tilde{r}_1(\nu, |\sigma|^2) > 0$, $\tilde{r}_2 = \tilde{r}_2(\nu, |\sigma|^2) > 0$ are independent of φ, δ and N.

The precise and complete formulation of Theorem 1.1 is divided between Theorem 3.21 and Theorem 3.22 below. In particular, its proof is founded on a discretization-uniform contraction bound from Theorem 3.9 and the finite-time error estimates from Proposition 3.17 and Proposition 3.18, which provide concrete instantiations of (1.8) and (1.5), in addition to being contributions of independent interest for (1.14) and (1.15).

Further, we notice that (1.17) together with the contraction inequality from Theorem 3.9 can be used to derive useful error estimates for the estimator $\frac{1}{n}\sum_{k=1}^n \varphi(\xi_{N,\delta}^k)$ as an approximation of the stationary average $\int \varphi(\xi')\mu_*(\mathrm{d}\xi')$. Indeed, in Remark 3.24 below we sketch the main steps involved in the derivation of the following bias estimate

$$\left| \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n \varphi\left(\xi_{N,\delta}^k(\xi_0)\right) - \int \varphi(\xi')\mu_*(\mathrm{d}\xi') \right) \right| \leq C_\varphi \left(\frac{1}{n\delta} + \delta^{r_1} + N^{-r_2} \right), \tag{1.19}$$

and the mean-squared error estimate

$$\mathbb{E}\left[\left| \frac{1}{n}\sum_{k=1}^n \varphi\left(\xi_{N,\delta}^k\right) - \int \varphi(\xi')\mu_*(\mathrm{d}\xi') \right|^2 \right] \leq C_\varphi \left(\frac{1}{n\delta} + \delta^{2r_1} + N^{-2r_2} \right), \tag{1.20}$$

for C_φ, r_1, r_2 as in (1.17). Clearly, estimates such as these have a direct significance in practical applications where one naturally computes the time-discrete average $\frac{1}{n}\sum_{k=1}^n \varphi(\xi_{N,\delta}^k)$ for a certain number n of states as a way of approximating the average of a given observable φ with respect to the (typically unknown) underlying stationary distribution μ_*.

Previous Literature, Elements of Our Analysis. There is an extensive literature on numerical analysis of stochastic systems. Some general background on this subject in the context of SDEs can be found in e.g. Kloeden & Platen (1992); Milstein & Tretyakov (2004), and for SPDEs we refer to Jentzen & Kloeden (2009); Lord *et al.* (2014). In this community, approximation results are typically characterized in terms of 'strong' and 'weak' convergence. The former notion of strong convergence concerns, in our notations, bounds on the quantity $|\xi^j_{N,\delta}(\xi_0) - \xi(j\delta;\xi_0)|_{L^2}$ either in mean or in probability. The latter notion of 'weak' convergence involves estimates for $\mathbb{E}(\phi(\xi^j_{N,\delta}(\xi_0)) - \phi(\xi(j\delta;\xi_0)))$ over different classes of test functions, see e.g. Kloeden & Platen (1992). When the test functions are Lipschitz continuous with respect to a certain notion of distance then weak convergence can be expressed in terms of bounds on the corresponding Wasserstein distance, cf. (2.52) below. Additionally, in this latter setting strong convergence implies weak convergence, but of course not visa-versa and typically weak rates are better than strong rates (cf. Davie & Gaines, 2001; Debussche & Printems, 2009).

The available approaches for weak convergence are mainly centered around the following observation. Expanding $\mathbb{E}(\phi(\xi^j_{N,\delta}(\xi_0)) - \phi(\xi(j\delta;\xi_0)))$ in a telescoping sum allows one to estimate the error using the Kolmogorov equation associated to the limiting dynamic. These approaches require some degree of regularity for solutions of the Kolmogorov equation. Additionally, note that one needs to show that these estimates are uniform in j in order to address long time accuracy. This typically entails obtaining a time decay for the corresponding solutions. A further difficulty for SPDEs is that our Kolmogorov equation is a parabolic PDE whose 'spatial' variable is infinite dimensional. In the setting we are concerned with here, due to the necessity and interest for noise acting in a limited subset of the phase space, the Kolmogorov equation has a degenerate second-order term. Furthermore, its drift term involves an unbounded operator and a strongly nonlinear term.

Details of the Kolmogorov approach to weak convergence vary e.g. according to the model of interest, the type of discretization considered and the topology in which numerical convergence is established, but frequently the estimates appear with time-interval length dependent bounds. For finite dimensional SDEs we mention the pioneering works Milshtein (1979); Talay (1984, 1986); Milshtein (1995), which were further refined in a significant body of work, see e.g. Talay & Tubaro (1990); Kloeden & Platen (1992); Bally & Talay (1996); Kohatsu-Higa (2001); Szepessy *et al.* (2001); Clément *et al.* (2006) and references therein. Analogous weak convergence results for SPDEs were more recently obtained in e.g. Davie & Gaines (2001); Buckwar & Shardlow (2005); De Bouard & Debussche (2006); Debussche & Printems (2009); Hausenblas (2010); Debussche (2011); Kovács *et al.* (2012, 2013); Wang & Gan (2013); Andersson & Larsson (2016); Andersson *et al.* (2016); Wang (2016); Conus *et al.* (2019); Jentzen & Kurniawan (2021) for equations that are linear or with globally Lipschitz nonlinearities, and Dörsek (2012); Bréhier & Debussche (2018); Cui & Hong (2019); Bréhier & Goudenège (2020); Cai *et al.* (2021); Cui *et al.* (2021) for more general nonglobally Lipschitz scenarios. Another set of works focused on obtaining long time approximation error bounds as in (1.17), (1.19) or (1.20), for either SDEs (Talay, 1990; Shardlow & Stuart, 2000; Talay, 2002; Mattingly *et al.*, 2010; Debussche & Faou, 2012; Abdulle *et al.*, 2014) or SPDEs (Bréhier, 2014; Bréhier & Kopec, 2017; Cui & Hong, 2018; Hong & Wang, 2019; Cui *et al.*, 2021; Bréhier, 2022).

As previously mentioned, our approach for proving Theorem 1.1 is instead based on the uniform Wasserstein contraction framework described above. Namely, by establishing a uniform Wasserstein contraction estimate as in (1.8) together with a finite-time error estimate as in (1.5). Regarding the latter, our bound is in fact given in terms of the strong discretization error in $L^p(\Omega; L^\infty_{\mathrm{loc},t} L^2_x)$ for a sufficiently small p, which is estimated from Proposition 3.17 and Proposition 3.18 below. Clearly, this approach is most likely not guaranteeing an optimal weak convergence rate in (1.18). Indeed, as we previously mentioned, it is generally expected that the weak order of convergence is larger than the strong order, and many of the references on weak convergence results mentioned above focused precisely on establishing this order improvement.

However, we emphasize that the main advantage of our approach lies in yielding a *uniform in time* weak error estimate, (1.18), in addition to providing long time error estimates for approximations of the limiting stationary distribution, i.e. (1.17), (1.19), (1.20). Notably, these are the first results of such type to be established for the stochastic Navier–Stokes equations. Previous works on numerical approximations for the SNSE focused on strong convergence in either probability (Carelli & Prohl, 2012; Bessaih *et al.*, 2014; Hausenblas & Randrianasolo, 2019; Breit & Dodgson, 2021; Breit & Prohl, 2022) or in mean (Dörsek, 2012; Brzeźniak *et al.*, 2013; Bessaih & Millet, 2019, 2021; Milstein & Tretyakov, 2021; Bessaih & Millet, 2022) for various space or time discretizations and noise types, but always for bounds on finite time windows $[0,T]$ with exponentially growing constants as a function of $T>0$.

Regarding the strong error bound implied by Proposition 3.17 and Proposition 3.18, we notice that it is given more explicitly, for sufficiently regular starting point ξ_0, as

$$\mathbb{E}\left[\sup_{j\leq J}\left|\xi^j_{N,\delta}(\xi_0)-\xi(j\delta;\xi_0)\right|^p_{L^2}\right]\leq C\left[\delta^{\tilde{p}p}+N^{-\frac{p}{2}}\right] \tag{1.21}$$

for any $J\in\mathbb{N}$, $\tilde{p}\in(0,1/2)$, and for $p>0$ sufficiently small depending only on the viscosity parameter ν and $|\sigma|^2_{L^2}=\sum_k|\sigma_k|^2_{L^2}$, where $C=C(p,\tilde{p},J)$ is a positive constant. This bound yields a strong convergence for the scheme (1.15) with respect to the topology in $L^\infty_{\mathrm{loc},t}L^2_x$, and with rates (almost) $1/2$ in time and 1 in space ($N\sim h^{-2}$ for a spatial grid with cell edge length h), see also Printems (2001, Definition 2.6). Additionally, as ν grows large, p can be almost 2. This temporal convergence rate is optimal due to $1/2$-Hölder time regularity of the solution, which is in turn implied by the regularity of the underlying stochastic forcing term. Optimal rates for other types of numerical approximations or noise terms for the 2D SNSE were also achieved in previous works under the velocity formulation, particularly Bessaih *et al.* (2014); Breit & Dodgson (2021); Breit & Prohl (2022) regarding convergence in probability, and Bessaih & Millet (2022) concerning strong $L^2(\Omega)$ convergence under a suitable smallness assumption on the noise. In relation to these recent results, we expect our method of proof for deriving (1.21) to be of independent interest. See more details in Section 3.3 below.

1.4 *Outlook and future work*

A number of avenues for future development suggest themselves as an outgrowth of the work here. Firstly, the model problems suggested in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020); Glatt-Holtz *et al.* (2021) provide a set of interesting challenges for numerical analysis. Here note that, while the results in Section 4 provide a first step toward addressing the case of 2D stochastic Navier–Stokes on a domain with boundaries, subtle details remain to complete the analogous program to the one that we fulfilled in the periodic setting in Section 3. Note furthermore that Section 3 addresses just one of a variety of possible numerical approximations of governing equations, and indeed each of the model equations in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020); Glatt-Holtz *et al.* (2021) would be expected to have their own bespoke natural approximation schemes. Another challenge for numerical accuracy would be to address the fully hypo-elliptic case. Here, to obtain a uniform rate of contraction one would presumably need to develop a discrete time analogue of the infinite dimensional Malliavin calculus based approaches developed in Hairer & Mattingly (2006, 2011a); Földes *et al.* (2015); Kuksin *et al.* (2020). Of course, other interesting parameter limit problems outside of numerical approximation may be addressed from our formalism as in our concurrent work (Glatt-Holtz *et al.*, 2022b).

Finally, it is notable that abstract frameworks developed in Section 2 have a scope of applicability reaching far beyond the SPDE models that we have focused on here. As already identified in Johndrow & Mattingly (2017), one may leverage the type of contractivity obtained from weak Harris results as a means of bias estimation in a variety of applications in computational statistics. Our upcoming contribution (Glatt-Holtz *et al.*, 2022a) expands on this insight particularly leveraging the use of uniformity identified here.

Organization

The rest of this manuscript is organized as follows. In Section 2 we present our main abstract results, namely our Wasserstein contraction criteria in Section 2.2 followed by our parameter convergence/stability at time ∞ given in Section 2.3. Section 3 presents our first application concerning the numerical analysis of a fully discrete scheme for the stochastic Navier–Stokes equations. Finally, Section 4 presents contraction estimates for the stochastic Navier–Stokes equations on a bounded domain.

2. Abstract results

Before presenting our general results in Section 2.2 and Section 2.3, we briefly recall in Section 2.1 some standard definitions regarding Markov processes and the Wasserstein distance on spaces of probability measures. For more details, we refer to e.g. Da Prato & Zabczyk (1996); Villani (2008).

2.1 *Preliminaries*

Let X be a Polish space. Throughout this manuscript, we denote by $\mathscr{B}(X)$ the σ-algebra of Borel subsets of X, and by $\mathrm{Pr}(X)$ the corresponding space of Borel probability measures, which we will often simply refer to as probability measures. We also denote by $\mathscr{M}_b(X)$ the family of all real-valued, bounded and Borel-measurable functions on X. Moreover, we fix the notation $\mathbb{R}^+$ for the interval $[0, \infty)$.

We recall that $P : X \times \mathscr{B}(X) \to [0, 1]$ is a *Markov kernel* if $P(\cdot, \mathscr{O})$ is measurable for each fixed $\mathscr{O} \in \mathscr{B}(X)$, and $P(u, \cdot)$ is a probability measure for each fixed $u \in X$. For any measure $\mu \in \mathrm{Pr}(X)$, we recall that its dual action on a Markov kernel P is given by

$$\mu P(\mathscr{O}) := \int_X P(u, \mathscr{O})\mu(\mathrm{d}u), \quad \mathscr{O} \in \mathscr{B}(X).$$

A measure $\mu \in \mathrm{Pr}(X)$ is said to be *invariant* with respect to a family of Markov kernels P_t, $t \geq 0$, if and only if $\mu P_t = \mu$ for every $t \geq 0$.

Moreover, a *Markovian transition function* is a family of Markov kernels P_t, $t \geq 0$, such that, for each $u \in X$ and $\mathscr{O} \in \mathscr{B}(X)$, $P_0(u, \mathscr{O}) = \mathbb{1}_{\mathscr{O}}(u)$, where $\mathbb{1}_{\mathscr{O}}$ denotes the indicator function of $\mathscr{O}$, and it satisfies the Chapman–Kolmogorov relation

$$P_{t+s}(u, \mathscr{O}) = P_t P_s(u, \mathscr{O}) := \int_X P_s(v, \mathscr{O}) P_t(u, \mathrm{d}v).$$

Given such Markovian transition function, its associated *Markov semigroup* is defined as the family of operators P_t, $t \geq 0$, acting on functions $\varphi \in \mathscr{M}_b(X)$ as

$$P_t\varphi(u) := \int_X \varphi(v)P_t(u, \mathrm{d}v), \quad u \in X. \tag{2.1}$$

Finally, we recall that a mapping $\rho : X \times X \to \mathbb{R}^+$ is called a *distance-like* function if it is symmetric, lower semicontinuous, and satisfies that $\rho(u, \tilde{u}) = 0$ if and only if $u = \tilde{u}$, see Hairer *et al.* (2011, Definition 4.3). For any such distance-like function ρ, its Wasserstein-like extension to $\mathrm{Pr}(X)$ is the mapping $\mathscr{W}_\rho : \mathrm{Pr}(X) \times \mathrm{Pr}(X) \to \mathbb{R}^+ \cup \{\infty\}$ defined as

$$\mathscr{W}_\rho(\mu, \tilde{\mu}) = \inf_{\Gamma \in \mathscr{C}(\mu, \tilde{\mu})} \int_{X \times X} \rho(u, \tilde{u})\Gamma(\mathrm{d}u, \mathrm{d}\tilde{u}), \tag{2.2}$$

where $\mathscr{C}(\mu, \tilde{\mu})$ denotes the family of all *couplings* of μ and $\tilde{\mu}$, i.e. all probability measures Γ on the product space $X \times X$ with marginals μ and $\tilde{\mu}$. We notice that when ρ is a metric on X, then its corresponding extension $\mathscr{W}_\rho$ coincides with the usual Wasserstein-1 distance, Villani (2008).

2.2 *Wasserstein contraction*

Our first general result, Theorem 2.1 below, provides a general set of assumptions on a given Markov semigroup that are sufficient for guaranteeing its contraction with respect to a suitable Wasserstein distance. Our formulation is inspired by the weak Harris theorem from Hairer *et al.* (2011, Theorem 4.8), which yields an analogous Wasserstein contraction under three main assumptions on the Markov semigroup. Namely, the existence of a Lyapunov function; a smallness condition for trajectories departing from certain level sets of the Lyapunov function; and a contractivity assumption between trajectories departing from points that are sufficiently 'close' to each other.

In our set of hypotheses, we focus on stochastic systems possessing an exponential Lyapunov structure, while allowing for more flexibility regarding the contractivity requirement, see (2.5) below. In particular, our 'contraction' coefficient is given as the product of a constant that is smaller than 1 with an exponential term depending on one of the starting points. This is tailored to reflect a typical situation in applications to SPDEs, particularly involving a dissipative structure. Indeed, this is demonstrated in the applications to the stochastic Navier–Stokes equations in Section 3.2 and Section 4 below.

THEOREM 2.1 Let X be a separable Banach space with norm $\|\cdot\|$. Consider an index set $\mathscr{I}$ that is a subspace of $\mathbb{R}^+$. and take $\{P_t\}_{t\in\mathscr{I}}$ to be a Markov semigroup on X satisfying

(A1) (Exponential Lyapunov structure) There exists a continuous function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\lim_{t\to\infty} \psi(t) = 0$, and also $\alpha_0 > 0$ such that for all $\alpha \in (0, \alpha_0]$, $t \in \mathscr{I}$ and $u_0 \in X$, the following inequality holds:

$$P_t \exp\left(\alpha\|u_0\|^2\right) \leq \exp\left(\alpha\left(\psi(t)\|u_0\|^2 + C_0\right)\right) \tag{2.3}$$

for some constant $C_0 > 0$, which is independent of t, u_0 and α.

Furthermore, we fix a collection Λ of distance-like functions $\rho : X \times X \to [0, 1]$ and consider the following set of assumptions on Λ and $\{P_t\}_{t\in\mathscr{I}}$:

(A2) (Eventual ρ-smallness of bounded sets) For every $M > 0$ and $\rho \in \Lambda$ there exists $T_1 = T_1(M, \rho) > 0$ and $\kappa_1 = \kappa_1(M) \in (0, 1)$, which is independent of ρ, such that

$$\sup_{t \in \mathscr{I}, t \geq T_1} \mathscr{W}_\rho(P_t(u_0, \cdot), P_t(v_0, \cdot)) \leq 1 - \kappa_1 \tag{2.4}$$

for every $u_0, v_0 \in X$ with $\|u_0\| \leq M$ and $\|v_0\| \leq M$.

(A3) For every $\kappa_2 \in (0, 1)$ and for every $r > 0$ there exists $\rho \in \Lambda$ for which the following holds:

(A3.i) (Eventual local ρ-contractivity) There exists $T_2 = T_2(\kappa_2, r) > 0$ such that

$$\sup_{t \in \mathscr{I}, t \geq T_2} \mathscr{W}_\rho(P_t(u_0, \cdot), P_t(v_0, \cdot)) \leq \kappa_2 \exp\left(r\|u_0\|^2 + r\|v_0\|^2\right) \rho(u_0, v_0) \tag{2.5}$$

for every $u_0, v_0 \in X$ with $\rho(u_0, v_0) < 1$.

(A3.ii) For all $\tau \geq 0$, there exists $C = C(\tau, \rho) > 0$ such that

$$\sup_{t \in \mathscr{I},\ t \in [0, \tau]} \mathscr{W}_\rho(P_t(u_0, \cdot), P_t(v_0, \cdot)) \leq C \exp\left(r\|u_0\|^2 + r\|v_0\|^2\right) \rho(u_0, v_0) \tag{2.6}$$

for every $u_0, v_0 \in X$ with $\rho(u_0, v_0) < 1$.

Then, under the assumptions (A1), (A2) and (A3.i) it follows that for every $m \geq 1$, there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exists $\rho \in \Lambda$, $\kappa \in (0, 1)$ and $T > 0$ for which the following inequality holds

$$\mathscr{W}_{\rho_\alpha}(\mu P_t, \tilde{\mu} P_t) \leq \kappa \mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu}) \tag{2.7}$$

for every $t \in \mathscr{I}$ with $t \geq T$, and for all $\mu, \tilde{\mu} \in \Pr(X)$. Here, for each $a > 0$, $\rho_a : X \times X \to \mathbb{R}^+$ is the distance-like function defined as

$$\rho_a(u, v) := \rho(u, v)^{1/2} \exp(a\|u\|^2 + a\|v\|^2), \quad u, v \in X. \tag{2.8}$$

Moreover, under additionally assumption (A3.ii) it follows that for every $m \geq 1$, there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exists $\rho \in \Lambda$, $T > 0$ and constants $C_1, C_2 > 0$ for which it holds that

$$\mathscr{W}_{\rho_\alpha}(\mu P_t, \tilde{\mu} P_t) \leq C_1 \, e^{-C_2 t} \mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu}) \tag{2.9}$$

for every $\mu, \tilde{\mu} \in \Pr(X)$ and all $t \in \mathscr{I}$ with $t \geq T$. Here, the constants C_1 and C_2 depend *only* on the parameters m, α, ρ, T, and the constants α_0, C_0, $\sup_{t \geq 0} \psi$ from assumption (A1).

REMARK 2.2 Note that, from the definition of ρ_a in (2.8), it follows immediately that, given any $\mu, \tilde{\mu}$, we have $\mathscr{W}_{\rho_{\alpha_1}}(\mu, \tilde{\mu}) \leq \mathscr{W}_{\rho_{\alpha_2}}(\mu, \tilde{\mu})$ for all $\alpha_1, \alpha_2 \in \mathbb{R}^+$ with $\alpha_1 \leq \alpha_2$. Therefore, inequality (2.7) implies $\mathscr{W}_{\rho_\alpha}(\mu P_t, \tilde{\mu} P_t) \leq \kappa \mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu}) \leq \kappa \mathscr{W}_{\rho_\alpha}(\mu, \tilde{\mu})$, yielding contraction for the Markov semigroup $\{P_t\}_{t \in \mathscr{I}}$ with respect to $\mathscr{W}_{\rho_\alpha}$ in $\Pr(X)$. The additional flexibility provided by the parameter m in (2.7) will be needed later in Theorem 2.5, specifically to deal with distance-like functions ρ satisfying a generalized form of triangle inequality as in (2.24), where α is multiplied by a factor γ in the right-hand side. By choosing m appropriately, we are then able to perform the calculations leading to the crucial inequality (2.31) below.

Proof. We start by noticing that, since each ρ_α is lower-semicontinuous and non-negative, it follows from Villani (2008, Theorem 4.8) that for every $\mu, \tilde{\mu} \in \mathscr{P}(X)$

$$\mathscr{W}_{\rho_\alpha}(\mu P_t, \tilde{\mu} P_t) \leq \int_{X\times X} \mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot))\,\Gamma(\mathrm{d}u_0, \mathrm{d}v_0), \tag{2.10}$$

for every coupling $\Gamma \in \mathscr{C}(\mu, \tilde{\mu})$. Therefore, to show (2.7) it suffices to obtain that for every $m \geq 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exists $\rho \in \Lambda$, $\kappa \in (0,1)$ and $T > 0$ for which the following holds

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \kappa \rho_{\alpha/m}(u_0, v_0) \tag{2.11}$$

for every $t \in \mathscr{I}$ with $t \geq T$, and every $u_0, v_0 \in X$.

Fix $m \geq 1$ and $u_0, v_0 \in X$. For some fixed $\rho \in \Lambda$ to be suitably chosen later in terms of m, α and α_0, C_0 from (A1) and following similar ideas from Hairer *et al.* (2011, Theorem 4.8), we split the proof into three cases:

(i) Let us first suppose that $\rho(u_0, v_0) = 1$ and $\|u_0\|^2 + \|v_0\|^2 \leq 6mC_0$, with $C_0 > 0$ as in assumption (A1). From the definition of $\mathscr{W}_{\rho_\alpha}$ in (2.2) and Hölder's inequality, it follows that for all $t \in \mathscr{I}$

$$\begin{aligned}
&\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \\
&\quad = \inf_{\Gamma \in \mathscr{C}(P_t(u_0,\cdot), P_t(v_0,\cdot))} \int_{X\times X} \rho(u,v)^{1/2} \exp\left(\alpha\|u\|^2 + \alpha\|v\|^2\right) \Gamma(\mathrm{d}u, \mathrm{d}v) \\
&\quad \leq \inf_{\Gamma \in \mathscr{C}(P_t(u_0,\cdot), P_t(v_0,\cdot))} \left(\int_{X\times X} \rho(\mathrm{d}u, \mathrm{d}v)\,\Gamma(\mathrm{d}u, \mathrm{d}v)\right)^{1/2} \left(\int_{X\times X} \exp(4\alpha\|u\|^2)\,\Gamma(\mathrm{d}u, \mathrm{d}v)\right)^{1/4} \\
&\qquad\qquad \cdot \left(\int_{X\times X} \exp(4\alpha\|v\|^2)\,\Gamma(\mathrm{d}u, \mathrm{d}v)\right)^{1/4} \\
&\quad = \mathscr{W}_\rho(P_t(u_0,\cdot), P_t(v_0,\cdot))^{1/2} \left(P_t \exp\left(4\alpha\|u_0\|^2\right)\right)^{1/4} \left(P_t \exp\left(4\alpha\|v_0\|^2\right)\right)^{1/4}.
\end{aligned} \tag{2.12}$$

We now invoke assumption (A2) with $M := (6mC_0)^{1/2}$ to estimate the first factor in (2.12), and assumption (A1) to estimate the last two factors, assuming $\alpha \in (0, \alpha_0/4]$. It thus follows that for every $t \in \mathscr{I}$ with $t \geq T_1$

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq (1-\kappa_1)^{1/2} \exp(2\alpha C_0) \exp\left(\alpha\psi(t)\left(\|u_0\|^2 + \|v_0\|^2\right)\right). \tag{2.13}$$

Since $\kappa_1 \in (0,1)$, we can take $\alpha_m \in (0, \alpha_0/4]$ sufficiently small such that for every $\alpha \in (0, \alpha_m]$ it holds

$$\tilde{\kappa}_1 := (1-\kappa_1)^{1/2} \exp(2\alpha C_0) \leq (1-\kappa_1)^{1/2} \exp(2\alpha_m C_0) < 1. \tag{2.14}$$

Moreover, since $\lim_{t\to\infty}\psi(t) = 0$, we can take $\widetilde{T}_1 \geq T_1$ sufficiently large such that $\psi(t) < 1/m$ for all $t \geq \widetilde{T}_1$, so that it follows from (2.13) that for any fixed $\alpha \in (0, \alpha_m]$ and for every $t \in \mathscr{I}$ with $t \geq \widetilde{T}_1$

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \tilde{\kappa}_1 \exp\left(\frac{\alpha}{m}\left(\|u_0\|^2 + \|v_0\|^2\right)\right) = \tilde{\kappa}_1 \rho_{\alpha/m}(u_0, v_0), \tag{2.15}$$

where the equality follows from the assumption that $\rho(u_0, v_0) = 1$.

(ii) Next, we assume that $\rho(u_0, v_0) = 1$ and $\|u_0\|^2 + \|v_0\|^2 > 6mC_0$. Since $\rho(u, v) \leq 1$ for all $u, v \in X$, it follows together with Hölder's inequality and assumption (A1) that for any fixed $\alpha \in (0, \alpha_m]$, with α_m as in (2.14), and for every $t \in \mathscr{I}$

$$\begin{aligned}
\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) &= \inf_{\Gamma \in \mathscr{C}(P_t(u_0,\cdot), P_t(v_0,\cdot))} \int_{X\times X} \rho(u,v)^{1/2} \exp\left(\alpha\|u\|^2 + \alpha\|v\|^2\right) \Gamma(\mathrm{d}u, \mathrm{d}v) \\
&\leq \left(P_t \exp\left(2\alpha\|u_0\|^2\right)\right)^{1/2} \left(P_t \exp\left(2\alpha\|v_0\|^2\right)\right)^{1/2} \\
&\leq \exp(2\alpha C_0) \exp\left(\alpha\psi(t)\left(\|u_0\|^2 + \|v_0\|^2\right)\right).
\end{aligned} \tag{2.16}$$

Notice that

$$\exp(2\alpha C_0) = \exp(-\alpha C_0) \exp\left(\frac{\alpha}{2m} 6mC_0\right) < \exp(-\alpha C_0) \exp\left(\frac{\alpha}{2m}\left(\|u_0\|^2 + \|v_0\|^2\right)\right).$$

Thus, from (2.16), we obtain

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \exp(-\alpha C_0) \exp\left(\left(\frac{\alpha}{2m} + \alpha\psi(t)\right)\left(\|u_0\|^2 + \|v_0\|^2\right)\right).$$

Therefore, taking $\widetilde{T} > 0$ sufficiently large such that $\psi(t) < 1/(2m)$ for all $t \geq \widetilde{T}$, we deduce that for all $t \in \mathscr{I}$ with $t \geq \widetilde{T}$

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \tilde{\kappa} \exp\left(\frac{\alpha}{m}\left(\|u_0\|^2 + \|v_0\|^2\right)\right) = \tilde{\kappa} \rho_{\alpha/m}(u_0, v_0), \tag{2.17}$$

where $\tilde{\kappa} := \exp(-\alpha C_0) < 1$.

(iii) Finally, let us suppose that $\rho(u_0, v_0) < 1$. Take $r := \alpha/m$, for a fixed $\alpha \in (0, \alpha_m]$, with $\alpha_m \in (0, \alpha_0/4]$ as chosen in (2.14). Moreover, take $\kappa_2 \in (0,1)$ satisfying $\kappa_2 < \exp(-\alpha_0 C_0)$, with α_0, C_0 from assumption (A1). For these choices of r and κ_2, we fix $\rho \in \Lambda$ as being the corresponding distance-like function for which assumption (A3) holds. Here we notice carefully that since ρ depends on r, which depends on α, which, in its turn, as seen from (2.14), depends on κ_1 from assumption (A2), it thus follows that ρ depends on κ_1. Therefore, the fact that κ_1 in assumption (A2) is *independent* of ρ is crucial for preventing a circular argument in the choice of $\rho \in \Lambda$.

Proceeding with the same estimate as in (2.12), we now invoke assumption (A3.i) to estimate the first factor, and assumption (A1) to estimate the remaining two factors. It thus follows that for every $t \in \mathscr{I}$ with $t \geq T_2$

$$\begin{aligned}
&\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot),P_t(v_0,\cdot)) \\
&\leq \left[\kappa_2 \exp\left(\frac{\alpha}{m}\|u_0\|^2+\frac{\alpha}{m}\|v_0\|^2\right)\rho(u_0,v_0)\right]^{1/2}\exp(2\alpha C_0)\exp\left(\alpha\psi(t)\left(\|u_0\|^2+\|v_0\|^2\right)\right) \\
&\leq \kappa_2^{1/2}\exp(2\alpha C_0)\rho(u_0,v_0)^{1/2}\exp\left(\left(\frac{\alpha}{2m}+\alpha\psi(t)\right)\left(\|u_0\|^2+\|v_0\|^2\right)\right). \qquad (2.18)
\end{aligned}$$

Since $\alpha \in (0,\alpha_0/4]$ and, by the choice of κ_2, we have $\kappa_2 < \exp(-\alpha_0 C_0)$, then

$$\tilde{\kappa}_2 := (\kappa_2\exp(4\alpha C_0))^{1/2} \leq (\kappa_2\exp(\alpha_0 C_0))^{1/2} < 1. \qquad (2.19)$$

Moreover, taking as before $\widetilde{T}_2 \geq T_2$ sufficiently large such that $\psi(t) < 1/(2m)$ for all $t \geq \widetilde{T}_2$, we obtain from (2.18) and (2.19) that for all $t \in \mathscr{I}$ with $t \geq \widetilde{T}_2$

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot),P_t(v_0,\cdot)) \leq \tilde{\kappa}_2\rho_{\alpha/m}(u_0,v_0). \qquad (2.20)$$

From (2.15), (2.17) and (2.20), it follows that for each fixed $m \geq 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0,\alpha_m]$ there exists $\rho \in \Lambda$ and $T > 0$ for which (2.11) holds with $\kappa := \max\{\tilde{\kappa}_1,\tilde{\kappa}_2,\tilde{\kappa}\}$ and for all $t \in \mathscr{I}$ with $t \geq T$. This finishes the first part of the proof.

We proceed to show inequality (2.9) under the additional assumption (A3.ii), with the same choices of κ_2 and r from step (*iii*) above. Again as a consequence of (2.10), it suffices to show that for every $m \geq 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0,\alpha_m]$ there exists $\rho \in \Lambda$, $T > 0$ and constants $C_1, C_2 > 0$ such that

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot),P_t(v_0,\cdot)) \leq C_1\, e^{-C_2 t}\rho_{\alpha/m}(u_0,v_0) \qquad (2.21)$$

for all $t \in \mathscr{I}$ with $t \geq T$, and for all $u_0, v_0 \in X$.

Fix $m \geq 1$. Take $K \geq 1$ such that the function ψ from assumption (A1) satisfies $\psi(t) \leq K$ for all $t \geq 0$. Take also $\alpha_{2mK} > 0$ as in (2.14) corresponding to the parameter $2mK$. Let us fix $\alpha \in (0,\alpha_{2mK}]$ and the corresponding $\kappa \in (0,1)$, $\rho \in \Lambda$ and $T > 0$ for which (2.7) holds. Clearly, we may assume $T \in \mathscr{I}$. Then, for any $t \in \mathscr{I}$ with $t \geq T$, we may write $t = jT + s$, with $j \in \mathbb{N}$, $j \geq 1$ and $s \in [0,T) \cap \mathscr{I}$. Notice that $jT \in \mathscr{I}$, for all $j \in \mathbb{N}$. Thus, invoking (2.11) j times, it follows that for all $u_0, v_0 \in X$

$$\begin{aligned}
\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot),P_t(v_0,\cdot)) &= \mathscr{W}_{\rho_\alpha}(P_{(j-1)T+s}(u_0,\cdot)P_T, P_{(j-1)T+s}(v_0,\cdot)P_T) \\
&\leq \kappa\mathscr{W}_{\rho_{\alpha/(2mK)}}(P_{(j-1)T+s}(u_0,\cdot),P_{(j-1)T+s}(v_0,\cdot)) \\
&\leq \kappa\mathscr{W}_{\rho_\alpha}(P_{(j-1)T+s}(u_0,\cdot),P_{(j-1)T+s}(v_0,\cdot)) \\
&\leq \ldots \leq \kappa^j\mathscr{W}_{\rho_{\alpha/(2mK)}}(P_s(u_0,\cdot),P_s(v_0,\cdot)). \qquad (2.22)
\end{aligned}$$

From a similar calculation as in (2.12), we have that for all $s \in [0, T) \cap \mathscr{I}$

$$\begin{aligned}&\mathscr{W}_{\rho_{\alpha/(2mK)}}(P_s(u_0,\cdot), P_s(v_0,\cdot))\\&\qquad \leq \mathscr{W}_\rho(P_s(u_0,\cdot), P_s(v_0,\cdot))^{1/2}\left(P_s \exp\left(2\frac{\alpha}{mK}\|u_0\|^2\right)\right)^{1/4}\left(P_s \exp\left(2\frac{\alpha}{mK}\|v_0\|^2\right)\right)^{1/4}.\end{aligned}$$

Hence, recalling the analogous choice of r in step (*iii*) above, namely $r := \alpha/(2mK)$, with $\alpha \in (0, \alpha_{2mK}]$ and α_{2mK} as in (2.14), it follows from assumption 2.1 along with assumption (A1) that

$$\begin{aligned}&\mathscr{W}_{\rho_{\alpha/(2mK)}}(P_s(u_0,\cdot), P_s(v_0,\cdot))\\&\quad \leq C\exp\left(\frac{\alpha}{4mK}\left(\|u_0\|^2+\|v_0\|^2\right)\right)\rho(u_0,v_0)^{1/2}\exp\left(\frac{\alpha}{mK}C_0\right)\exp\left(\frac{\alpha}{2mK}\psi(s)\left(\|u_0\|^2+\|v_0\|^2\right)\right)\\&\quad \leq C\exp\left(\frac{\alpha}{4m}\left(\|u_0\|^2+\|v_0\|^2\right)\right)\rho(u_0,v_0)^{1/2}\exp\left(\frac{\alpha}{mK}C_0\right)\exp\left(\frac{\alpha}{2m}\left(\|u_0\|^2+\|v_0\|^2\right)\right)\\&\quad \leq C\rho_{\alpha/m}(u_0,v_0). \end{aligned} \tag{2.23}$$

Plugging (2.23) into (2.22) yields

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \kappa^j C\rho_{\alpha/m}(u_0,v_0).$$

Since $t = jT + s < (j+1)T$, then $j > (t/T) - 1$ and, consequently, $\kappa^j < \kappa^{\frac{t}{T}-1}$. Thus,

$$\mathscr{W}_{\rho_\alpha}(P_t(u_0,\cdot), P_t(v_0,\cdot)) \leq \kappa^{\frac{t}{T}-1}C\rho_{\alpha/m}(u_0,v_0) = \frac{C}{\kappa}e^{t\frac{\ln\kappa}{T}}\rho_{\alpha/m}(u_0,v_0)$$

for all $t \in \mathscr{I}$ with $t \geq T$, and every $u_0, v_0 \in X$. Therefore, (2.21) holds with $C_1 = C/\kappa$ and $C_2 = -(\ln\kappa)/T$. This concludes the proof. □

REMARK 2.3 Clearly, if P_t, $t \in \mathscr{I}$, is a Markov semigroup satisfying the assumptions of Theorem 2.1, and for which there exists an associated invariant measure $\mu_* \in \Pr(X)$, i.e. $\mu_* P_t = \mu_*$ for all $t \in \mathscr{I}$, then inequality (2.7) implies that μ_* must also be the unique invariant measure. Moreover, fixing $\rho \in \Lambda$ to be the distance-like function for which (2.7) holds, it follows similarly as in Hairer *et al.* (2011, Corollary 4.11) that if there exists a complete metric $\tilde{\rho}$ on X such that $\tilde{\rho} \leq \sqrt{\rho}$ and such that P_t is a Feller semigroup on $(X, \tilde{\rho})$, then together with assumptions (A1), (A2) and (A3.i) one can also guarantee the existence of such invariant measure. Indeed, an explicit example of $\tilde{\rho}$ can be easily identified for the specific ρ we consider in our applications and given in (3.26) below, see Corollary 3.10.

REMARK 2.4 From the proof of Theorem 2.1 it follows that the contraction constants from (2.7) and (2.9) become worse with decreasing values of α. Indeed, since κ from (2.7) is given by $\kappa := \max\{\tilde{\kappa}_1, \tilde{\kappa}_2, \tilde{\kappa}\}$, then from the definitions of $\tilde{\kappa}_1, \tilde{\kappa}, \tilde{\kappa}_2$ given in steps (i), (ii), (iii), respectively, one obtains that $\kappa \to 1$ as $\alpha \to 0$.

2.3 *Uniform in time weak convergence*

The following general result provides the specific set of assumptions needed for achieving a long time bias estimate similar to (1.7), though in a more general setting than considered in Section 1.1. Indeed, our estimate is given with respect to the Wasserstein distance induced by the distance-like function ρ_α

defined in (2.8) above, thus not necessarily a metric. The main assumptions are given by: a generalized triangle inequality satisfied by ρ_α, (H1); the existence of an invariant measure for each member of the given parametrized family of Markov kernels, (H2); a finite-time error estimate for the approximating processes, (H3); and a parameter-uniform Wasserstein contraction for the given family of Markov kernels. Under these assumptions, the proof follows essentially the same steps of argumentation leading to (1.7).

THEOREM 2.5 Let X be a separable Banach space with norm $\|\cdot\|$. Fix a collection Λ of distance-like functions $\rho : X \times X \to [0,1]$. Consider a family of Markov kernels P_t^θ on X indexed by $t \in \mathbb{R}^+$ and a parameter θ varying in some set Θ. Assume the following set of conditions:

(H1) There exists a constant $\gamma \geq 1$ such that for every $\rho \in \Lambda$ and $\alpha > 0$ the distance-like function $\rho_\alpha : X \times X \to \mathbb{R}^+$ defined in (2.8) satisfies

$$\rho_\alpha(u,v) \leq C\left[\rho_{\gamma\alpha}(u,w) + \rho_{\gamma\alpha}(w,v)\right] \tag{2.24}$$

for all $u, v, w \in X$ and for some constant $C > 0$ (which may depend on ρ, α and γ).

(H2) For each $\theta \in \Theta$, there exists a probability measure μ_θ on X, which is invariant under P_t^θ, $t \in \mathbb{R}^+$.

(H3) There exist $\theta_0 \in \Theta$ and $\alpha' > 0$ such that for each $\alpha \in (0, \alpha']$ and for each $\rho \in \Lambda$ there exist functions $g_{\theta_0} : \Theta \to \mathbb{R}^+$, $R : \mathbb{R}^+ \to \mathbb{R}^+$, and a measurable function $f : X \to \mathbb{R}^+ \cup \{\infty\}$ such that

$$\mathscr{W}_{\rho_\alpha}\left(P_t^\theta(u,\cdot), P_t^{\theta_0}(u,\cdot)\right) \leq R(t) f(u) g_{\theta_0}(\theta), \tag{2.25}$$

for all $\theta \in \Theta$, $t \in \mathbb{R}^+$ and $u \in X$.

(H4) For every $m \geq 1$, there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exists $\rho \in \Lambda$, $T > 0$ and constants $C_1, C_2 > 0$ for which the following inequality holds:

$$\sup_{\theta \in \Theta \setminus \{\theta_0\}} \mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \tilde{\mu} P_t^\theta\right) \leq C_1\, e^{-tC_2} \mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu}) \tag{2.26}$$

for every $\mu, \tilde{\mu} \in \Pr(X)$ and $t \geq T$, with $\theta_0 \in \Theta$ as in (H3).

Then, there exists $\alpha_* > 0$ such that for each fixed $\alpha \in (0, \alpha_*]$ there exists $\rho \in \Lambda$, $\widetilde{T} > 0$ and a constant $C > 0$ for which it holds that

$$\mathscr{W}_{\rho_\alpha}(\mu_\theta, \mu_{\theta_0}) \leq C R(\widetilde{T}) g_{\theta_0}(\theta) \int_X f(u) \mu_{\theta_0}(\mathrm{d}u), \tag{2.27}$$

for every $\theta \in \Theta$.

REMARK 2.6 The crucial condition (2.24) of Theorem 2.5 is not hard to verify in practice. The main underlying condition for the collection of distance like functions Λ is that they satisfy a generalized triangle inequality, namely that for $\rho \in \Lambda$, we have the bound $\rho(u,w) \leq C(\rho(u,v) + \rho(u,w))$, for a constant C independent of any $u, v, w \in X$. See Proposition A.1 below for our precise formulation and (3.26) in Section 3.2, (4.8) in Section 4.2 where we put this result into practice.

Proof. Due to assumptions (H1) and (H2), together with Proposition A.2, it follows that for each $\rho \in \Lambda$ and $\alpha > 0$ there exists a constant $C > 0$ such that for all $t \in \mathbb{R}^+$ and $\theta \in \Theta \backslash \{\theta_0\}$

$$\mathscr{W}_{\rho_\alpha}\left(\mu_\theta, \mu_{\theta_0}\right) = \mathscr{W}_{\rho_\alpha}\left(\mu_\theta P_t^\theta, \mu_{\theta_0} P_t^{\theta_0}\right) \leq C\left[\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_\theta P_t^\theta, \mu_{\theta_0} P_t^\theta\right) + \mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0} P_t^\theta, \mu_{\theta_0} P_t^{\theta_0}\right)\right]. \quad (2.28)$$

Now invoking assumption (H4) with $m = \gamma$ to estimate the first term in the right-hand side of (2.28), we obtain that for any fixed $\alpha \in (0, \alpha_\gamma/\gamma]$ and corresponding $\rho \in \Lambda$, $T > 0$ and constants $C_1, C_2 > 0$, we have

$$\mathscr{W}_{\rho_\alpha}\left(\mu_\theta, \mu_{\theta_0}\right) \leq C\left[C_1\, e^{-tC_2} \mathscr{W}_{\rho_\alpha}\left(\mu_\theta, \mu_{\theta_0}\right) + \mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0} P_t^\theta, \mu_{\theta_0} P_t^{\theta_0}\right)\right] \quad (2.29)$$

for all $t \geq T$. Take $\widetilde{T} \geq T$ such that

$$CC_1\, e^{-\widetilde{T}C_2} < \frac{1}{2}. \quad (2.30)$$

Thus, taking $t = \widetilde{T}$ in (2.29) and rearranging terms, we deduce that

$$\mathscr{W}_{\rho_\alpha}\left(\mu_\theta, \mu_{\theta_0}\right) \leq 2C \mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0} P_{\widetilde{T}}^\theta, \mu_{\theta_0} P_{\widetilde{T}}^{\theta_0}\right). \quad (2.31)$$

Moreover, similarly as in (2.10), it follows from Villani (2008, Theorem 4.8) that

$$\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0} P_{\widetilde{T}}^\theta, \mu_{\theta_0} P_{\widetilde{T}}^{\theta_0}\right) \leq \int_{X\times X} \mathscr{W}_{\rho_{\gamma\alpha}}\left(P_{\widetilde{T}}^\theta(u,\cdot), P_{\widetilde{T}}^{\theta_0}(v,\cdot)\right) \Gamma(\mathrm{d}u, \mathrm{d}v), \quad (2.32)$$

for every coupling $\Gamma \in \mathscr{C}(\mu_{\theta_0}, \mu_{\theta_0})$. Take $\Gamma(\mathrm{d}u, \mathrm{d}v) = \delta_u(\mathrm{d}v)\mu_{\theta_0}(\mathrm{d}u)$, where $\delta_u \in \mathrm{Pr}(X)$ denotes the Dirac measure concentrated at $u \in X$. It is not difficult to check that $\Gamma \in \mathscr{C}(\mu_{\theta_0}, \mu_{\theta_0})$. Therefore,

$$\begin{aligned}\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0} P_{\widetilde{T}}^\theta, \mu_{\theta_0} P_{\widetilde{T}}^{\theta_0}\right) &\leq \int_{X\times X} \mathscr{W}_{\rho_{\gamma\alpha}}\left(P_{\widetilde{T}}^\theta(u,\cdot), P_{\widetilde{T}}^{\theta_0}(v,\cdot)\right) \delta_u(\mathrm{d}v)\mu_{\theta_0}(\mathrm{d}u) \\ &= \int_X \mathscr{W}_{\rho_{\gamma\alpha}}\left(P_{\widetilde{T}}^\theta(u,\cdot), P_{\widetilde{T}}^{\theta_0}(u,\cdot)\right) \mu_{\theta_0}(\mathrm{d}u). \quad (2.33)\end{aligned}$$

Let us assume additionally that $\alpha \leq \alpha'/\gamma$, with α' as in (H3), so that inequality (2.25) from (H3) holds with respect to $\gamma\alpha$, namely

$$\mathscr{W}_{\rho_{\gamma\alpha}}\left(P_t^\theta(u,\cdot), P_t^{\theta_0}(u,\cdot)\right) \leq R(t) f(u) g_{\theta_0}(\theta), \quad (2.34)$$

for all $\theta \in \Theta$, $t \in \mathbb{R}^+$, $u \in X$, and where we are fixing ρ as in (2.29). It thus follows from (2.31), (2.33) and (2.34) with $t = \widetilde{T}$ that

$$\mathscr{W}_{\rho_\alpha}(\mu_\theta, \mu_{\theta_0}) \leq 2C\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_{\theta_0}P^\theta_{\widetilde{T}}, \mu_{\theta_0}P^{\theta_0}_{\widetilde{T}}\right) \leq 2CR(\widetilde{T})g_{\theta_0}(\theta)\int_X f(u)\,\mu_{\theta_0}(\mathrm{d}u). \tag{2.35}$$

This shows (2.27), and concludes the proof. □

Next we notice that a finite time error estimate as in (2.25), assuming $g_{\theta_0}(\theta) \to 0$ as $\theta \to \theta_0$, combined with a uniform Wasserstein contraction for the approximating family $\{P^\theta_t\}_{t\geq 0}$, $\theta \in \Theta\backslash\{\theta_0\}$, as in (2.26) yields a Wasserstein contraction result for the limiting process $\{P^{\theta_0}_t\}_{t\geq 0}$. This is made precise as follows.

LEMMA 2.7 Fix the same setting from Theorem 2.5 and assume that hypotheses (H1)–(H4) hold. Regarding (H3), suppose additionally that for each $\alpha \in (0, \alpha']$ and $\rho \in \Lambda$ the corresponding function g_{θ_0} is such that $\lim_{\theta\to\theta_0} g_{\theta_0}(\theta) = 0$. Then, for every $m \geq 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exists $\rho \in \Lambda$, $\widetilde{T} > 0$ and constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ for which the following inequality holds:

$$\mathscr{W}_{\rho_\alpha}\left(\mu P^{\theta_0}_t, \tilde{\mu}P^{\theta_0}_t\right) \leq \widetilde{C}_1\, e^{-t\widetilde{C}_2}\mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu}) \tag{2.36}$$

for every $t \geq \widetilde{T}$ and all $\mu, \tilde{\mu} \in \Pr(X)$ satisfying

$$\int_X f(u)\mu(\mathrm{d}u) + \int_X f(u)\tilde{\mu}(\mathrm{d}u) < \infty, \tag{2.37}$$

where f is the function from (H3).

Proof. Fix any $m \geq 1$ and μ, $\tilde{\mu}$ satisfying (2.37). Invoking (H1) and Proposition A.2 twice, we obtain

$$\mathscr{W}_{\rho_\alpha}\left(\mu P^{\theta_0}_t, \tilde{\mu}P^{\theta_0}_t\right) \leq C\left[\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu P^{\theta_0}_t, \mu P^\theta_t\right) + \mathscr{W}_{\rho_{\gamma^2\alpha}}\left(\mu P^\theta_t, \tilde{\mu}P^\theta_t\right) + \mathscr{W}_{\rho_{\gamma^2\alpha}}\left(\tilde{\mu}P^\theta_t, \tilde{\mu}P^{\theta_0}_t\right)\right] \tag{2.38}$$

for any $\theta \in \Theta\backslash\{\theta_0\}$. Now we assume $\alpha > 0$ is sufficiently small, then proceed as in (2.32)–(2.33) and invoke (H3) to estimate the first and third terms in the right-hand side of (2.38), and (H4) to estimate the second term. It thus follows that for any such α there exists $\rho \in \Lambda$ and $T > 0$ such that

$$\mathscr{W}_{\rho_\alpha}\left(\mu P^{\theta_0}_t, \tilde{\mu}P^{\theta_0}_t\right) \leq C\left[R(t)g_{\theta_0}(\theta)\left(\int_X f(u)\mu(\mathrm{d}u) + \int_X f(u)\tilde{\mu}(\mathrm{d}u)\right) + C_1\, e^{-tC_2}\mathscr{W}_{\rho_{\alpha/m}}(\mu, \tilde{\mu})\right]$$

for all $t \geq T$, where C_1, C_2 are the same as in (2.26). Thus, taking the limit as $\theta \to \theta_0$ and recalling the assumptions that $\lim_{\theta\to\theta_0} g_{\theta_0}(\theta) = 0$ and (2.37), we deduce (2.36). □

Next, we show that, under the same assumptions from Theorem 2.5 together with some natural conditions on the functions appearing in the right-hand side of the finite-time error estimate (2.25), it follows that the given parametrized family of Markov kernels P^θ_t, $\theta \in \Theta$, converges uniformly in $t \geq 0$ towards $P^{\theta_0}_t$ in the Wasserstein topology determined by $\mathscr{W}_{\rho_\alpha}$.

THEOREM 2.8 Fix the same setting from Theorem 2.5 and assume that hypotheses (H1)–(H4) hold. Additionally, regarding (H3), suppose that for each $\alpha \in (0, \alpha']$ and $\rho \in \Lambda$ the corresponding functions R, g_{θ_0} and f satisfy:

(H5) R is continuous and strictly increasing;

(H6) g_{θ_0} is bounded, and $\lim_{\theta \to \theta_0} g_{\theta_0}(\theta) = 0$;

(H7) $\int_X f(u)\mu_{\theta_0}(\mathrm{d}u) < \infty$, with μ_{θ_0} as in (H2).

Then, there exists $\hat{\alpha} > 0$ such that for each fixed $\alpha \in (0, \hat{\alpha}]$ there exists $\rho \in \Lambda$ for which the following inequality holds for every $\theta \in \Theta$ and $\mu \in \Pr(X)$ with $\int_X f(u)\mu(\mathrm{d}u) < \infty$:

$$\sup_{t \in \mathbb{R}^+} \mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) \le \tilde{g}_{\theta_0}(\theta) \left[\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) + \int_X f(u)\mu(\mathrm{d}u) + \int_X f(u)\mu_{\theta_0}(\mathrm{d}u)\right], \tag{2.39}$$

where, if R is bounded,

$$\tilde{g}_{\theta_0}(\theta) = C g_{\theta_0}(\theta),$$

and if R is unbounded

$$\tilde{g}_{\theta_0}(\theta) = C \max\left\{\exp\left(-C_2 R^{-1}(C g_{\theta_0}(\theta)^{-q})\right), g_{\theta_0}(\theta), g_{\theta_0}(\theta)^{1-q}\right\}, \tag{2.40}$$

for $g_{\theta_0}(\theta) \neq 0$, and $\tilde{g}_{\theta_0}(\theta) = 0$ otherwise. Here, q is any fixed number in $(0, 1)$, $C > 0$ is a constant that may depend on α, ρ and γ from (H1), but is independent of θ, μ; and C_2 is the constant from (2.26).

Consequently, if additionally $\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) < \infty$, then

$$\lim_{\theta \to \theta_0} \sup_{t \in \mathbb{R}^+} \mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) = 0. \tag{2.41}$$

Proof. From Proposition A.2, along with assumptions (H1) and (H2), we obtain that for every $\mu \in \Pr(X)$, $\alpha > 0$, $\rho \in \Lambda$, $t \in \mathbb{R}^+$ and $\theta \in \Theta \backslash \{\theta_0\}$

$$\begin{aligned} \mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) &\le C\left[\mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu P_t^\theta, \mu_\theta P_t^\theta\right) + \mathscr{W}_{\rho_{\gamma\alpha}}\left(\mu_\theta, \mu P_t^{\theta_0}\right)\right] \\ &\le C\left[\mathscr{W}_{\rho_{\gamma^2\alpha}}\left(\mu P_t^\theta, \mu_{\theta_0} P_t^\theta\right) + \mathscr{W}_{\rho_{\gamma^2\alpha}}\left(\mu_{\theta_0} P_t^\theta, \mu_\theta P_t^\theta\right) + \mathscr{W}_{\rho_{\gamma^2\alpha}}(\mu_\theta, \mu_{\theta_0}) + \mathscr{W}_{\rho_{\gamma^2\alpha}}\left(\mu_{\theta_0} P_t^{\theta_0}, \mu P_t^{\theta_0}\right)\right]. \end{aligned} \tag{2.42}$$

Now we invoke assumption (H4) to estimate the first and second terms in the right-hand side of (2.42), (2.27) from Theorem 2.5 to estimate the third term and Lemma 2.7 to estimate the fourth term. Here we notice that we can apply (H4) and Lemma 2.7 with any choice of $m \ge 1$ to estimate the first, second and fourth terms. But for estimating the third term via (2.27), as we recall from the proof of Theorem 2.5, we must invoke (H4) with $m = \gamma$. Since the distance-like function $\rho \in \Lambda$ for which (2.26) in (H4) and (2.36) in Lemma 2.7 hold depends in particular on the choice of m, we thus also estimate the first, second and fourth terms under $m = \gamma$. This yields that for $\alpha > 0$ sufficiently small there exists $\rho \in \Lambda$, $T > 0$, and constants $C_1, C_2 > 0$ such that

$$\begin{aligned}\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) &\leq C\left[C_1\, e^{-tC_2}\mathscr{W}_{\rho_{\gamma\alpha}}(\mu,\mu_{\theta_0}) + C_1\, e^{-tC_2}\mathscr{W}_{\rho_{\gamma\alpha}}(\mu_{\theta_0},\mu_\theta) + \mathscr{W}_{\rho_{\gamma^2\alpha}}(\mu_\theta,\mu_{\theta_0})\right] \\ &\leq C\left[e^{-tC_2}\mathscr{W}_{\rho_{\gamma\alpha}}(\mu,\mu_{\theta_0}) + \mathscr{W}_{\rho_{\gamma^2\alpha}}(\mu_\theta,\mu_{\theta_0})\right] \\ &\leq C\left[e^{-tC_2}\mathscr{W}_{\rho_{\gamma\alpha}}(\mu,\mu_{\theta_0}) + g_{\theta_0}(\theta)\int_X f(u)\mu_{\theta_0}(\mathrm{d}u)\right],\end{aligned} \tag{2.43}$$

for every $t \geq T, \theta \in \Theta$ and $\mu \in \Pr(X)$ such that $\int_X f(u)\mu(\mathrm{d}u) < \infty$. Here, $C > 0$ is a constant that may depend on γ, α, T, ρ, but is independent of t, θ, μ.

On the other hand, proceeding similarly as in (2.32)–(2.35), we obtain directly from assumption (H3) that, by taking α smaller if necessary,

$$\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) \leq R(t) g_{\theta_0}(\theta)\int_X f(u)\mu(\mathrm{d}u), \tag{2.44}$$

for all $\theta \in \Theta$ and $t \in \mathbb{R}^+$. We now combine inequalities (2.43) and (2.44) to yield the desired uniform in $t \in \mathbb{R}^+$ estimate (2.39).

Let us first suppose that R is a bounded function. In this case, it follows immediately from (2.44) that

$$\sup_{t\in\mathbb{R}^+}\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) \leq C g_{\theta_0}(\theta)\int_X f(u)\mu(\mathrm{d}u),$$

for every $\theta \in \Theta$ and for some constant $C > 0$, as desired.

Now let us assume that R is unbounded. Fix $q \in (0,1)$ and let $t_* > 0$ such that $R(t_*) \neq 0$. Let $\tilde{R}(\tau) := R(\tau)/R(t_*)$, $\tau \in \mathbb{R}^+$, and $\tilde{g}_{\theta_0}(\theta) := g_{\theta_0}(\theta)/\overline{g_{\theta_0}}$, $\theta \in \Theta$, where $\overline{g_{\theta_0}} := \sup_{\theta\in\Theta} g_{\theta_0}(\theta)$. Here we recall our assumptions that R is strictly increasing and continuous, and g_{θ_0} is bounded, which implies that $\tilde{R}$ is also strictly increasing and continuous, and $\overline{g_{\theta_0}} < \infty$.

Fix $\theta \in \Theta$ and assume first that $g_{\theta_0}(\theta) \neq 0$, so that $\tilde{g}_{\theta_0}(\theta) \neq 0$. Notice that $\tilde{R}(t_*) = 1$, and $\tilde{g}_{\theta_0}(\theta) \leq 1$, so that $\tilde{R}(t_*) \leq \tilde{g}_{\theta_0}(\theta)^{-q}$. Since $\tilde{R} : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous and unbounded, there exists $\tau_* \in \mathbb{R}^+$ such that $\tilde{R}(\tau_*) = \tilde{g}_{\theta_0}(\theta)^{-q}$. And since $R, \tilde{R}$ are strictly increasing, then their corresponding inverses R^{-1}, $\tilde{R}^{-1}$ are well-defined, so that $\tau_* = \tilde{R}^{-1}(\tilde{g}_{\theta_0}(\theta)^{-q}) = R^{-1}(g_{\theta_0}(\theta)^{-q}R(t_*)/\overline{g_{\theta_0}})$. From (2.44), we thus obtain that for every $t \leq \tau_*$

$$\begin{aligned}\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) \leq R(\tau_*)g_{\theta_0}(\theta)\int_X f(u)\mu(\mathrm{d}u) &= \tilde{R}(\tau_*)\tilde{g}_{\theta_0}(\theta)R(t_*)\overline{g_{\theta_0}}\int_X f(u)\mu(\mathrm{d}u) \\ &= \tilde{g}_{\theta_0}(\theta)^{1-q}R(t_*)\overline{g_{\theta_0}}\int_X f(u)\mu(\mathrm{d}u) \\ &\leq C g_{\theta_0}(\theta)^{1-q}\int_X f(u)\mu(\mathrm{d}u),\end{aligned} \tag{2.45}$$

and for every $t \leq T$

$$\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) \leq R(T) g_{\theta_0}(\theta) \int_X f(u)\mu(\mathrm{d}u). \tag{2.46}$$

On the other hand, it follows from (2.43) that for every $t > \max\{\tau_*, T\}$

$$\begin{aligned}\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) &\leq C\left[e^{-\tau_* C_2}\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) + g_{\theta_0}(\theta)\int_X f(u)\mu_{\theta_0}(\mathrm{d}u)\right] \\ &\leq C\max\left\{\exp\left(-C_2 R^{-1}(g_{\theta_0}(\theta)^{-q} R(t_*)/\overline{g_{\theta_0}})\right), g_{\theta_0}(\theta)\right\}\left[\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) + \int_X f(u)\mu_{\theta_0}(\mathrm{d}u)\right]. \end{aligned} \tag{2.47}$$

From (2.45), (2.46) and (2.47) we conclude that

$$\begin{aligned}\sup_{t\in\mathbb{R}^+}\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) &\leq C\max\left\{\exp\left(-C_2 R^{-1}(g_{\theta_0}(\theta)^{-q} R(t_*)/\overline{g_{\theta_0}})\right), g_{\theta_0}(\theta), g_{\theta_0}(\theta)^{1-q}\right\} \\ &\quad\cdot\left[\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) + \int_X f(u)\mu_{\theta_0}(\mathrm{d}u) + \int_X f(u)\mu(\mathrm{d}u)\right], \end{aligned} \tag{2.48}$$

for all $\theta \in \Theta$ such that $g_{\theta_0}(\theta) \neq 0$. Further, if $g_{\theta_0}(\theta) = 0$, then it follows directly from (2.44) that

$$\mathscr{W}_{\rho_\alpha}\left(\mu P_t^\theta, \mu P_t^{\theta_0}\right) = 0 \quad \text{for all } t \in \mathbb{R}^+.$$

This concludes the proof of (2.39). Finally, the validity of (2.41) is clear under (2.39) and the additional condition $\mathscr{W}_{\rho_{\gamma\alpha}}(\mu, \mu_{\theta_0}) < \infty$. □

REMARK 2.9 Note that the order of convergence implied by the uniform in time estimate (2.40) is possibly smaller than the order of convergence from the finite-time error assumption in (H3), (2.25).

REMARK 2.10 In the particular case that assumption (H3) holds with $R(t) = \widetilde{C} e^{tC'}$ for some positive constants $\widetilde{C}, C'$, it follows that, under assumptions (H1)–(H3), inequality (2.39) holds with

$$\tilde{g}_{\theta_0}(\theta) = C\max\left\{g_{\theta_0}(\theta)^{\frac{C_2 q}{C'}}, g_{\theta_0}(\theta), g_{\theta_0}(\theta)^{1-q}\right\} \tag{2.49}$$

for any fixed $q < 1$ and for some constant $C > 0$. Indeed, this follows directly from the general expression of $\tilde{g}_{\theta_0}(\theta)$ in (2.40) for this specific case of R. This particular situation appears in the application to a numerical discretization of the 2D SNSE presented in Section 3 below. From (2.49), it is clear that larger values of C' imply a smaller order of convergence of the numerical scheme in the sense specified by (2.39).

To conclude this section, we have the following immediate corollary of Theorem 2.8 yielding uniform-in-time weak convergence for stochastic processes associated to the Markov kernels $P_t^\theta, \theta \in \Theta$, with respect to Lipschitz test functions.

COROLLARY 2.11 Fix the same setting and assumptions from Theorem 2.8. For each $u_0 \in X$ and $\theta \in \Theta$, let $u_\theta(t; u_0)$, $t \in \mathbb{R}^+$, be a stochastic process such that $\mathscr{L}(u_\theta(t; u_0)) = P_t^\theta(u_0, \cdot)$ for every $t \in \mathbb{R}^+$, where $\mathscr{L}(u_\theta(t; u_0))$ denotes the law of $u_\theta(t; u_0)$. Then, there exists $\hat{\alpha} > 0$ such that for each $\alpha \in (0, \hat{\alpha}]$ there exists $\rho \in \Lambda$ for which the following inequality holds for every $\theta \in \Theta$, $u_0 \in X$ such that $f(u_0) < \infty$, and every ρ_α-Lipschitz function $\varphi : X \to \mathbb{R}$ with Lipschitz constant L_φ:

$$\sup_{t \in \mathbb{R}^+} \left| \mathbb{E}\left[\varphi(u_\theta(t; u_0)) - \varphi(u_{\theta_0}(t; u_0)) \right] \right| \leq L_\varphi \tilde{g}_{\theta_0}(\theta) \left[\mathscr{W}_{\rho_{\gamma\alpha}}(\delta_{u_0}, \mu_{\theta_0}) + f(u_0) + \int_X f(u) \mu_{\theta_0}(\mathrm{d}u) \right], \tag{2.50}$$

where $\tilde{g}_{\theta_0}(\theta)$ is as given in (2.40).

Consequently, if $\mathscr{W}_{\rho_{\gamma\alpha}}(\delta_{u_0}, \mu_{\theta_0}) < \infty$ then

$$\lim_{\theta \to \theta_0} \sup_{t \in \mathbb{R}^+} \left| \mathbb{E}\left[\varphi(u_\theta(t; u_0)) - \varphi(u_{\theta_0}(t; u_0)) \right] \right| = 0. \tag{2.51}$$

Proof. Let $\hat{\alpha} > 0$ such that (2.39) holds for each $\alpha \in (0, \hat{\alpha}]$ and corresponding $\rho \in \Lambda$, and let us fix any such α and ρ. Fix also $\theta \in \Theta$, $u_0 \in X$ such that $f(u_0) < \infty$, with f as in (2.25), and let $\varphi : X \to \mathbb{R}$ be a ρ_α-Lipschitz function with Lipschitz constant denoted as L_φ. Thus, for every $t \in \mathbb{R}^+$ and coupling $\Gamma \in \mathscr{C}(P_t^\theta(u_0, \cdot), P_t^{\theta_0}(u_0, \cdot))$, we have

$$\begin{aligned} \left| \mathbb{E}\left[\varphi(u_\theta(t; u_0)) - \varphi(u_{\theta_0}(t; u_0)) \right] \right| &= \left| \int_X \varphi(u) P_t^\theta(u_0, \mathrm{d}u) - \int_X \varphi(\tilde{u}) P_t^{\theta_0}(u_0, \mathrm{d}\tilde{u}) \right| \\ &= \left| \int_X \left[\varphi(u) - \varphi(\tilde{u}) \right] \Gamma(\mathrm{d}u, \mathrm{d}\tilde{u}) \right| \leq L_\varphi \int_X \rho_\alpha(u, \tilde{u}) \Gamma(\mathrm{d}u, \mathrm{d}\tilde{u}). \end{aligned}$$

Taking the infimum over $\Gamma \in \mathscr{C}(P_t^\theta(u_0, \cdot), P_t^{\theta_0}(u_0, \cdot))$, we deduce that

$$\left| \mathbb{E}\left[\varphi(u_\theta(t; u_0)) - \varphi(u_{\theta_0}(t; u_0)) \right] \right| \leq L_\varphi \mathscr{W}_{\rho_\alpha} \left(P_t^\theta(u_0, \cdot), P_t^{\theta_0}(u_0, \cdot) \right). \tag{2.52}$$

Next, we take the supremum over $t \in \mathbb{R}^+$ and invoke Theorem 2.8 to further estimate the right-hand side. It thus follows from (2.39) with $\mu = \delta_{u_0}$ that

$$\sup_{t \in \mathbb{R}^+} \left| \mathbb{E}\left[\varphi(u_\theta(t; u_0)) - \varphi(u_{\theta_0}(t; u_0)) \right] \right| \leq L_\varphi \tilde{g}_{\theta_0}(\theta) \left[\mathscr{W}_{\rho_{\gamma\alpha}}(\delta_{u_0}, \mu_{\theta_0}) + f(u_0) + \int_X f(u) \mu_{\theta_0}(\mathrm{d}u) \right],$$

with $\tilde{g}_{\theta_0}(\theta)$ as given in (2.40). This shows (2.50). Clearly, (2.51) follows immediately from (2.50) and the assumption that $\mathscr{W}_{\rho_{\gamma\alpha}}(\delta_{u_0}, \mu_{\theta_0}) < \infty$. This concludes the proof. □

3. Numerical approximation of the 2D stochastic Navier–Stokes equations

We now turn to the application of the abstract results from the previous section to the 2D stochastic Navier–Stokes equations (SNSE) and a corresponding space-time numerical discretization. In Section 3.1, we introduce some preliminary material regarding the form of the 2D stochastic Navier–Stokes equations that we consider here, along with the specific space-time discretization to be analyzed. In Section 3.2, we verify the general set of assumptions from Theorem 2.1 for a suitable class Λ of distance-like functions, defined in (3.26) below, to prove Wasserstein contraction for the Markov semigroup generated by this discretization. Here, as mentioned above in the introduction, Section 1, we emphasize that the contraction coefficients obtained for the discretized system are *independent* of any discretization parameters. This fact is crucial for obtaining a weak convergence result for the numerical scheme as a consequence of Theorem 2.8, which we present later in Section 3.4. We also provide in Section 3.4 an estimate of the bias between the long time statistics of the discrete system and the continuous one as an application of Theorem 2.5. Before in Section 3.3, we present some pathwise finite-time error estimates that are used to verify the required assumption 2.5 from Theorem 2.5 and Theorem 2.8 for these last two results.

3.1 *Mathematical setting and moment bounds*

3.1.1 *Two-dimensional stochastic Navier–Stokes equations.* Let $\mathbb{T}^2 \simeq (\mathbb{R}/2\pi\mathbb{Z})^2$ be the two-dimensional torus. We consider the homogeneous Lebesgue space $\dot{L}^2 = \dot{L}^2(\mathbb{T}^2) = \{\xi \in L^2(\mathbb{T}^2) : \int_{\mathbb{T}^2} \xi(x)\mathrm{d}x = 0\}$, endowed with the standard inner product and norm of $L^2(\mathbb{T}^2)$, which we denote by $(\cdot,\cdot)$ and $|\cdot|$, respectively. Recall that any function $\xi \in \dot{L}^2$ can be written as the Fourier expansion $\xi(x) = \sum_{\kappa\in\mathbb{Z}^2\backslash\{\mathbf{0}\}} \hat{\xi}(\kappa)\, e^{i\kappa\cdot x}$, with $\hat{\xi}$ denoting the Fourier transform of ξ, given by $\hat{\xi}(\kappa) := (2\pi)^{-2}\int_{\mathbb{T}^2} e^{-i\kappa\cdot x}\xi(x)\mathrm{d}x$.

We also consider, for each $s \geq 0$, the homogeneous Sobolev space $\dot{H}^s = \dot{H}^s(\mathbb{T}^2) = \{\xi \in \dot{L}^2(\mathbb{T}^2) : \|\xi\|_{\dot{H}^s} < \infty\}$, where $\|\cdot\|_{\dot{H}^s}$ is the norm induced by the inner product $(\cdot,\cdot)_{\dot{H}^s}$, given by

$$(\xi_1,\xi_2)_{\dot{H}^s} = (2\pi)^2 \sum_{\kappa\in\mathbb{Z}^2\backslash\{\mathbf{0}\}} |\kappa|^{2s}\hat{\xi}_1(\kappa)\overline{\hat{\xi}_2(\kappa)},$$

where $\overline{\cdot}$ denotes complex conjugation. Clearly, for every $s_1 < s_2$, we have $\dot{H}^{s_2} \subset \dot{H}^{s_1}$. Moreover, note that $\dot{H}^0$ coincides with $\dot{L}^2$, with $|\xi| = \|\xi\|_{\dot{H}^0}$. Also, $|\nabla\xi| = \|\xi\|_{\dot{H}^1}$ and $|\Delta\xi| = \|\xi\|_{\dot{H}^2}$.

Fix a stochastic basis $\mathscr{S} = (\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$, i.e. a filtered probability space equipped with a finite family $\{W^k\}_{k=1}^d$ of standard independent real-valued Brownian motions on Ω that are adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$. We consider the stochastically forced 2D Navier–Stokes equations (SNSE) in vorticity form in $\mathbb{T}^2$ and driven by a white in time and colored in space additive noise, namely

$$\mathrm{d}\xi + (-\nu\Delta\xi + \mathbf{u}\cdot\nabla\xi)\,\mathrm{d}t = \sum_{k=1}^d \sigma_k\,\mathrm{d}W^k, \quad \mathbf{u} = \mathscr{K} * \xi, \tag{3.1}$$

where $\xi = \xi(\mathbf{x},t)$, $(\mathbf{x},t) \in \mathbb{T}^2 \times (0,\infty)$, represents the unknown random vorticity field; $\mathbf{u} = \mathbf{u}(\mathbf{x},t)$ represents the random velocity field, which is determined from the vorticity through the Biot–Savart kernel $\mathscr{K}$ in (3.1), so that $\nabla^\perp\cdot\mathbf{u} = (-\partial_y,\partial_x)\cdot(u_1,u_2) = \xi$ and $\nabla\cdot\mathbf{u} = 0$, see e.g. Majda & Bertozzi (2001). Note in particular that the Fourier coefficients of $\mathbf{u}$ and ξ satisfy the relation $\hat{\mathbf{u}}(\kappa) = -i\frac{\kappa^\perp}{|\kappa|^2}\hat{\xi}(\kappa)$ for any $\kappa = (\kappa_1,\kappa_2) \in \mathbb{Z}^2\backslash\{\mathbf{0}\}$, with $\kappa^\perp := (-\kappa_2,\kappa_1)$, which yields an explicit way of determining $\mathbf{u}$ from ξ in actual numerical implementations.

Moreover, $\sigma_1, \ldots, \sigma_d$ are given functions in $\dot{L}^2$. We will sometimes use the abbreviated notation $\sigma\, \mathrm{d}W$ for $\sum_{k=1}^d \sigma_k\, \mathrm{d}W^k$. Also, we assume that (3.1) is in nondimensional form, so that the parameter ν equals $(Re)^{-1}$, where Re denotes the Reynolds number associated to the fluid flow.

Equation (3.1) is sometimes also written in the following convenient functional form

$$\mathrm{d}\xi + (\nu A\xi + B(\xi, \xi))\, \mathrm{d}t = \sigma\, \mathrm{d}W, \tag{3.2}$$

where $A = (-\Delta) : \dot{H}^2 \to \dot{L}^2$, and $B : \dot{H}^1 \times \dot{H}^1 \to (\dot{H}^1)'$ is the bilinear mapping defined as $B(\xi, \xi) = \mathbf{u} \cdot \nabla\xi = (\mathscr{K} * \xi) \cdot \nabla\xi$. Here, $(\dot{H}^1)'$ denotes the dual space of $\dot{H}^1$. For each $s \geq 0$, we define the corresponding power of A as $A^s : D(A^s) \to \dot{L}^2$, given by $A^s\xi(x) = \sum_{\kappa\in\mathbb{Z}^2\setminus\{\mathbf{0}\}} |\kappa|^{2s}\hat{\xi}(\kappa)\, e^{i\kappa\cdot x}$, where $D(A^s) = \dot{H}^{2s}$. Notice that $|A^s\xi| = \|\xi\|_{\dot{H}^{2s}}$. Further, we recall that A is a positive and self-adjoint operator with compact inverse. As such, it possesses a nondecreasing sequence of positive eigenvalues $\{\lambda_k\}_{k\in\mathbb{N}}$ with $\lambda_k \sim k$ asymptotically, so that $\lambda_k \to \infty$ as $k \to \infty$, associated to a sequence of eigenfunctions $\{e_k\}_{k\in\mathbb{N}}$ that form an orthonormal basis of $\dot{L}^2$.

Regarding the noise term in (3.1), we adopt the following additional notation. For each $s \geq 0$, we denote by $\dot{\boldsymbol{H}}^s$ the d-fold product of $\dot{H}^s$, and define, for each $\sigma = (\sigma_1, \ldots, \sigma_d) \in \dot{\boldsymbol{H}}^s$, $\|\sigma\|_{\dot{H}^s}^2 := \sum_{k=1}^d \|\sigma_k\|_{\dot{H}^s}^2$. Similarly, we consider $\dot{\boldsymbol{L}}^2 = \dot{\boldsymbol{H}}^0$ and denote $|\sigma| := \|\sigma\|_{\dot{H}^0}$ for all $\sigma \in \dot{\boldsymbol{L}}^2$. We then set $\sigma W := \sum_{k=1}^d \sigma_k W^k$, so that, for any $\sigma \in \dot{\boldsymbol{H}}^s$, σW is a Brownian motion on $\dot{H}^s$ with covariance operator tQ_s, where $Q_s : \dot{H}^s \to \dot{H}^s$ is given by

$$Q_s\xi = \sum_{k=1}^d (\sigma_k, \xi)_{\dot{H}^s}\sigma_k, \quad \xi \in \dot{H}^s. \tag{3.3}$$

We notice that Q_s is a compact and symmetric operator with $\mathrm{Tr}(Q_s) = \|\sigma\|_{\dot{H}^s}^2$, where we recall that $\mathrm{Tr}(Q_s) = \sum_{j=1}^\infty (Q_s\tilde{e}_j, \tilde{e}_j)_{\dot{H}^s}$ for any orthonormal basis $\{\tilde{e}_j\}_{j\geq 1}$ of $\dot{H}^s$.

With a slight abuse of notation, we also regard a given $\sigma \in \dot{\boldsymbol{H}}^s$ as a mapping $\sigma : \mathbb{R}^d \to \dot{H}^s$, defined as $\sigma(w_1, \ldots, w_d) = \sum_{k=1}^d \sigma_k w_k$ for all $(w_1, \ldots, w_d) \in \mathbb{R}^d$. Clearly, σ is thus a bounded linear operator on $\mathbb{R}^d$ with operator norm bounded from above by $\|\sigma\|_{\dot{H}^s}$. Moreover, we denote by $\sigma^{-1} : range(\sigma) \to \mathbb{R}^d$ its corresponding pseudo-inverse, which is a bounded operator.[1]

In what follows, we will be interested in pathwise, i.e. probabilistically strong, solutions of (3.1), which are defined with respect to a fixed stochastic basis $\mathscr{S} = (\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$ as considered above. We have the following well-posedness result regarding this type of solutions.

PROPOSITION 3.1 Let $\mathscr{S} = (\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$ be a stochastic basis. Then, given any sequence $\{\sigma_k\}_{k=1}^d$ in $\dot{L}^2$ and any $\mathscr{F}_0$-measurable random variable $\xi_0 \in L^2(\Omega, \dot{L}^2)$, there exists a unique $\dot{L}^2$-valued random process ξ with

$$\xi \in L^2\left(\Omega; L^2_{loc}\left([0, \infty); \dot{H}^1\right) \cap C([0, \infty); \dot{L}^2)\right),$$

[1] Note that since $range(\sigma)$ is a finite dimensional subset of $\dot{L}^2$, then it is closed. This implies that the pseudo-inverse σ^{-1} is a bounded operator (see e.g. Sheffield, 1956, Theorem 3.8).

which is $\mathscr{F}_t$-adapted, solves (3.1) weakly in $\dot{L}^2$ and satisfies the initial condition $\xi(0) = \xi_0$ almost surely. Moreover, ξ depends continuously on the initial data, i.e. for each $\xi_0 \in \dot{L}^2$, the mapping $\xi_0 \mapsto \xi(t; \xi_0, \{W^k\}_{k=1}^d)$ is continuous in $\dot{L}^2$ for any $t \in [0, \infty)$ and any fixed realization $\{W^k(\cdot, \omega)\}_k$, $\omega \in \Omega$.

Within this additive noise setting, a proof of Proposition 3.1 is given by following the standard argument of defining a change of variables $\xi = \overline{\xi} + \zeta$, where

$$\frac{\mathrm{d}\overline{\xi}}{\mathrm{d}t} - \nu\Delta\overline{\xi} + (\mathscr{K} * (\overline{\xi} + \zeta)) \cdot \nabla(\overline{\xi} + \zeta) = 0 \quad \text{and} \quad \mathrm{d}\zeta - \nu\Delta\zeta\,\mathrm{d}t = \sigma\,\mathrm{d}W.$$

For each realization of ζ, $\overline{\xi}$ thus satisfies a deterministic equation for which one can show well-posedness by following similar arguments as for the 2D Navier–Stokes equations, see e.g. Constantin & Foias (1988); Temam (2001). Whereas ζ satisfies a linear SPDE whose well-posedness is well-established, see e.g. Da Prato & Zabczyk (2014). We remark, however, that well-posedness has also been established under much more general noise settings, see e.g. Mikulevicius & Rozovskii (2004); Glatt-Holtz & Ziane (2009).

With the notation introduced in Section 2.1 and Proposition 3.1, we denote the transition function associated to (3.1) by $P_t = P_t(\xi_0, \mathscr{O})$, for each $t \geq 0$, initial point $\xi_0 \in \dot{L}^2$ and Borel set $\mathscr{O} \in \mathscr{B}(\dot{L}^2)$, defined as

$$P_t(\xi_0, \mathscr{O}) := \mathbb{P}(\xi(t; \xi_0) \in \mathscr{O}), \tag{3.4}$$

where $\xi(t; \xi_0)$, $t \geq 0$, is the unique solution of (3.1) satisfying $\xi(0) = \xi_0$ almost surely, in the sense given in Proposition 3.1. The corresponding Markov semigroup P_t, $t \geq 0$, is defined for each $\varphi \in \mathscr{M}_b(\dot{L}^2)$ as

$$P_t\varphi(\xi_0) := \mathbb{E}\varphi(\xi(t; \xi_0)), \quad \xi_0 \in \dot{L}^2. \tag{3.5}$$

Since $\xi(\cdot; \xi_0)$ is continuous with respect to the initial data ξ_0, it follows that $\{P_t\}_{t\geq 0}$ is also a Feller Markov semigroup in $\dot{L}^2$. Namely, denoting by $\mathscr{C}_b(\dot{L}^2)$ the space of real-valued, bounded and continuous functions on $\dot{L}^2$, we have $P_t\varphi \in \mathscr{C}_b(\dot{L}^2)$ for every $\varphi \in \mathscr{C}_b(\dot{L}^2)$.

We recall that existence of an invariant measure μ_* with respect to the semigroup P_t, $t \geq 0$, is a well-established result, in fact valid for much more general noise structures than specified in (3.1), see e.g. Flandoli (1994). On the other hand, showing uniqueness of the invariant measure requires extra assumptions on the noise term, see e.g. Flandoli & Maslowski (1995); Da Prato & Zabczyk (1996); Mattingly (1999); Bricmont *et al.* (2001); E *et al.* (2001); Kuksin & Shirikyan (2001); Bricmont *et al.* (2002); Kuksin & Shirikyan (2002); Mattingly (2002); Shirikyan *et al.* (2002); Mattingly (2003); Hairer & Mattingly (2006, 2008, 2011a); Kuksin & Shirikyan (2012); Debussche (2013); Glatt-Holtz *et al.* (2017, 2018). Following a similar assumption from previous works, here we consider that the number d of stochastically forced directions in (3.1) is sufficiently large depending on the 'size' of the parameter ν and the coefficients σ_k, see (3.29) below. Our setup would also accommodate infinitely many driving noise terms provided they decay sufficiently fast to preserve sufficient smoothness in the solutions and with the appropriate modifications in the definition of σ.

We expect certain other noise structures to yield similar results regarding the space-time discretization (3.20) below. For example, we could consider the degenerate type of stochastic forcing as analyzed in Hairer & Mattingly (2006, 2008); Földes *et al.* (2015); Glatt-Holtz *et al.* (2018). This would however require more sophisticated Malliavin calculus techniques that we intend to pursue in future work, as we pointed out in Section 1.4 above.

Let us also recall a few basic inequalities and properties of the bilinear term $\mathbf{u}\cdot\nabla\xi$ in (3.1). For any divergence-free $\mathbf{u}\in(\dot{H}^1)^2$, it follows with integration by parts that

$$(\mathbf{u}\cdot\nabla\xi,\widetilde{\xi}) = -(\mathbf{u}\cdot\nabla\widetilde{\xi},\xi) \quad \text{for all } \xi,\widetilde{\xi}\in\dot{H}^1, \tag{3.6}$$

which implies the orthogonality property

$$(\mathbf{u}\cdot\nabla\xi,\xi) = ((\mathscr{K}*\xi)\cdot\nabla\xi,\xi) = 0 \quad \text{for all } \xi\in\dot{H}^1. \tag{3.7}$$

Moreover, the following inequalities follow by standard arguments involving Hölder and interpolation inequalities, see e.g. Constantin & Foias (1988); Temam (2001):

$$|\left((\mathscr{K}*\xi_1)\cdot\nabla\xi_2,\xi_3\right)| \le c|\xi_1||\nabla\xi_2||\xi_3|^{1-a}|\nabla\xi_3|^a, \quad a\in[0,1], \tag{3.8}$$

$$|\left((\mathscr{K}*\xi_1)\cdot\nabla\xi_2,\xi_3\right)| \le c|\xi_1|^{1/2}|\nabla\xi_1|^{1/2}|\nabla\xi_2||\xi_3|, \tag{3.9}$$

$$|\left(\nabla\left[(\mathscr{K}*\xi_1)\cdot\nabla\xi_2\right],\nabla\xi_2\right)| \le c|\xi_1|^{3/4}|\Delta\xi_1|^{1/4}|\nabla\xi_2||\xi_2|^{1/4}|\Delta\xi_2|^{3/4}, \tag{3.10}$$

for some positive absolute constant c, and for all ξ such that the norms above make sense.

REMARK 3.2 Throughout the next sections, we adopt the following convention regarding constants. With lower-case letters c, $\tilde{c}$, we denote a positive absolute constant, i.e. independent of any parameters whatsoever. Whereas with upper-case letters $C,\tilde{C},C_0,C_1,C_2$, we denote a positive constant that depends *at most* on the parameters $\nu,|\sigma|,|\nabla\sigma|,|A\sigma|$, the parameters $\varepsilon>0$ and $s\in(0,1]$ from the definition of the family of distances in (3.26) below, along with other parameters that are specific to certain statements. These will be made explicit within each statement. Most importantly, these constants will always be independent of any discretization parameters. Under this convention regarding their dependences, we allow the values of these constants to vary from line to line.

3.1.2 *Spectral Galerkin discretization.* We start by fixing some notation. As before, we denote the eigenvalues and eigenfunctions of $A=(-\Delta):\dot{H}^2\to\dot{L}^2$ by $\{\lambda_k\}_{k\in\mathbb{N}}$ and $\{e_k\}_{k\in\mathbb{N}}$, respectively. Then, for each $N\in\mathbb{N}$, we denote by $\Pi_N:\dot{L}^2\to\dot{L}^2$ the projection operator onto the subspace $\Pi_N\dot{L}^2$ of $\dot{L}^2$ given by the span of the first N eigenfunctions of $(-\Delta)$. Therefore, $I-\Pi_N$ is the projection operator onto the complement space $(I-\Pi_N)\dot{L}^2$. We have the following Poincaré-type inequality, see e.g. Constantin & Foias (1988); Temam (2001):

$$|\nabla(I-\Pi_N)\xi|^2 \ge \lambda_{N+1}|(I-\Pi_N)\xi|^2 \quad \text{for all } \xi\in\dot{H}^1. \tag{3.11}$$

The spectral Galerkin in space approximation of (3.1) in $\Pi_N\dot{L}^2$ is given by

$$\mathrm{d}\xi_N + [-\nu\Delta\xi_N + \Pi_N(\mathbf{u}_N\cdot\nabla\xi_N)]\,\mathrm{d}t = \Pi_N\sigma\,\mathrm{d}W, \quad \mathbf{u}_N = \mathscr{K}*\xi_N. \tag{3.12}$$

The existence and uniqueness of probabilistically strong solutions of (3.12) satisfying a given initial condition follows analogously to the proof of Proposition 3.1. For completeness, we state this result below.

PROPOSITION 3.3 Fix a stochastic basis $\mathscr{S} = (\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$. Then, given any family $\{\sigma_k\}_{k=1}^d$ of functions in $\dot{L}^2$ and any $\mathscr{F}_0$-measurable random variable $\xi_0 \in L^2(\Omega, \dot{L}^2)$, there exists a unique $\Pi_N \dot{L}^2$-valued random process ξ_N with

$$\xi_N \in L^2\Big(\Omega; C\Big([0,\infty); \Pi_N \dot{L}^2\Big)\Big),$$

which is $\mathscr{F}_t$-adapted, solves (3.12) weakly in $\dot{L}^2$ and satisfies the initial condition $\xi(0) = \Pi_N \xi_0$ almost surely. Moreover, ξ_N depends continuously on the initial data, i.e. for each $\xi_0 \in \dot{L}^2$, the mapping $\xi_0 \mapsto \xi_N(t; \Pi_N\xi_0, \{W^k\}_{k=1}^d)$ is continuous in $\dot{L}^2$ for any $t \in [0,\infty)$ and any fixed realization $\{W^k(\cdot,\omega)\}_k$, $\omega \in \Omega$.

We next state a collection of results regarding solutions of the Galerkin system (3.12) as well of the limiting system (3.1) that will be particularly useful in Section 3.3 and Section 3.4 below.

The following proposition presents some further moment bounds for solutions of the Galerkin scheme (3.12) and the fully continuous system (3.1). The proof follows from similar arguments as in Kuksin & Shirikyan (2012, Corollary 2.4.11, Proposition 2.4.12), where for handling the nonlinear term in each case we invoke (3.7), (3.10) and the following inequality, which follows similarly as in Kuksin & Shirikyan (2012, Lemma 2.1.20)

$$\left| \Big((\mathscr{K} * \xi) \cdot \nabla \xi, A^k \xi \Big) \right| \leq c |\xi|^{\frac{k+2}{2(k+1)}} \|\xi\|_{\dot{H}^1}^{\frac{k+2}{2k}} \|\xi\|_{\dot{H}^k}^{\frac{k(4k+1)-2}{2k(k+1)}},$$

for all $\xi \in \dot{H}^k$, $k \geq 2$.

PROPOSITION 3.4 Fix any $\xi_0 \in \dot{H}^1$ and $\sigma \in \dot{\boldsymbol{H}}^1$. Let $\xi_N(t)$, $t \geq 0$, be the solution of (3.12) satisfying $\xi_N(0) = \Pi_N \xi_0$ almost surely. Then, for every $T > 0$ and $m \in \mathbb{N}$, it holds

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[\sup_{t\in[0,T]} \left(|\nabla \xi_N(t)|^2 + \nu \int_0^t |\Delta \xi_N(s)|^2 \, \mathrm{d}s \right)^m \right] \leq C\Big(1 + |\xi_0|^{4m} + |\nabla \xi_0|^{2m}\Big), \tag{3.13}$$

for some constant $C = C(m, \nu, T, |\nabla\sigma|)$.

Moreover, given any $\xi_0 \in \dot{L}^2$ and $\sigma \in \dot{\boldsymbol{H}}^k$, $k \in \mathbb{Z}^+$, it follows that for every $T > 0$ and $m \in \mathbb{N}$ there exists $p = p(k)$ such that

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[\sup_{t\in[0,T]} \left(t^k \|\xi_N(t)\|_{\dot{H}^k}^2 + \nu \int_0^t s^k \|\xi_N(s)\|_{\dot{H}^{k+1}}^2 \, \mathrm{d}s \right)^m \right] \leq \widetilde{C}\left(1 + |\xi_0|^{2mp}\right), \tag{3.14}$$

for some constant $\widetilde{C} = \widetilde{C}(m, k, \nu, T, \|\sigma\|_{\dot{H}^k})$. More precisely, $p(0) = 1$, $p(1) = 2$ and $p(k) = (3k+1)(k+2)/(3k+2)$ for every $k \geq 2$.

Furthermore, let $\xi(t)$, $t \geq 0$, be the solution of (3.1) satisfying $\xi(0) = \xi_0$ almost surely. Then, analogous inequalities to (3.13) and (3.14) hold with $\xi_N(t)$ replaced by $\xi(t)$.

The next two propositions provide, respectively, some exponential moment bounds, and exponential Lyapunov inequalities for systems (3.12) and (3.1). The proofs are available within similar contexts in e.g. Hairer & Mattingly (2006, 2008); Kuksin & Shirikyan (2012); Debussche (2013); Glatt-Holtz (2014); Glatt-Holtz *et al.* (2017), while also following as entirely analogous continuous versions of the proofs of Proposition 3.15 and Proposition 3.11 below.

We note in particular that in inequalities (3.16) and (3.61) we obtain the supremum in time inside the expectation thanks to an argument involving exponential martingales combined with Doob's martingale inequality, which is described in detail later in the proof of Proposition 3.15. Note moreover that the terms $-\alpha|\sigma|^2 t$ in (3.16) and $\frac{n}{4}\ln(1-4\alpha\delta|\sigma|^2)$ in (3.61) imply an exponentially growing upper bound in t and n, respectively, for the exponential of the remaining two terms inside the parentheses in expected value.

PROPOSITION 3.5 Fix any $\sigma \in \dot{\boldsymbol{L}}^2$ and $\xi_0 \in \dot{L}^2$. Fix also $N \in \mathbb{N}$ and let $\xi_N(t)$, $t \geq 0$, be the solution of (3.20) satisfying $\xi_N(0) = \Pi_N \xi_0$ almost surely. Then, for all $\alpha \in \mathbb{R}$ satisfying

$$0 < \alpha \leq \frac{\nu}{2|\sigma|^2}, \tag{3.15}$$

the following inequality holds

$$\mathbb{E} \sup_{t \geq 0} \exp\left(\alpha|\xi_N(t)|^2 + \alpha\nu \int_0^t |\nabla \xi_N(s)|^2 \, ds - \alpha|\sigma|^2 t\right) \leq 2\exp\left(\alpha|\xi_0|^2\right). \tag{3.16}$$

Moreover, let $\xi(t)$, $t \geq 0$, be the solution of (3.1) satisfying $\xi(0) = \xi_0$ almost surely. Then, an analogous inequality to (3.16) holds with $\xi_N(t)$ replaced by $\xi(t)$.

PROPOSITION 3.6 Fix any $\sigma \in \dot{\boldsymbol{L}}^2$ and $\xi_0 \in \dot{L}^2$. Fix also $N \in \mathbb{N}$ and let $\xi_N(t)$, $t \geq 0$, be the solution of (3.20) satisfying $\xi_N(0) = \Pi_N \xi_0$ almost surely. Consider $\alpha \in \mathbb{R}$ satisfying (3.15). Then, the following inequality holds

$$\mathbb{E}\exp\left(\alpha|\xi_N(t)|^2\right) \leq \exp\left(\alpha\left(e^{-\nu t}|\xi_0|^2 + \frac{|\sigma|^2}{\nu}\right)\right) \qquad \text{for all } t \geq 0. \tag{3.17}$$

Moreover, let $\xi(t)$, $t \geq 0$, be the solution of (3.1) satisfying $\xi(0) = \xi_0$ almost surely. Then, an analogous inequality to (3.17) holds with $\xi_N(t)$ replaced by $\xi(t)$.

The following result shows Hölder regularity in time for solutions of the Galerkin system (3.12). We note that a similar result was shown in Carelli & Prohl (2012, Lemma 2.3) involving the velocity formulation of (3.1) subject to a suitable multiplicative noise structure, and resulting in Hölder regularity for the associated solution with respect to a weaker norm than presented here. A proof is included in Section B.

THEOREM 3.7 Fix any $\sigma \in \dot{\boldsymbol{H}}^1$ and $\xi_0 \in \dot{H}^2$. Let $\xi_N = \xi_N(t)$ be the solution of (3.12) satisfying $\xi_N(0) = \Pi_N \xi_0$ almost surely. Then, for every $T > 0$, $m \in \mathbb{N}$ and $\tilde{p} \in (0, 1/2)$,

$$\sup_{N \in \mathbb{N}} \mathbb{E}\left[|\xi_N(t) - \xi_N(s)|^m\right] \leq C|t-s|^{m\tilde{p}}\left(1 + |\xi_0|^{4m} + |\nabla \xi_0|^{2m}\right) \tag{3.18}$$

and

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[|\nabla\xi_N(t)-\nabla\xi_N(s)|^m\right] \leq C|t-s|^{m\tilde{p}}\Big(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}+|A\xi_0|^m\Big) \tag{3.19}$$

for all $s,t\in[0,T]$, where $C=C(m,\tilde{p},T,\nu,|\sigma|,|\nabla\sigma|)$. Moreover, let $\xi(t)$, $t\geq 0$, be the solution of (3.1) satisfying $\xi(0)=\xi_0$ almost surely. Then, analogous inequalities to (3.18) and (3.19) hold with $\xi(t)$ replaced by $\xi(t)$.

3.1.3 *Space-time discretization.* We now introduce, for each fixed time step $\delta>0$, a fully space-time discrete approximation of (3.1) given by a semi-implicit in time Euler discretization of the Galerkin system (3.12), namely

$$\begin{gathered}\xi_{N,\delta}^n = \xi_{N,\delta}^{n-1} + \delta\Big[\nu\Delta\xi_{N,\delta}^n - \Pi_N\Big(\mathbf{u}_{N,\delta}^{n-1}\cdot\nabla\xi_{N,\delta}^n\Big)\Big] + \sum_{k=1}^d \Pi_N\sigma_k\Big(W^k(t_n)-W^k(t_{n-1})\Big),\\ \mathbf{u}_{N,\delta}^{n-1} = \mathscr{K}*\xi_{N,\delta}^{n-1},\end{gathered} \tag{3.20}$$

where each $\xi_{N,\delta}^n$ represents the approximation of ξ_N, and thus of ξ, at time $t_n=n\delta$, for all $n\in\mathbb{N}$.

Since $\{W^k\}_{k=1}^d$ is a sequence of independent real-valued Brownian motions, we can write

$$W^k(t_n)-W^k(t_{n-1}) \overset{\mathscr{L}}{=} \eta_n^k\delta^{1/2} \tag{3.21}$$

for a sequence $\eta_n^k:\Omega\to\mathbb{R}$, $n\in\mathbb{N}$, $k=1,\dots,d$, of independent and identically distributed Gaussian random variables with mean zero and covariance 1.

As in (3.2) above, we adopt the abbreviated notation

$$\eta_n=(\eta_n^1,\dots,\eta_n^d),\quad \Pi_N\sigma=(\Pi_N\sigma_1,\dots,\Pi_N\sigma_d)\quad \text{and } \Pi_N\sigma\eta_n=\sum_{k=1}^d \Pi_N\sigma_k\eta_n^k,$$

so that (3.20) is compactly written as

$$\xi_{N,\delta}^n = \xi_{N,\delta}^{n-1} + \delta\Big[\nu\Delta\xi_{N,\delta}^n - \Pi_N\Big(\mathbf{u}_{N,\delta}^{n-1}\cdot\nabla\xi_{N,\delta}^n\Big)\Big] + \Pi_N\sigma\eta_n\delta^{1/2},\quad \mathbf{u}_{N,\delta}^{n-1}=\mathscr{K}*\xi_{N,\delta}^{n-1}. \tag{3.22}$$

We notice that, for each $n\in\mathbb{N}$, $\sigma\eta_n$ is a Gaussian random variable in $\dot{L}^2$ with zero mean and covariance operator given by Q_0 defined in (3.3).

Remark on notation: To avoid overburdening notation, in the subsequent sections we will frequently denote $\xi_{N,\delta}^n$ and $\mathbf{u}_{N,\delta}^n$ simply as ξ^n and $\mathbf{u}^n$, respectively, for any $n\in\mathbb{N}$.

The following proposition establishes pathwise well-posedness of (3.20) for a given initial data. Its proof follows by standard arguments that are analogous as in the deterministic case (see e.g. Foias *et al.*, 1991, Section 4.2), so we only present the main ideas.

PROPOSITION 3.8 Let $\mathscr{S}=(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P},\{W^k\}_{k=1}^d)$ be a stochastic basis. Then, given any family $\{\sigma_k\}_{k=1}^d$ of functions in $\dot{L}^2$ and any $\dot{L}^2$-valued and $\mathscr{F}_0$-measurable random variable $\xi_0\in L^2(\Omega,\dot{L}^2)$, there exists a unique $\Pi_N \dot{L}^2$-valued discrete random process $\{\xi_{N,\delta}^n\}_{n\in\mathbb{Z}^+}$ with $\xi_{N,\delta}^n \in L^2(\Omega, \dot{L}^2)$, for all $n \in \mathbb{Z}^+$, and which is $\{\mathscr{F}_{t_n}\}_{n\in\mathbb{N}}$-adapted, solves (3.20) in $\dot{L}^2$ and satisfies the initial condition $\xi_{N,\delta}^0 = \Pi_N \xi_0$ almost surely. Moreover, $\xi_{N,\delta}^n$ depends continuously on the initial data, i.e. for each $\xi_0 \in \dot{L}^2$, the mapping $\xi_0 \mapsto \xi_{N,\delta}^n(\Pi_N \xi_0, \{W^k\}_{k=1}^d)$ is continuous in $\dot{L}^2$ for any $n \in \mathbb{Z}^+$ and any fixed realization $\{W^k(\cdot,\omega)\}_k$, $\omega \in \Omega$.

Proof. Since (3.22) is a linear equation with respect to ξ^n on the finite-dimensional space $\Pi_N \dot{L}^2$, existence of pathwise strong solutions follows immediately from uniqueness. For showing uniqueness, suppose that for a given fixed realization of $\{W^k\}_{k=1}^d$ and given $\xi^{n-1} \in \Pi_N \dot{L}^2$, there exist two solutions ξ_1^n and ξ_2^n of (3.20). Then $\zeta_n = \xi_1^n - \xi_2^n$ satisfies

$$\zeta_n = \nu\delta\Delta\zeta_n - \delta\Pi_N\Big(\Big(\mathscr{K} * \xi_N^{n-1}\Big)\cdot\nabla\zeta_n\Big). \tag{3.23}$$

Taking the inner product of (3.23) with ζ_n in $\dot{L}^2$ and invoking (3.7), it follows that $|\zeta_n|^2 + \nu\delta|\nabla\zeta_n|^2 = 0$, which implies that $\zeta_n = 0$, and thus $\xi_1^n = \xi_2^n$.

The fact that $\{\xi^n\}_{n\in\mathbb{N}}$ is $\{\mathscr{F}_{t_n}\}_{n\in\mathbb{N}}$-adapted follows immediately from (3.20), since ξ_0 is $\mathscr{F}_0$-measurable and $\{W^k(t_n)\}_{k=1}^d$, $n \in \mathbb{N}$, is $\{\mathscr{F}_{t_n}\}_{n\in\mathbb{N}}$-adapted. The continuity of $\{\xi^n\}_{n\in\mathbb{Z}^+}$ with respect to initial data for a fixed realization of $\{W^k\}_{k=1}^d$ follows by considering the equation for the difference $\xi^n(\Pi_N\xi_0) - \xi^n(\Pi_N\tilde{\xi}_0)$ for any $\xi_0, \tilde{\xi}_0 \in \dot{L}^2$, and proceeding with standard estimates by invoking (3.8) and (3.34) below. We omit further details. $\square$

For each fixed time step $\delta > 0$ and number N of Galerkin modes, we denote the Markov transition function associated to n steps of the discrete scheme (3.22) by $P_n^{N,\delta} = P_n^{N,\delta}(\xi_0, \mathscr{O})$, $\xi_0 \in \dot{L}^2$, $\mathscr{O} \in \mathscr{B}(\dot{L}^2)$. This is defined as

$$P_n^{N,\delta}(\xi_0, \mathscr{O}) := \mathbb{P}\Big(\xi_{N,\delta}^n(\Pi_N\xi_0) \in \mathscr{O}\Big), \tag{3.24}$$

where $\xi_{N,\delta}^n(\Pi_N\xi_0)$ is the unique solution of (3.1) starting from the initial datum $\Pi_N\xi_0$, in the sense given in Proposition 3.8. The corresponding Markov semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, is thus defined for each $\varphi \in \mathscr{M}_b(\dot{L}^2)$ as

$$P_n^{N,\delta}\varphi(\xi_0) := \mathbb{E}\varphi\Big(\xi_{N,\delta}^n(\Pi_N\xi_0)\Big), \quad \xi_0 \in \dot{L}^2, \quad n \in \mathbb{Z}^+. \tag{3.25}$$

Similarly as pointed out in Section 3.1.1 for the Markov semigroup P_t, $t \geq 0$, it follows as a consequence of the continuity of the solution $\xi_{N,\delta}^n(\Pi_N\xi_0)$ with respect to the initial datum $\Pi_N\xi_0$, guaranteed by Proposition 3.8 above, that $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, is a Feller Markov semigroup in $\dot{L}^2$.

3.2 *Discretization-uniform Wasserstein contraction*

In this section, we apply Theorem 2.1 to show a Wasserstein contraction result for the Markov semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, associated to the numerical scheme (3.22), defined in (3.25), for any fixed parameters $N \in \mathbb{N}$, $\delta > 0$. Within the setting of Theorem 2.1, we consider $(X, \|\cdot\|) = (\dot{L}^2, |\cdot|)$, $\mathscr{I} = \delta\mathbb{Z}^+$ and $\{P_t\}_{t\in\mathscr{I}}$ given by $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$. Moreover, we consider the class of distance-like functions $\Lambda = \{\rho_{\varepsilon,s} :$

$\varepsilon > 0,\ 0 < s \leq 1\}$, with each $\rho_{\varepsilon,s}$ defined as

$$\rho_{\varepsilon,s}(\xi,\widetilde{\xi}) = 1 \wedge \frac{|\xi - \widetilde{\xi}|^s}{\varepsilon}, \quad \xi,\widetilde{\xi} \in \dot{L}^2. \tag{3.26}$$

Here, in fact, each $\rho_{\varepsilon,s}$ is an actual metric on $\dot{L}^2$, as it can be easily verified. The parameter ε is appropriately tuned so as to produce a local contraction in (3.59) below, in view of assumption 2.1 from Theorem 2.1. Thus, in a certain sense, ε can be understood as representing the small spatial scales in the dynamics specified by (3.20) and (3.1), respectively.

As in (2.8), for each $a > 0$ we denote the corresponding Lyapunov-weighted version of $\rho_{\varepsilon,s}$ by $\rho_{\varepsilon,s,a}$, defined as

$$\rho_{\varepsilon,s,a}(\xi,\widetilde{\xi}) = \rho_{\varepsilon,s}(\xi,\widetilde{\xi})^{1/2} \exp\left(a|\xi|^2 + a|\widetilde{\xi}|^2\right), \quad \xi,\widetilde{\xi} \in \dot{L}^2. \tag{3.27}$$

Moreover, we denote the Wasserstein-like extensions to $\mathrm{Pr}(\dot{L}^2)$ corresponding to $\rho_{\varepsilon,s}$ and $\rho_{\varepsilon,s,a}$, as defined in (2.2), by $\mathscr{W}_{\varepsilon,s}$ and $\mathscr{W}_{\varepsilon,s,a}$, respectively.

The validity of assumptions (A1)–(A3) from Theorem 2.1 is verified in Proposition 3.11, Proposition 3.12 and Proposition 3.13 below. This leads us to the Wasserstein contraction result Proposition 3.9 below, whose proof we postpone to the end of this section. With the purpose of later applying Theorem 2.8 to yield uniform weak convergence of the numerical scheme (3.20) towards the continuous system (3.1), we state Theorem 3.9 in terms of a suitable continuous family of Markov kernels corresponding to the discrete semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$. Namely, we define for each $t \in \mathbb{R}^+$

$$\mathscr{P}_t^{N,\delta} := P_n^{N,\delta} \quad \text{if } t \in [n\delta, (n+1)\delta),\ n \in \mathbb{Z}^+. \tag{3.28}$$

We notice that the family $\mathscr{P}_t^{N,\delta}$, $t \in \mathbb{R}^+$, may not define a Markov semigroup. However, the semigroup property is not required in the general weak convergence result, Theorem 2.8.

THEOREM 3.9 Fix $\delta_0 > 0$. Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \dot{\boldsymbol{L}}^2$ such that

$$\Pi_K \dot{L}^2 \subset range(\sigma), \quad \text{and} \quad \lambda_{K+1} \geq \frac{c}{\nu} \max\left\{\frac{1}{\delta_0}, \frac{\delta_0^2|\sigma|^4}{\nu^3}, \frac{|\sigma|^4}{\nu^5}\right\} \tag{3.29}$$

for some absolute constant $c > 0$. For each $N \in \mathbb{N}$ and $0 < \delta \leq \delta_0$, let $\mathscr{P}_t^{N,\delta}$, $t \in \mathbb{R}^+$, be the corresponding family of Markov kernels defined in (3.28). Then, for every $m > 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exist $\varepsilon > 0$, $s \in (0,1]$, $T > 0$, and constants $C_1, C_2 > 0$ for which the following holds

$$\sup_{N\in\mathbb{N},\, 0<\delta\leq\delta_0} \mathscr{W}_{\varepsilon,s,\alpha}(\mu\mathscr{P}_t^{N,\delta}, \tilde{\mu}\mathscr{P}_t^{N,\delta}) \leq C_1\, e^{-tC_2}\mathscr{W}_{\varepsilon,s,\alpha/m}(\mu,\tilde{\mu}) \tag{3.30}$$

for every $\mu, \tilde{\mu} \in \mathrm{Pr}(\dot{L}^2)$ and $t \geq T$.

In view of Remark 2.3, Theorem 3.9 together with the Feller property of $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, implies the existence of a unique associated invariant measure. We state this result below.

COROLLARY 3.10 Consider the assumptions of Theorem 3.9. Then, for each fixed discretization parameters $N \in \mathbb{N}$ and $\delta > 0$, there exists a unique invariant measure $\mu_*^{N,\delta}$ of the discrete Markov semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, and consequently of $\mathscr{P}_t^{N,\delta}$, $t \in \mathbb{R}^+$.

Proof. The uniqueness of the invariant measure follows immediately from inequality (3.30). For the existence, as recalled in Remark 2.3, it follows similarly as in Hairer *et al.* (2011, Corollary 4.11) that it suffices to show there exists a complete metric $\tilde{\rho}$ on $\dot{L}^2$ such that $\tilde{\rho} \leq \sqrt{\rho_{\varepsilon,s}}$ and for which $\{P_n^{N,\delta}\}_{n\in\mathbb{Z}^+}$ is a Feller semigroup on $(\dot{L}^2, \tilde{\rho})$. Here, $\varepsilon > 0$ and $s \in (0, 1]$ are any parameters such that (3.30) holds. This is achieved, for example, by $\tilde{\rho} = \sqrt{\rho_{\varepsilon,s}} = \rho_{\sqrt{\varepsilon},s/2}$, which is a metric on $\dot{L}^2$ that is equivalent to the distance induced by the norm $|\cdot|$, so that the known Feller property of $\{P_n^{N,\delta}\}_{n\in\mathbb{Z}^+}$ on $(\dot{L}^2, |\cdot|)$ also holds in $(\dot{L}^2, \sqrt{\rho_{\varepsilon,s}})$. Clearly, if $\mu_*^{N,\delta}$ is an invariant measure for $\{P_n^{N,\delta}\}_{n\in\mathbb{Z}^+}$, then from the definition (3.28) it follows immediately that $\mu_*^{N,\delta}$ is also an invariant measure for $\{\mathscr{P}_t^{N,\delta}\}_{t\in\mathbb{R}^+}$. □

To prove Theorem 3.9, we start by verifying the existence of an exponential Lyapunov structure as in assumption 2.1 of Theorem 2.1.

PROPOSITION 3.11 Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \leq \delta_0$, $\sigma \in \dot{\boldsymbol{L}}^2$ and $\xi_0 \in \dot{L}^2$. Let $\{\xi_{N,\delta}^n\}_{n\in\mathbb{Z}^+}$ be the solution of (3.22) corresponding to the parameters N, δ and satisfying $\xi_{N,\delta}^0 = \Pi_N \xi_0$ almost surely. Then, for all $\alpha \in \mathbb{R}$ satisfying

$$0 < \alpha \leq \frac{\nu}{4|\sigma|^2}, \tag{3.31}$$

it holds that

$$\mathbb{E} \exp\left(\alpha |\xi_{N,\delta}^n|^2\right) \leq \exp\left(\alpha\left(\frac{2|\xi_0|^2}{(1+\nu\lambda_1\delta)^n} + C\right)\right) \quad \text{for all } n \in \mathbb{Z}^+, \tag{3.32}$$

for some positive constant C depending only on $\nu, |\sigma|, \delta_0$.

Consequently, recalling the definition of the Markov semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, in (3.25), it follows that for all $n \in \mathbb{Z}^+$

$$P_n^{N,\delta} \exp\left(\alpha|\xi_0|^2\right) \leq \exp\left(\alpha\left(ce^{-n\delta\widetilde{C}}|\xi_0|^2 + C\right)\right), \tag{3.33}$$

where $\widetilde{C}$ is a positive constant depending only on ν, δ_0.

Proof. Throughout the proof we adopt the simplified notation $\xi_{N,\delta}^j = \xi^j$, $j \in \mathbb{Z}^+$, mentioned in Section 3.1.3 above.

Fix $n \in \mathbb{N}$. For each $j \in \{1, \ldots, n\}$, we take the inner product of the first equation in (3.22) with ξ^j in $\dot{L}^2$ and invoke the Hilbert space identity

$$2(\xi - \widetilde{\xi}, \xi) = |\xi|^2 + |\xi - \widetilde{\xi}|^2 - |\widetilde{\xi}|^2 \quad \text{for all } \xi, \widetilde{\xi} \in \dot{L}^2, \tag{3.34}$$

together with the orthogonality property (3.7), to obtain that

$$|\xi^j|^2 + |\xi^j - \xi^{j-1}|^2 - |\xi^{j-1}|^2 + 2\nu\delta|\nabla\xi^j|^2 = 2\delta^{1/2}(\Pi_N\sigma\eta_j, \xi^j) = 2\delta^{1/2}(\sigma\eta_j, \xi^j). \tag{3.35}$$

In view of obtaining a well-defined martingale in (3.38) below, we add and subtract $2\delta^{1/2}(\sigma\eta_j, \xi^{j-1})$ in the right-hand side and estimate

$$
\begin{aligned}
2\delta^{1/2}\Big(\sigma\eta_j, \xi^j\Big) &= 2\delta^{1/2}\Big(\sigma\eta_j, \xi^j - \xi^{j-1}\Big) + 2\delta^{1/2}\Big(\sigma\eta_j, \xi^{j-1}\Big) \\
&\leq |\xi^j - \xi^{j-1}|^2 + \delta|\sigma\eta_j|^2 + 2\delta^{1/2}\Big(\sigma\eta_j, \xi^{j-1}\Big),
\end{aligned}
$$

so that, from (3.35) and Poincaré inequality,

$$
(1+\nu\delta)|\xi^j|^2 + \nu\delta|\nabla\xi^j|^2 \leq |\xi^j|^2 + 2\nu\delta|\nabla\xi^j|^2 \leq |\xi^{j-1}|^2 + \delta|\sigma\eta_j|^2 + 2\delta^{1/2}\Big(\sigma\eta_j, \xi^{j-1}\Big), \tag{3.36}
$$

for all $j \in \{1, \ldots, n\}$.

Fix $m \in \mathbb{N}$ with $m \geq n$. Denoting $b := (1+\nu\delta)^{-1}$, we obtain after multiplying both sides of (3.36) by b^{m-j+1} and summing over $j = 1, \ldots, n$ that

$$
b^{m-n}|\xi^n|^2 + \nu\delta\sum_{j=1}^{n} b^{m-j+1}|\nabla\xi^j|^2 \leq b^m|\xi_0|^2 + \sum_{j=1}^{n}\delta b^{m-j+1}|\sigma\eta_j|^2 + M_n, \tag{3.37}
$$

where $\{M_n\}_{n\in\mathbb{N}}$ is the martingale defined as

$$
M_n := 2\delta^{1/2}\sum_{j=1}^{n} b^{m-j+1}\Big(\sigma\eta_j, \xi^{j-1}\Big) \quad \text{for all } n \in \mathbb{N}, \tag{3.38}
$$

with corresponding quadratic variation given by

$$
\langle M\rangle_n = 4\delta\sum_{j=1}^{n}\sum_{k=1}^{d} b^{2(m-j+1)}\Big(\sigma_k, \xi^{j-1}\Big)^2. \tag{3.39}
$$

We then estimate $\langle M\rangle_n$ as

$$
\begin{aligned}
\langle M\rangle_n &\leq 4\delta|\sigma|^2\sum_{j=1}^{n} b^{2(m-j+1)}|\xi^{j-1}|^2 = 4\delta|\sigma|^2\sum_{j=0}^{n-1} b^{2(m-j)}|\xi^j|^2 \\
&\leq 4\delta|\sigma|^2\left(b^{2m}|\xi_0|^2 + \sum_{j=1}^{n-1} b^{2(m-j)}|\nabla\xi^j|^2\right) \\
&\leq 4\delta|\sigma|^2\left(b^{m+1}|\xi_0|^2 + \sum_{j=1}^{n} b^{m-j+1}|\nabla\xi^j|^2\right),
\end{aligned}
$$

where in the last line we used that $b \leq 1$ and $m \geq n \geq 1$. Thus, for all $\alpha \in \mathbb{R}$ satisfying (3.31) we obtain that

$$\alpha \langle M \rangle_n \leq b^m |\xi_0|^2 + \nu\delta \sum_{j=1}^{n} b^{m-j+1} |\nabla \xi^j|^2. \tag{3.40}$$

Now, adding and subtracting $\alpha\langle M\rangle_n$ to the right-hand side of (3.37) and invoking (3.40) it follows that

$$b^{m-n}|\xi^n|^2 \leq 2b^m|\xi_0|^2 + \sum_{j=1}^{n} \delta b^{m-j+1}|\sigma\eta_j|^2 + M_n - \alpha\langle M\rangle_n. \tag{3.41}$$

Multiplying by α, taking exponentials and expected values on both sides of (3.41) and applying Hölder's inequality, we deduce that

$$\begin{aligned}\mathbb{E}\exp\left(\alpha b^{m-n}|\xi^n|^2\right) &\leq \exp\left(2\alpha b^m|\xi_0|^2\right)\mathbb{E}\left[\left(\prod_{j=1}^{n}\exp\left(\alpha\delta b^{m-j+1}|\sigma\eta_j|^2\right)\right)\exp\left(\alpha M_n - \alpha^2\langle M\rangle_n\right)\right] \\ &\leq \exp\left(2\alpha b^m|\xi_0|^2\right)\left[\prod_{j=1}^{n}\mathbb{E}\exp\left(2\alpha\delta b^{m-j+1}|\sigma\eta_j|^2\right)\right]^{1/2}\left(\mathbb{E}\widetilde{M}_n\right)^{1/2},\end{aligned} \tag{3.42}$$

where

$$\widetilde{M}_n = \exp\left(2\alpha M_n - 2\alpha^2\langle M\rangle_n\right), \tag{3.43}$$

and we used the independence of the random variables $\sigma\eta_j$, $j = 1, \ldots, n$, to write the second factor in the right-hand side of (3.42).

From (3.38), let us denote $z_n := b^{m-n+1}(\sigma\eta_n, \xi^{n-1})$, and consider the regular conditional probability of z_n given $\mathscr{F}_{t_{n-1}}$, i.e. $\mu_n(\omega, A) := \mathbb{P}(z_n(\omega) \in A \,|\, \mathscr{F}_{t_{n-1}}) = \mathbb{E}[\mathbb{1}_{z_n^{-1}(A)} \,|\, \mathscr{F}_{t_{n-1}}]$, for $\omega \in \Omega$, $A \in \mathscr{B}(\mathbb{R})$, see e.g. Dudley (2002, Section 10.2). It is not difficult to show that, for each fixed $\omega \in \Omega$, $\mu_n(\omega, \cdot)$ is a Gaussian probability measure on $\mathscr{B}(\mathbb{R})$ with zero mean and variance $b^{2(m-n+1)}(Q_0\xi^{n-1}, \xi^{n-1}) = \sum_{k=1}^{d} b^{2(m-n+1)}(\sigma_k, \xi^{n-1})^2$, where Q_0 is as defined in (3.3). Using this fact, one can easily show that $\{\widetilde{M}_n\}_{n\in\mathbb{N}}$ is a martingale with respect to the filtration $\{\mathscr{F}_{t_n}\}_{n\in\mathbb{N}}$, and $\mathbb{E}\widetilde{M}_n = 1$ for all n (see e.g. Lamba *et al.*, 2007, Appendix). Moreover, since $\sigma\eta_j \sim \mathscr{N}(0, Q_0)$, $j = 1, \ldots, n$, from a general result on Gaussian probability measures on Hilbert spaces (Da Prato & Zabczyk, 2014, Proposition 2.17) it follows that

$$\mathbb{E}\exp(\gamma|\sigma\eta_j|^2) \leq \frac{1}{(1 - 2\gamma|\sigma|^2)^{1/2}} \quad \text{for all } \gamma < \frac{1}{2|\sigma|^2}, \tag{3.44}$$

where we recall that $|\sigma|^2 = \operatorname{Tr}(Q_0)$. In particular, since $\alpha \leq \nu/(4|\sigma|^2)$ by assumption (3.31), and since $b^{m-j+1} \leq b = 1/(1+\nu\delta)$, we have that (3.44) holds with $\gamma = 2\alpha\delta b^{m-j+1}$. Thus, from (3.42),

$$\mathbb{E}\exp\left(\alpha b^{m-n}|\xi^n|^2\right) \leq \exp\left(2\alpha b^m |\xi_0|^2\right)\prod_{j=1}^{n}\frac{1}{(1-4\alpha\delta b^{m-j+1}|\sigma|^2)^{1/4}} \quad \text{for all } m \geq n \geq 1. \tag{3.45}$$

In particular, if $m = n$ then

$$\begin{aligned}\mathbb{E}\exp\left(\alpha|\xi^n|^2\right) &\leq \exp\left(2\alpha b^n |\xi_0|^2\right)\prod_{j=1}^{n}\frac{1}{(1-4\alpha\delta b^{j}|\sigma|^2)^{1/4}}\\ &= \exp\left(2\alpha b^n |\xi_0|^2\right)\exp\left(-\frac{1}{4}\sum_{j=1}^{n}\ln\left(1-4\alpha\delta b^j|\sigma|^2\right)\right).\end{aligned} \tag{3.46}$$

Since $-\ln(1-x) \leq x(1-x)^{-1}$ for all $x \in (0,1)$, we obtain

$$-\frac{1}{4}\sum_{j=1}^{n}\ln\left(1-4\alpha\delta b^j|\sigma|^2\right) \leq \frac{1}{4}\sum_{j=1}^{n}\frac{4\alpha\delta b^j|\sigma|^2}{1-4\alpha\delta b^j|\sigma|^2}. \tag{3.47}$$

Moreover, since $\alpha \leq \nu/(4|\sigma|^2)$ and $b^j \leq b = 1/(1+\nu\delta)$, it follows that $[1-4\alpha\delta b^j|\sigma|^2]^{-1} \leq 1+\nu\delta$, so that

$$\begin{aligned}\frac{1}{4}\sum_{j=1}^{n}\frac{4\alpha\delta b^j|\sigma|^2}{1-4\alpha\delta b^j|\sigma|^2} &\leq (1+\nu\delta)\alpha\delta|\sigma|^2\sum_{j=1}^{n}b^j\\ &\leq (1+\nu\delta)\alpha\delta|\sigma|^2\frac{b}{1-b} = (1+\nu\delta)\frac{\alpha|\sigma|^2}{\nu} \leq (1+\nu\delta_0)\frac{\alpha|\sigma|^2}{\nu}.\end{aligned} \tag{3.48}$$

Therefore, from (3.46)–(3.48), it follows that

$$\mathbb{E}\exp\left(\alpha|\xi^n|^2\right) \leq \exp\left(\alpha(2b^n|\xi_0|^2 + C)\right),$$

where $C = (1+\nu\delta_0)|\sigma|^2/(\nu)$. This shows (3.32).

For the remaining inequality, (3.33), we use the fact that for any constant $0 < a < 1$ we have $\ln(1+x) \geq ax$ for all $x \in [0,(1/a)-1]$. Since $\delta \leq \delta_0$, we take $a := 1/(1+\nu\delta_0)$ and obtain that $\ln(1+\nu\delta) \geq \nu\delta/(1+\nu\delta_0)$, so that

$$\frac{2}{(1+\nu\delta)^n} = 2\exp\left(-n\ln(1+\nu\delta)\right) \leq 2\exp\left(-\frac{\nu}{1+\nu\delta_0}n\delta\right). \tag{3.49}$$

From (3.32) and (3.49), it thus follows that, for every $0 < \alpha \leq \nu/(4|\sigma|^2)$,

$$P_n^{N,\delta} \exp\left(\alpha|\xi_0|^2\right) = \mathbb{E}\exp\left(\alpha|\xi^n|^2\right) \leq \exp\left(\alpha\left(\frac{2|\xi_0|^2}{(1+\nu\delta)^n} + C\right)\right)$$
$$\leq \exp\left(\alpha\left(2\exp\left(-\frac{\nu}{1+\nu\delta_0}n\delta\right)|\xi_0|^2 + C\right)\right),$$

which shows (3.33) and concludes the proof. □

For showing the remaining assumptions (A1) and (A3) from Theorem 2.1, we follow a similar asymptotic coupling strategy from previous works, see e.g. Bricmont *et al.* (2001); E *et al.* (2001); Kuksin & Shirikyan (2001); E & Liu (2002); Hairer (2002); Kuksin & Shirikyan (2002); Mattingly (2002); Mattingly (2003); Debussche & Odasso (2005); Hairer & Mattingly (2006, 2008); Hairer *et al.* (2011); Hairer & Mattingly (2011a); Kuksin & Shirikyan (2012); Földes *et al.* (2015); Butkovsky *et al.* (2020). The idea consists in introducing the following modified equation for a given $\xi_0 \in \dot{L}^2$ and corresponding solution $\xi^n_{N,\delta} = \xi^n_{N,\delta}(\Pi_N\xi_0) = \xi^n(\Pi_N\xi_0)$, $n \in \mathbb{N}$, of (3.20). Namely, we consider $\widetilde{\xi}^n_{N,\delta} = \widetilde{\xi}^n$, $n \in \mathbb{N}$, satisfying

$$\widetilde{\xi}^n = \widetilde{\xi}^{n-1} + \delta\left[\nu\Delta\widetilde{\xi}^n - \Pi_N(\widetilde{\mathbf{u}}^{n-1}\cdot\nabla\widetilde{\xi}^n) - \beta\Pi_K\left(\widetilde{\xi}^n - \xi^n(\Pi_N\xi_0)\right)\right] + \sum_{k=1}^d \Pi_N\sigma_k\nu(W^k(t_n) - W^k(t_{n-1})), \quad (3.50)$$

$$\widetilde{\mathbf{u}}^{n-1} = \mathscr{K} * \widetilde{\xi}^{n-1}. \quad (3.51)$$

Here, the extra term $-\beta\delta\Pi_K(\widetilde{\xi}^n - \xi^n(\Pi_N\xi_0))$ has the purpose of enforcing a suitable control over 'large' scales, with $K \in \mathbb{N}$ representing the number of controlled modes, and $\beta > 0$ a control parameter, both to be appropriately chosen in (3.77) below.

Analogously as in Proposition 3.8, we can show that system (3.50)–(3.51) is well-posed in the pathwise sense. We omit the technical details. Therefore, for each $N \in \mathbb{N}$, $\delta > 0$ and $\xi_0 \in \dot{L}^2$, we may define

$$\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0, \mathscr{O}) = \mathbb{P}\left(\widetilde{\xi}^n_{N,\delta}(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0) \in \mathscr{O}\right) \quad \text{for all } n \in \mathbb{Z}^+,\ \widetilde{\xi}_0 \in \dot{L}^2 \text{ and } \mathscr{O} \in \mathscr{B}(\dot{L}^2), \quad (3.52)$$

where $\widetilde{\xi}^n_{N,\delta}(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0)$ is the unique (strong) solution of (3.50)–(3.51) with respect to a fixed stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$, and which satisfies the initial condition $\widetilde{\xi}^0_{N,\delta} = \Pi_N\widetilde{\xi}_0$ almost surely. Moreover, for every bounded and measurable function $\varphi : \dot{L}^2 \to \mathbb{R}$, we denote

$$\widetilde{P}_n^{N,\delta,\xi_0}\varphi(\widetilde{\xi}_0) = \mathbb{E}\varphi\left(\widetilde{\xi}^n_{N,\delta}(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0)\right) \quad \text{for all } n \in \mathbb{Z}^+ \text{ and } \widetilde{\xi}_0 \in \dot{L}^2. \quad (3.53)$$

Given any $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$, the idea consists in utilizing the family $\widetilde{P}_n^{N,\delta,\xi_0}$, $n \in \mathbb{Z}^+$, to estimate the Wasserstein distance $\mathscr{W}_{\varepsilon,s}$ between $P_n^{N,\delta}(\xi_0, \cdot)$ and $P_n^{N,\delta}(\widetilde{\xi}_0, \cdot)$ as

$$\mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot), P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) \leq \mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot), \widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot)\Big) + \mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot), P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big), \tag{3.54}$$

which holds since $\mathscr{W}_{\varepsilon,s}$ is a metric in $\mathrm{Pr}(\dot{L}^2)$. We then estimate each term on the right-hand side of (3.54) by analyzing system (3.50)–(3.51) under two different perspectives. The first term is estimated by establishing a suitable contraction between the solution $\widetilde{\xi}_{N,\delta}^n(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0)$ of (3.50)–(3.51) and the solution $\xi_{N,\delta}^n(\Pi_N\xi_0)$ of (3.20). This is possible due to the presence of the control term $-\beta\Pi_K(\widetilde{\xi}_{N,\delta}^n - \xi_{N,\delta}^n(\Pi_N\xi_0))$ in (3.50), and provided the number $K \in \mathbb{N}$ of controlled modes and the tuning parameter $\beta > 0$ are chosen sufficiently large (see (3.77) below).

For the second term in the right-hand side of (3.54), due to uniqueness of pathwise strong solutions of (3.20) we deduce that the solution $\widetilde{\xi}_{N,\delta}^n(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0, W)$ of (3.50)–(3.51) corresponding to the Wiener process W coincides with the solution $\xi_{N,\delta}^n(\Pi_N\widetilde{\xi}_0; \widehat{W})$ of (3.20) corresponding to the following shifted process

$$\widehat{W}(t) = W(t) + \int_0^t \sum_{j=1}^{\infty} \psi_j \mathbb{1}_{[t_{j-1},t_j)}(\tau)\, \mathrm{d}\tau, \tag{3.55}$$

where

$$\psi_j = -\beta\sigma^{-1}\Pi_K\Big(\widetilde{\xi}_{N,\delta}^j\Big(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0, W\Big) - \xi_{N,\delta}^j(\Pi_N\xi_0; W)\Big) \quad \forall j. \tag{3.56}$$

Here we recall that σ^{-1} denotes the pseudo-inverse of σ (see Section 3.1.1). The second term in the right-hand side of (3.54) can then be estimated by the *total variation distance* (see (3.89) below) between the laws of the processes W and $\widehat{W}$. This is in turn estimated via a Girsanov-type result. We note carefully that in order to have the expression in (3.56) well-defined, particularly in what concerns the domain of definition of σ^{-1}, we assume that $\Pi_K\dot{L}^2 \subset range(\sigma)$.

Under this approach, we prove here the following results validating assumptions 2.1 and 2.1 of Theorem 2.1 for the Markov semigroup $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$, and the class of distances $\Lambda = \{\rho_{\varepsilon,s} : \varepsilon > 0, s \in (0,1]\}$ defined in (3.26) above.

PROPOSITION 3.12 Fix $\delta_0 > 0$ and suppose there exists $K \in \mathbb{N}$ and $\sigma \in \dot{\boldsymbol{L}}^2$ such that (3.29) holds. Then, for every $M > 0$, $\varepsilon > 0$ and $s \in (0,1]$, there exist a time $T_1 = T_1(M, \varepsilon, s) > 0$ and a coefficient $\kappa_1 = \kappa_1(M) \in (0,1)$, which is independent of ε, s, such that

$$\sup_{N\in\mathbb{N},\, 0<\delta\leq\delta_0}\ \sup_{n\geq T_1/\delta} \mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot), P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) \leq 1 - \kappa_1 \tag{3.57}$$

for all $\delta_0 > 0$, and for every $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$ with $|\xi_0| \leq M$ and $|\widetilde{\xi}_0| \leq M$.

PROPOSITION 3.13 Fix $\delta_0 > 0$ and suppose there exists $K \in \mathbb{N}$ and $\sigma \in \dot{\boldsymbol{L}}^2$ such that (3.29) holds. Then, for every $\kappa_2 \in (0,1)$ and for every $r > 0$ there exists $s \in (0,1]$ for which the following holds:

(i) For every $\varepsilon > 0$ and $\delta_0 > 0$, there exists a constant $C = C(\varepsilon, s, \delta_0) > 0$ such that

$$\sup_{N\in\mathbb{N},\,0<\delta\le\delta_0,\,n\in\mathbb{Z}^+} \mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot), P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) \le C\exp\Big(r|\xi_0|^2 + r|\widetilde{\xi}_0|^2\Big)\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0) \tag{3.58}$$

for every $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$ with $\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0) < 1$.

(ii) For every $\delta_0 > 0$, there exist a parameter $\varepsilon = \varepsilon(\kappa_2, r) > 0$ and a time $T_2 = T_2(\kappa_2, r) > 0$ such that

$$\sup_{N\in\mathbb{N},\,0<\delta\le\delta_0}\ \sup_{n\ge T_2/\delta} \mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot), P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) \le \kappa_2\exp\Big(r|\xi_0|^2 + r|\widetilde{\xi}_0|^2\Big)\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0) \tag{3.59}$$

for every $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$ with $\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0) < 1$.

REMARK 3.14 We notice that item (*i*) of Proposition 3.13 gives a slightly stronger result than required in the general assumption (A3.ii) of Theorem 2.1. Indeed, inequality (3.58) is valid over all $n \in \mathbb{Z}^+$ and $\varepsilon > 0$. In contrast, (2.6) concerns only a finite time interval $[0,\tau]$ and a particular choice of distance-like function $\rho_{\varepsilon,s}$, which thus entails both a particular choice of $s \in (0,1]$ and $\varepsilon > 0$.

Before proceeding with the proofs of Proposition 3.12 and Proposition 3.13, we establish some preliminary facts and terminology that are necessary for following the outline described under (3.54) above. We start with the following result establishing suitable exponential moment bounds for solutions of (3.22).

LEMMA 3.15 Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \le \delta_0$, $\sigma \in \dot{\boldsymbol{L}}^2$ and $\xi_0 \in \dot{L}^2$. Let $\{\xi_{N,\delta}^n\}_{n\in\mathbb{Z}^+}$ be the solution of (3.22) corresponding to the parameters N, δ, and satisfying $\xi_{N,\delta}^0 = \Pi_N\xi_0$ almost surely. Then, there exists an absolute constant $c > 0$ such that for all $\alpha \in \mathbb{R}$ satisfying

$$0 < \alpha \le \frac{c}{|\sigma|^2}\min\left\{\nu, \frac{1}{\delta_0}\right\}, \tag{3.60}$$

the following inequality holds

$$\mathbb{E}\sup_{n\ge1}\exp\left(\alpha\left|\xi_{N,\delta}^n\right|^2 + \alpha\nu\delta\sum_{j=1}^n\left|\nabla\xi_{N,\delta}^j\right|^2 + \frac{n}{4}\ln(1-4\alpha\delta|\sigma|^2)\right) \le \tilde{c}\exp\Big(C\alpha|\xi_0|^2\Big), \tag{3.61}$$

and, consequently,

$$\mathbb{E}\exp\left(\alpha\left|\xi_{N,\delta}^n\right|^2 + \alpha\nu\delta\sum_{j=1}^n\left|\nabla\xi_{N,\delta}^j\right|^2\right) \le \tilde{c}\exp\Big(C\alpha|\xi_0|^2\Big)\exp\Big(\tilde{c}\alpha|\sigma|^2 n\delta\Big) \quad \text{for all } n \in \mathbb{N}. \tag{3.62}$$

Here, $C = \tilde{c}(1 + \nu\delta_0)$ and $\tilde{c} > 0$ is an absolute constant.

Proof. Proceeding as in (3.35)–(3.36) above, and summing (3.36) over $j = 1, \dots, n$, we obtain

$$|\xi^n|^2 - |\xi_0|^2 + 2\nu\delta \sum_{j=1}^{n} |\nabla \xi^j|^2 \leq \delta \sum_{j=1}^{n} |\sigma \eta_j|^2 + M_n, \tag{3.63}$$

where $\{M_n\}_{n\in\mathbb{N}}$ is the martingale defined as

$$M_n := 2\delta^{1/2} \sum_{j=1}^{n} \left(\sigma \eta_j, \xi^{j-1}\right) \tag{3.64}$$

with corresponding quadratic variation given by

$$\langle M \rangle_n = 4\delta \sum_{j=1}^{n} \sum_{k=1}^{d} \left(\sigma_k, \xi^{j-1}\right)^2. \tag{3.65}$$

We estimate $\langle M \rangle_n$ as

$$\begin{aligned} \langle M \rangle_n \leq 4\delta|\sigma|^2 \sum_{j=1}^{n} |\xi^{j-1}|^2 &= 4\delta|\sigma|^2 \left(|\xi_0|^2 + \sum_{j=1}^{n-1} |\xi^j|^2\right) \\ &\leq 4\delta|\sigma|^2|\xi_0|^2 + 4\delta|\sigma|^2 \sum_{j=1}^{n-1} |\nabla \xi^j|^2. \end{aligned} \tag{3.66}$$

Thus, under assumption (3.60) on α with a suitable absolute constant c it follows that

$$\alpha \langle M \rangle_n \leq \nu\delta|\xi_0|^2 + \nu\delta \sum_{j=1}^{n-1} |\nabla \xi^j|^2.$$

Adding and subtracting $\alpha\langle M \rangle_n$ in (3.63) yields

$$|\xi^n|^2 + \nu\delta \sum_{j=1}^{n} |\nabla \xi^j|^2 \leq (1 + \nu\delta)|\xi_0|^2 + \delta \sum_{j=1}^{n} |\sigma \eta_j|^2 + M_n - \alpha\langle M \rangle_n. \tag{3.67}$$

We now subtract $R_n := -\frac{n}{2\alpha}\mathrm{Tr}(\ln(1 - 2\alpha\delta Q_0))$ from both sides of (3.67), where Q_0 is defined in (3.3). Then, multiplying by $\alpha/2$, taking exponentials, the supremum over $n \in \{1, \dots, m\}$ for some $m \in \mathbb{N}$ and expected values, it follows that

$$\mathbb{E} \sup_{1\le n\le m} \exp\left(\frac{\alpha}{2}|\xi^n|^2 + \frac{\alpha\nu\delta}{2}\sum_{j=1}^{n}|\nabla\xi^j|^2 - \frac{\alpha}{2}R_n\right)$$
$$\le \exp\left(\frac{\alpha}{2}(1+\nu\delta)|\xi_0|^2\right)\left[\mathbb{E}\sup_{1\le n\le m}\exp\left(\frac{D_n}{2}\right)\right]^{1/2}\left[\mathbb{E}\sup_{1\le n\le m}\exp\left(\frac{E_n}{2}\right)\right]^{1/2}, \quad (3.68)$$

where

$$D_n = 2\alpha M_n - 2\alpha^2\langle M\rangle_n \quad (3.69)$$

and

$$E_n = \alpha\delta\sum_{j=1}^{n}|\sigma\eta_j|^2 - \alpha R_n. \quad (3.70)$$

Similarly as in (3.43), we have that $\{\exp(D_n)\}_{n\in\mathbb{N}}$ is a martingale with respect to the filtration $\{\mathscr{F}_{t_n}\}_{n\in\mathbb{N}}$, and $\mathbb{E}\exp(D_n) = 1$ for all n.

Clearly, each $\exp(E_n)$ is measurable with respect to $\mathscr{F}_{t_n}$. To conclude that $\{\exp(E_n)\}_{n\in\mathbb{N}}$ is a martingale, it remains to show that $\mathbb{E}|\exp(E_n)| = \mathbb{E}\exp(E_n) < \infty$ and $\mathbb{E}(\exp(E_{n+1})|\mathscr{F}_{t_n}) = \exp(E_n)$ for all $n\in\mathbb{N}$. Since $\alpha < c(\delta_0|\sigma|^2)^{-1}$ and $\sigma\eta_n \sim \mathscr{N}(0, Q_0)$, it follows by invoking once again (Da Prato & Zabczyk, 2014, Proposition 2.17) that for all $\delta \le \delta_0$

$$\mathbb{E}\exp(\alpha\delta|\sigma\eta_n|^2) = \exp\left(-\frac{1}{2}\mathrm{Tr}\ln(1-2\alpha\delta Q_0)\right) \qquad \text{for all } n\in\mathbb{N}.$$

Hence,

$$\mathbb{E}(\exp(E_{n+1})\,|\,\mathscr{F}_{t_n}) = \exp\left(\alpha\delta\sum_{j=1}^{n}|\sigma\eta_j|^2\right)\mathbb{E}\exp\left(\alpha\delta|\sigma\eta_{n+1}|^2 + \frac{(n+1)}{2}\mathrm{Tr}\ln(1-2\alpha\delta Q_0)\right)$$
$$= \exp\left(\alpha\delta\sum_{j=1}^{n}|\sigma\eta_j|^2 + \frac{n}{2}\mathrm{Tr}\ln(1-2\alpha\delta Q_0)\right) = \exp(E_n).$$

This implies that, for all $n\in\mathbb{N}$,

$$\mathbb{E}[\exp(E_n)] = \mathbb{E}[\mathbb{E}[\exp(E_n)\,|\,\mathscr{F}_{t_1}]] = \mathbb{E}[\exp(E_1)] = \exp\left(\alpha\delta|\sigma\eta_1|^2 + \frac{1}{2}\mathrm{Tr}\ln(1-2\alpha\delta Q_0)\right) = 1. \quad (3.71)$$

Therefore, $\{\exp(E_n)\}_{n\in\mathbb{N}}$ is a martingale and, moreover, $\mathbb{E}\exp(E_n) = 1$ for all n.

With these facts, we proceed to further estimate the right-hand side of (3.68) by noticing that

$$\mathbb{E} \sup_{1\leq n\leq m} \exp\left(\frac{E_n}{2}\right) = \int_0^\infty \mathbb{P}\left(\sup_{1\leq n\leq m} \exp(E_n) \geq z^2\right) \mathrm{d}z$$
$$\leq 1 + \int_1^\infty \mathbb{P}\left(\sup_{1\leq n\leq m} \exp(E_n) \geq z^2\right) \mathrm{d}z \leq 1 + \int_1^\infty \frac{\mathbb{E}\exp(E_m)}{z^2}\,\mathrm{d}z = 2, \quad (3.72)$$

where the last inequality follows from Doob's martingale inequality, while in the final equality we used (3.71). Analogously, we can show that

$$\mathbb{E} \sup_{1\leq n\leq m} \exp\left(\frac{D_n}{2}\right) \leq 2. \quad (3.73)$$

Plugging estimates (3.72) and (3.73) into (3.68), it follows that

$$\mathbb{E} \sup_{1\leq n\leq m} \exp\left(\frac{\alpha}{2}|\xi^n|^2 + \frac{\alpha\nu\delta}{2}\sum_{j=1}^n |\nabla\xi^j|^2 - \frac{\alpha}{2}R_n\right) \leq 2\exp\left(\frac{\alpha}{2}(1+\nu\delta)|\xi_0|^2\right), \quad (3.74)$$

for all $m \in \mathbb{N}$. Replacing $\alpha/2$ by α, and noticing that $\mathrm{Tr}(\ln(1-4\alpha\delta Q_0)) \geq \ln(1-4\alpha\delta\mathrm{Tr}(Q_0))$ (see Da Prato & Zabczyk, 2014, Proposition 2.17) and $\mathrm{Tr}(Q_0) = |\sigma|^2$, we obtain that for all $N \in \mathbb{N}$ and $\delta \leq \delta_0$

$$\mathbb{E} \sup_{1\leq n\leq m} \exp\left(\alpha|\xi^n|^2 + \alpha\nu\delta\sum_{j=1}^n |\nabla\xi^j|^2 + \frac{n}{4}\ln(1-4\alpha\delta|\sigma|^2)\right) \leq 2\exp\left(\frac{\alpha}{2}(1+\nu\delta_0)|\xi_0|^2\right) \quad (3.75)$$

for all $m \in \mathbb{N}$. Now we conclude (3.61) from (3.75) by invoking the Monotone Convergence theorem.

For the final inequality (3.62), first notice that (3.61) clearly implies

$$\mathbb{E}\exp\left(\alpha|\xi^n|^2 + \alpha\nu\delta\sum_{j=1}^n |\nabla\xi^j|^2\right) \leq \frac{\tilde{c}\exp\left(C\alpha|\xi_0|^2\right)}{(1-4\alpha\delta|\sigma|^2)^{n/4}} \quad \text{for all } n \in \mathbb{N}. \quad (3.76)$$

Now we use the elementary fact that $\ln(1-x) \geq -ex$ for every $0 \leq x \leq 1/e$. Thus, choosing the constant c in (3.60) appropriately so that $4\alpha\delta_0|\sigma|^2 \leq 1/e$, we obtain that for all $\delta \leq \delta_0$

$$(1-4\alpha\delta|\sigma|^2)^{-n/4} = \exp\left(-\frac{n}{4}\ln(1-4\alpha\delta|\sigma|^2)\right) \leq \exp(\tilde{c}\alpha|\sigma|^2 n\delta).$$

Plugging this inequality into (3.76), we deduce (3.62). This concludes the proof. □

Next, we have the following contraction result.

LEMMA 3.16 Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \leq \delta_0$, $\sigma \in \dot{\boldsymbol{L}}^2$ and $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$. Let $\widetilde{\xi}^n_{N,\delta} = \widetilde{\xi}^n_{N,\delta}(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0)$, $n \in \mathbb{Z}^+$, be the solution of (3.50)–(3.51) corresponding to the parameters N, δ, ξ_0, and satisfying $\widetilde{\xi}^0_{N,\delta} = \Pi_N \widetilde{\xi}_0$ almost surely. Suppose $K \in \mathbb{N}$ and $\beta > 0$ from (3.50) satisfy

$$\nu \lambda_{K+1} \geq 2\beta \tag{3.77}$$

and

$$\beta \geq c \max\left\{ \frac{1}{\delta_0}, \frac{\delta_0^2 |\sigma|^4}{\nu^3}, \frac{|\sigma|^4}{\nu^5} \right\} \tag{3.78}$$

for some absolute constant $c > 0$. Then, for every $n \in \mathbb{Z}^+$

$$\mathbb{E}\left[\left| \widetilde{\xi}^n_{N,\delta}\left(\Pi_N \widetilde{\xi}_0; \Pi_N \xi_0\right) - \xi^n_{N,\delta}(\Pi_N \xi_0) \right|^2 \right] \leq \tilde{c} \frac{\exp\left(C(\nu^3\beta)^{-1/2}|\xi_0|^2\right)}{(1+\beta\delta)^{3n/4}} |\widetilde{\xi}_0 - \xi_0|^2, \tag{3.79}$$

where $\tilde{c}$ is an absolute constant, and $C > 0$ is a constant depending only on ν, δ_0.

Proof. Denote $\zeta^n = \zeta^n_{N,\delta} := \widetilde{\xi}^n_{N,\delta}(\Pi_N \widetilde{\xi}_0; \Pi_N \xi_0) - \xi^n_{N,\delta}(\Pi_N \xi_0)$ and $\mathbf{v}^n = \mathbf{v}^n_{N,\delta} = \widetilde{\mathbf{u}}^n_{N,\delta} - \mathbf{u}^n_{N,\delta}$. Subtracting (3.20) from (3.50), we obtain

$$\zeta^n = \zeta^{n-1} + \delta\Big[\nu \Delta \zeta^n - \Pi_N(\mathbf{v}^{n-1} \cdot \nabla \zeta^n) - \Pi_N(\mathbf{v}^{n-1} \cdot \nabla \xi^n) - \Pi_N(\mathbf{u}^{n-1} \cdot \nabla \zeta^n) - \beta \Pi_K \zeta^n\Big]. \tag{3.80}$$

Taking the inner product of (3.80) with ζ^n and invoking (3.7) and (3.34), it follows that

$$|\zeta^n|^2 - |\zeta^{n-1}|^2 + |\zeta^n - \zeta^{n-1}|^2 + 2\nu\delta|\nabla \zeta^n|^2 = -2\delta(\mathbf{v}^{n-1} \cdot \nabla \xi^n, \zeta^n) - 2\beta\delta\left|\Pi_K \zeta^n\right|^2. \tag{3.81}$$

Invoking (3.8) with $a = 1/2$ and Young's inequality, we estimate the nonlinear term above as

$$\begin{aligned} 2\delta|(\mathbf{v}^{n-1} \cdot \nabla \xi^n, \zeta^n)| &\leq \tilde{c}\delta|\zeta^{n-1}||\nabla \xi^n||\zeta^n|^{1/2}|\nabla \zeta^n|^{1/2} \\ &\leq \tilde{c}\frac{\delta}{(\nu\beta)^{1/2}}|\zeta^{n-1}|^2|\nabla \xi^n|^2 + \beta\delta|\zeta^n|^2 + \nu\delta|\nabla \zeta^n|^2, \end{aligned} \tag{3.82}$$

for some absolute constant $\tilde{c} > 0$. Thus, from (3.81),

$$|\zeta^n|^2 - |\zeta^{n-1}|^2 + |\zeta^n - \zeta^{n-1}|^2 + \nu\delta|\nabla \zeta^n|^2 \leq \tilde{c}\frac{\delta}{(\nu\beta)^{1/2}}|\zeta^{n-1}|^2|\nabla \xi^n|^2 + \beta\delta|\zeta^n|^2 - 2\beta\delta\left|\Pi_K \zeta^n\right|^2.$$

With inequality (3.11), we estimate the last term in the left-hand side as

$$\begin{aligned} \nu\delta|\nabla \zeta^n|^2 = \nu\delta\left(\left|\nabla \Pi_K \zeta^n\right|^2 + |\nabla(I - \Pi_K)\zeta^n|^2\right) &\geq \nu\delta\left(|\nabla \Pi_K \zeta^n|^2 + \lambda_{K+1}|(I - \Pi_K)\zeta^n|^2\right) \\ &\geq \nu\delta\left|\nabla \Pi_K \zeta^n\right|^2 + 2\beta\delta\left|(I - \Pi_K)\zeta^n\right|^2, \end{aligned} \tag{3.83}$$

where in the last inequality we invoked the hypotheses that $\nu\lambda_{K+1} \geq 2\beta$, (3.77). After rearranging terms, we deduce that

$$|\zeta^n|^2 - |\zeta^{n-1}|^2 + |\zeta^n - \zeta^{n-1}|^2 + \nu\delta|\nabla \Pi_K \zeta^n|^2 + \beta\delta|\zeta^n|^2 \leq \tilde{c}\frac{\delta}{(\nu\beta)^{1/2}}|\zeta^{n-1}|^2|\nabla\xi^n|^2. \tag{3.84}$$

In particular, after ignoring the third and fourth terms from the left-hand side of (3.84), we obtain

$$(1+\beta\delta)|\zeta^n|^2 \leq \left(1 + \tilde{c}\delta\frac{|\nabla\xi^n|^2}{(\nu\beta)^{1/2}}\right)|\zeta^{n-1}|^2 \quad \text{for all } n \in \mathbb{N}.$$

Therefore, by induction,

$$|\zeta^n|^2 \leq \frac{|\zeta_0|^2}{(1+\beta\delta)^n}\prod_{j=1}^n\left(1 + \tilde{c}\delta\frac{|\nabla\xi^j|^2}{(\nu\beta)^{1/2}}\right) \leq \frac{|\zeta_0|^2}{(1+\beta\delta)^n}\exp\left(\sum_{j=1}^n \tilde{c}\delta\frac{|\nabla\xi^j|^2}{(\nu\beta)^{1/2}}\right), \tag{3.85}$$

where in the last inequality we used that $1 + x \leq e^x$, for all $x \in \mathbb{R}$. Taking expected values on both sides of (3.85), we thus obtain

$$\mathbb{E}[|\zeta^n|^2] \leq \frac{|\zeta_0|^2}{(1+\beta\delta)^n}\mathbb{E}\exp\left(\sum_{j=1}^n \tilde{c}\delta\frac{|\nabla\xi^j|^2}{(\nu\beta)^{1/2}}\right). \tag{3.86}$$

Now let $\alpha = \tilde{c}(\nu^3\beta)^{-1/2}$. From assumption (3.78) on β it is clear that α satisfies condition (3.60) from Proposition 3.15. Thus, from (3.76) and (3.86), we obtain that

$$\mathbb{E}[|\zeta^n|^2] \leq \frac{|\zeta_0|^2}{(1+\beta\delta)^n}\frac{\tilde{c}\exp\left(C\alpha|\xi_0|^2\right)}{(1-4\alpha\delta|\sigma|^2)^{n/4}}. \tag{3.87}$$

Moreover, we can assume that the constant c in assumption (3.78) on β is large enough so that $\alpha = \tilde{c}(\nu^3\beta)^{-1/2} \leq 1/(8\delta_0|\sigma|^2)$ and $\beta \geq 1/\delta_0$. We then estimate

$$\begin{aligned} 1 - 4\alpha\delta|\sigma|^2 &\geq 1 - \frac{\delta}{2\delta_0} = 1 - \frac{1}{2\delta_0\beta}\beta\delta \\ &\geq 1 - \frac{1}{1+\delta_0\beta}\beta\delta \geq 1 - \frac{1}{1+\beta\delta}\beta\delta = \frac{1}{1+\beta\delta}, \end{aligned}$$

where in the last inequality we used that $\delta \leq \delta_0$. Therefore,

$$(1+\beta\delta)^n(1-4\alpha\delta|\sigma|^2)^{n/4} \geq (1+\beta\delta)^n(1+\beta\delta)^{-n/4} = (1+\beta\delta)^{3n/4},$$

so that from (3.87) we deduce

$$\mathbb{E}[|\zeta^n|^2] \leq |\zeta_0|^2 \frac{\tilde{c}\exp\left(C\alpha|\xi_0|^2\right)}{(1+\beta\delta)^{3n/4}}. \tag{3.88}$$

This shows (3.79) and concludes the proof. □

We next recall some additional notions of distance in the space of probability measures on any measurable space (X, Σ_X), along with some useful related inequalities. These will be particularly helpful in further estimating the second term in the right-hand side of (3.54), i.e. the cost-of-control term. First, we recall that the *total variation* distance between any two measures $\mu, \tilde{\mu} \in \Pr(X)$ is defined as

$$\|\mu - \tilde{\mu}\|_{\mathrm{TV}} := \sup_{A\in\Sigma_X} |\mu(A) - \tilde{\mu}(A)|. \tag{3.89}$$

Given another measurable space (Y, Σ_Y) and a measurable function $\phi : X \to Y$, it follows immediately from definition (3.89) that

$$\left\|\phi^*\mu - \phi^*\tilde{\mu}\right\|_{\mathrm{TV}} \leq \|\mu - \tilde{\mu}\|_{\mathrm{TV}}, \tag{3.90}$$

where here $\phi^*\mu \in \Pr(Y)$ denotes the *pushforward measure* of μ by the function ϕ, i.e. $\phi^*\mu(A) := \mu(\phi^{-1}(A))$ for all $A \in \Sigma_Y$.

Secondly, we recall that the *Kullback–Leibler divergence* is defined as

$$D_{\mathrm{KL}}(\tilde{\mu}\|\mu) := \int_X \ln\left(\frac{\mathrm{d}\tilde{\mu}}{\mathrm{d}\mu}(\xi)\right)\tilde{\mu}(\mathrm{d}\xi), \tag{3.91}$$

for any $\mu, \tilde{\mu} \in \Pr(X)$ such that $\tilde{\mu}$ is absolutely continuous with respect to μ, so that the Radon–Nikodym derivative $d\tilde{\mu}/d\mu$ is well-defined. When $\tilde{\mu}$ is not absolutely continuous with respect to μ, we set $D_{\mathrm{KL}}(\tilde{\mu}\|\mu) := +\infty$.

Regarding the definitions in (3.89) and (3.91), we will make use of two useful inequalities from Butkovsky *et al.* (2020) providing estimates on the distance between the law of a d-dimensional Wiener process W and the corresponding shifted process

$$\widehat{W}(t) = W(t) + \int_0^t \varphi(\tau)\,\mathrm{d}\tau \tag{3.92}$$

for some progressively measurable process $\varphi(t)$, $t \geq 0$. In the proofs below, these inequalities will be applied with $\varphi(t) = \sum_{j=1}^{\infty} \psi_j \mathbb{1}_{[t_{j-1},t_j)}(t)$, for ψ_j as given in (3.56). Specifically, denoting by $\mathscr{L}(W)$ and $\mathscr{L}(\widehat{W})$ the laws of W and $\widehat{W}$, respectively, it follows from Butkovsky *et al.* (2020, Theorem A.2) that

$$D_{\mathrm{KL}}(\mathscr{L}(\widehat{W})\|\mathscr{L}(W)) \leq \frac{1}{2}\mathbb{E}\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t. \tag{3.93}$$

And from Butkovsky *et al.* (2020, Theorem A.5, (A.13)), we have that for any $a \in (0,1]$[2]

$$\left\| \mathscr{L}(\widehat{W}) - \mathscr{L}(W) \right\|_{\mathrm{TV}} \leq 2^{\frac{1-a}{1+a}} \left\{ \mathbb{E}\left[\left(\int_0^\infty |\varphi(t)|^2 \, \mathrm{d}t \right)^a \right] \right\}^{\frac{1}{1+a}}. \tag{3.94}$$

We also recall the following inequality providing an explicit relation between these two definitions (see e.g. Tsybakov, 2009, inequality (2.25)):

$$\|\mu - \tilde{\mu}\|_{\mathrm{TV}} \leq 1 - \frac{1}{2} \exp\left(-D_{\mathrm{KL}}(\tilde{\mu} \| \mu)\right) \tag{3.95}$$

for all $\mu, \tilde{\mu} \in \Pr(X)$.

To further connect these definitions with the Wasserstein-like distances defined in (2.2) on $\Pr(\dot{L}^2)$, we notice that for any distance-like function $\rho : \dot{L}^2 \times \dot{L}^2 \to \mathbb{R}^+$ such that $\rho(\xi, \widetilde{\xi}) \leq 1$ for all $\xi, \widetilde{\xi} \in \dot{L}^2$, it follows as an immediate consequence of the coupling lemma (Kuksin & Shirikyan, 2012, Lemma 1.2.24) that

$$\mathscr{W}_\rho(\mu, \tilde{\mu}) \leq \|\mu - \tilde{\mu}\|_{\mathrm{TV}} \qquad \text{for all } \mu, \tilde{\mu} \in \Pr(\dot{L}^2). \tag{3.96}$$

With these notations and facts in place, we now proceed with the proofs of Proposition 3.12 and Proposition 3.13.

Proof of Proposition 3.12. Fix $M > 0$, $\varepsilon > 0$, $s \in (0,1]$, and let $\xi_0, \widetilde{\xi}_0 \in \dot{L}^2$ such that $|\xi_0| \leq M$ and $|\widetilde{\xi}_0| \leq M$.

We start with the triangle inequality as in (3.54) and provide an estimate of each term in the right-hand side by following the strategy described in the introduction to this section. For the first term, it follows from the definition of $\mathscr{W}_{\varepsilon,s}$ according to (2.2) and (3.26), along with Hölder's inequality, that

$$\begin{aligned}
\mathscr{W}_{\varepsilon,s}\left(P_n^{N,\delta}(\xi_0, \cdot), \widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0, \cdot)\right) &\leq \mathbb{E}\left(1 \wedge \frac{\left|\xi^n(\varPi_N \xi_0) - \widetilde{\xi}^n(\varPi_N \xi_0; \varPi_N \widetilde{\xi}_0)\right|^s}{\varepsilon}\right) \\
&\leq \frac{1}{\varepsilon}\left(\mathbb{E}\left[\left|\xi^n(\varPi_N \xi_0) - \widetilde{\xi}^n(\varPi_N \xi_0; \varPi_N \widetilde{\xi}_0)\right|^2\right]\right)^{s/2}.
\end{aligned} \tag{3.97}$$

By assumption, there exists $K \in \mathbb{N}$ such that (3.29) holds. In particular, the second condition in (3.29) implies that we can take $\beta > 0$ satisfying assumptions (3.77) and (3.78) of Lemma 3.16. It thus follows

[2] In Butkovsky *et al.* (2020, Theorem A.5, (A.13)), it is actually assumed $a \in (0,1)$. In fact, inequality (3.94) also holds with $a = 1$, although a slightly sharper bound is valid in this case due to (3.93) and Pinsker's inequality (see e.g. Tsybakov, 2009, Lemma 2.5.(i)). Namely, $\left\| \mathscr{L}(\widehat{W}) - \mathscr{L}(W) \right\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}(\mathscr{L}(\widehat{W}) \| \mathscr{L}(W))} \leq \frac{1}{2}\left(\mathbb{E} \int_0^\infty |\varphi(t)|^2 \, \mathrm{d}t\right)^{1/2}$.

from (3.79) and (3.97) that

$$\mathscr{W}_{\varepsilon,s}\left(P_n^{N,\delta}(\xi_0,\cdot),\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot)\right) \leq \frac{|\xi_0-\widetilde{\xi}_0|^s}{\varepsilon}\frac{\tilde{c}^s \exp\left(Cs(\nu^3\beta)^{-1/2}|\xi_0|^2\right)}{(1+\beta\delta)^{3ns/8}} \tag{3.98}$$

$$= \tilde{c}^s\frac{M^s}{\varepsilon}\frac{\exp\left(Cs(\nu^3\beta)^{-1/2}M^2\right)}{(1+\beta\delta)^{3ns/8}}, \tag{3.99}$$

for some absolute constant $\tilde{c} > 0$ and some constant $C > 0$ depending only on ν, δ_0.

We proceed to estimate the second term in the right-hand side of (3.54). Let us denote by $\xi^n(\Pi_N\widetilde{\xi}_0; W)$ and $\widetilde{\xi}^n(\Pi_N\widetilde{\xi}_0; \Pi_N\xi_0, W)$ the solutions of (3.20) and (3.50)–(3.51), respectively, starting from $\Pi_N\widetilde{\xi}_0 \in \Pi_N\dot{L}^2$ and corresponding to the family $W = \{W^k\}_{k=1}^d$ of independent real-valued Brownian motions W^k, $k = 1,\dots,d$. Then, denoting by $\mathscr{L}(Z)$ the law of a random variable Z, we can equivalently write the Markov transition kernels defined in (3.24) and (3.52) as

$$P_n^{N,\delta}(\widetilde{\xi}_0,\cdot) = \mathscr{L}\left(\xi^n(\Pi_N\widetilde{\xi}_0; W)\right) \quad \text{and} \quad \widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot) = \mathscr{L}\left(\widetilde{\xi}^n(\Pi_N\widetilde{\xi}_0;\Pi_N\xi_0, W)\right),$$

respectively. From inequality (3.96), we thus have

$$\begin{aligned}\mathscr{W}_{\varepsilon,s}\left(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\right) &\leq \left\|\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot) - P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\right\|_{\mathrm{TV}} \\ &= \left\|\mathscr{L}(\widetilde{\xi}^n(\Pi_N\widetilde{\xi}_0;\Pi_N\xi_0,W)) - \mathscr{L}(\xi^n(\Pi_N\widetilde{\xi}_0;W))\right\|_{\mathrm{TV}}.\end{aligned} \tag{3.100}$$

Let $\widehat{W} = \{\widehat{W}^k\}_{k=1}^d$ be the family of shifted independent Brownian motions defined in (3.55)–(3.56). Here notice that, in the definition of ψ_j in (3.56), $\sigma^{-1}\Pi_K$ is well-defined due to the assumption that $\Pi_K\dot{L}^2 \subset \mathit{range}(\sigma)$ in (3.29). Moreover, due to the uniqueness of pathwise strong solutions of (3.20), as shown in Proposition 3.8, it follows that $\xi^n(\Pi_N\widetilde{\xi}_0;\widehat{W}) = \widetilde{\xi}^n(\Pi_N\widetilde{\xi}_0;\Pi_N\xi_0,W)$ for all $n \in \mathbb{N}$ almost surely. Hence, from (3.100),

$$\mathscr{W}_{\varepsilon,s}(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)) \leq \left\|\mathscr{L}(\xi^n(\Pi_N\widetilde{\xi}_0;\widehat{W})) - \mathscr{L}(\xi^n(\Pi_N\widetilde{\xi}_0;W))\right\|_{\mathrm{TV}}. \tag{3.101}$$

It is not difficult to show that $W \in \mathscr{C}([0,n\delta];\mathbb{R}^d) \mapsto \xi^n(\Pi_N\widetilde{\xi}_0;W) \in \dot{L}^2$ is a continuous mapping with respect to the topology in $\dot{L}^2$ generated by $|\cdot|$, and the one in the space $\mathscr{C}([0,n\delta];\mathbb{R}^d)$ of continuous $\mathbb{R}^d$-valued functions on $[0,n\delta]$ derived from the supremum norm on this interval. It thus follows from (3.90) that

$$\begin{aligned}\left\|\mathscr{L}\left(\xi^n(\Pi_N\widetilde{\xi}_0;\widehat{W})\right) - \mathscr{L}\left(\xi^n(\Pi_N\widetilde{\xi}_0;W)\right)\right\|_{\mathrm{TV}} &= \left\|\xi^n(\Pi_N\widetilde{\xi}_0;\cdot)^*\mathscr{L}(\widehat{W}) - \xi^n(\Pi_N\widetilde{\xi}_0;\cdot)^*\mathscr{L}(W)\right\|_{\mathrm{TV}} \\ &\leq \left\|\mathscr{L}(\widehat{W}) - \mathscr{L}(W)\right\|_{\mathrm{TV}}.\end{aligned} \tag{3.102}$$

Together with inequality (3.95), we thus have

$$\left\| \mathscr{L}\Big(\xi^n(\Pi_N\widetilde{\xi}_0;\widehat{W})\Big) - \mathscr{L}\Big(\xi^n(\Pi_N\widetilde{\xi}_0;W)\Big)\right\|_{\mathrm{TV}} \leq 1 - \frac{1}{2}\exp\left(-D_{\mathrm{KL}}(\mathscr{L}(\widehat{W})\|\mathscr{L}(W))\right). \tag{3.103}$$

Recalling from (3.55) that

$$\widehat{W}(t) = W(t) + \int_0^t \varphi(\tau)\mathrm{d}\tau, \quad \text{ where } \varphi(\tau) := \sum_{j=1}^\infty \psi_j \mathbb{1}_{[t_{j-1},t_j)}(\tau),$$

with ψ_j as defined in (3.56), we now invoke (3.93) and obtain that

$$D_{\mathrm{KL}}(\mathscr{L}(\widehat{W})\|\mathscr{L}(W)) \leq \frac{1}{2}\mathbb{E}\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t = \mathbb{E}\delta\sum_{j=1}^\infty |\psi_j|^2. \tag{3.104}$$

Now invoking the fact that the pseudo-inverse σ^{-1} is bounded, and once again Lemma 3.16, we further estimate the right-hand side above as

$$\begin{aligned}
\mathbb{E}\delta\sum_{j=1}^\infty |\psi_j|^2 &\leq \delta\beta^2\|\sigma^{-1}\|^2\sum_{j=1}^\infty \mathbb{E}\left|\widetilde{\xi}^j\Big(\Pi_N\widetilde{\xi}_0;\Pi_N\xi_0,W\Big) - \xi^j(\Pi_N\xi_0;W)\right|^2 \\
&\leq \delta\beta^2\|\sigma^{-1}\|^2\sum_{j=1}^\infty |\xi_0-\widetilde{\xi}_0|^2\frac{\tilde{c}\exp\left(C(\nu^3\beta)^{-1/2}|\xi_0|^2\right)}{(1+\beta\delta)^{3j/4}}
\end{aligned} \tag{3.105}$$

for some constant $C>0$, and where $\|\sigma^{-1}\|$ denotes the operator norm of σ^{-1}. Notice that

$$\begin{aligned}
\beta\delta\sum_{j=1}^\infty \frac{1}{(1+\beta\delta)^{3j/4}} &= \beta\delta\frac{1}{1-(1+\beta\delta)^{-3/4}} \leq \beta\delta\frac{1}{1-(1+\beta\delta)^{-1/2}} = \beta\delta\frac{(1+\beta\delta)^{1/2}}{(1+\beta\delta)^{1/2}-1} \\
&= \beta\delta\frac{(1+\beta\delta)^{1/2}[(1+\beta\delta)^{1/2}+1]}{\beta\delta} \leq 2(1+\beta\delta),
\end{aligned}$$

so that, from (3.105),

$$\frac{1}{2}\mathbb{E}\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t = \mathbb{E}\delta\sum_{j=1}^\infty|\psi_j|^2 \leq \tilde{c}\beta(1+\beta\delta)\|\sigma^{-1}\|^2|\xi_0-\widetilde{\xi}_0|^2\exp\Big(C(\nu^3\beta)^{-1/2}|\xi_0|^2\Big) \tag{3.106}$$

$$\leq \tilde{c}\beta(1+\beta\delta)\|\sigma^{-1}\|^2M^2\exp\Big(C(\nu^3\beta)^{-1/2}M^2\Big). \tag{3.107}$$

From (3.101)–(3.104) and (3.107), we thus have

$$\begin{aligned}&\mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big)\\&\qquad\le 1-\frac{1}{2}\exp\Big\{-\tilde{c}\beta(1+\beta\delta)\|\sigma^{-1}\|^2M^2\exp\Big(C(\nu^3\beta)^{-1/2}M^2\Big)\Big\}.\end{aligned}\tag{3.108}$$

Hence, combining inequality (3.54) with the estimates (3.99) and (3.108), it follows that

$$\begin{aligned}\mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big)&\le \tilde{c}^s\frac{M^s}{\varepsilon}\frac{\exp\big(Cs(\nu^3\beta)^{-1/2}M^2\big)}{(1+\beta\delta)^{3ns/8}}\\&\quad+1-\frac{1}{2}\exp\Big\{-\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2M^2\exp\Big(C(\nu^3\beta)^{-1/2}M^2\Big)\Big\},\end{aligned}\tag{3.109}$$

where we have used that $\delta\le\delta_0$ to further estimate the right-hand side of (3.108).

To arrive at (3.57), we use the fact that for any constant $0<a<1$ we have $\ln(1+x)\ge ax$ for all $x\in[0,(1/a)-1]$. In particular, taking $a=1/(1+\beta\delta_0)$ we have $\beta\delta\le\beta\delta_0=(1/a)-1$, so that

$$\frac{1}{(1+\beta\delta)^{3ns/8}}=\exp\left(-\frac{3ns}{8}\ln(1+\beta\delta)\right)\le\exp\left(-\frac{3s}{8}\frac{\beta}{1+\beta\delta_0}n\delta\right).\tag{3.110}$$

Hence, we can fix a time $T_1>0$ depending on $M,s,\varepsilon,\beta,\sigma,\delta_0$ such that for all $n\in\mathbb{N}$ with $n\delta\ge T_1$ the first term in the right-hand side of (3.109) can be estimated as

$$\begin{aligned}\tilde{c}^s\frac{M^s}{\varepsilon}\frac{\exp\big(Cs(\nu^3\beta)^{-1/2}M^2\big)}{(1+\beta\delta)^{3ns/8}}&\le\tilde{c}^s\frac{M^s}{\varepsilon}\exp\Big(Cs(\nu^3\beta)^{-1/2}M^2\Big)\exp\left(-\frac{3s}{8}\frac{\beta}{1+\beta\delta_0}n\delta\right)\\&\le\frac{1}{4}\exp\Big\{-\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2M^2\exp\Big(C(\nu^3\beta)^{-1/2}M^2\Big)\Big\}.\end{aligned}\tag{3.111}$$

From (3.109) and (3.111), we thus conclude

$$\begin{aligned}\mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big)&\le 1-\frac{1}{4}\exp\Big\{-\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2M^2\exp\Big(C(\nu^3\beta)^{-1/2}M^2\Big)\Big\}\\&=:1-\kappa_1.\end{aligned}$$

This shows (3.57) and concludes the proof. □

Proof of Proposition 3.13. Fix $\kappa_2\in(0,1)$ and $r>0$. Let $\xi_0,\widetilde{\xi}_0\in\dot{L}^2$ satisfying $\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0)<1$, for $\rho_{\varepsilon,s}$ as defined in (3.26). Note that this implies $\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0)=|\xi_0-\widetilde{\xi}_0|^s/\varepsilon$.

We proceed analogously as in the proof of Proposition 3.12, starting with the inequality (3.54). For the first term in the right-hand side of (3.54), we first estimate as in (3.98). Then, choose $s\in(0,1]$ such that

$$\frac{Cs}{(\nu^3\beta)^{1/2}}\le r,\tag{3.112}$$

with $C > 0$ as in (3.98) and (3.106). Here we recall that $\beta > 0$ is fixed so that assumptions (3.77) and (3.78) of Lemma 3.16 hold, which is possible due to the second condition in the standing assumption (3.29).

With this choice of s, it follows from (3.98) that

$$\mathscr{W}_{\varepsilon,s}\Big(P_n^{N,\delta}(\xi_0,\cdot),\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot)\Big) \leq \tilde{c}^s \frac{|\xi_0-\widetilde{\xi}_0|^s}{\varepsilon}\frac{\exp\left(r|\xi_0|^2\right)}{(1+\beta\delta)^{3ns/8}}. \tag{3.113}$$

For the second term in the right-hand side of (3.54), we proceed as in (3.100)–(3.102), and then invoke (3.94) to obtain that for any $a \in (0,1]$

$$\mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) \leq \big\|\mathscr{L}(\widehat{W})-\mathscr{L}(W)\big\|_{\mathrm{TV}} \leq 2^{\frac{1-a}{1+a}}\left\{\mathbb{E}\left[\left(\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t\right)^a\right]\right\}^{\frac{1}{1+a}}. \tag{3.114}$$

By Hölder's inequality, together with estimate (3.106), we have

$$\begin{aligned}
2^{\frac{1-a}{1+a}}&\left\{\mathbb{E}\left[\left(\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t\right)^a\right]\right\}^{\frac{1}{1+a}} \leq 2^{\frac{1-a}{1+a}}\left[\mathbb{E}\int_0^\infty |\varphi(t)|^2\,\mathrm{d}t\right]^{\frac{a}{1+a}}\\
&\leq 2^{\frac{1-a}{1+a}}\left(\tilde{c}\beta(1+\beta\delta)\|\sigma^{-1}\|^2\right)^{\frac{a}{1+a}}|\xi_0-\widetilde{\xi}_0|^{\frac{2a}{1+a}}\exp\left(\frac{Ca}{(1+a)(\nu^3\beta)^{1/2}}|\xi_0|^2\right).
\end{aligned} \tag{3.115}$$

In particular, choosing $a \in (0,1]$ such that $2a/(1+a) = s$, with $s \in (0,1]$ as fixed in (3.112), it follows from (3.114), (115) and (3.112) that

$$\begin{aligned}
\mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_n^{N,\delta,\xi_0}(\widetilde{\xi}_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot)\Big) &\leq 2^{1-s}\left(\tilde{c}\beta(1+\beta\delta)\|\sigma^{-1}\|^2\right)^{s/2}|\xi_0-\widetilde{\xi}_0|^s\exp\left(r|\xi_0|^2\right)\\
&\leq 2^{1-s}\left(\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2\right)^{s/2}|\xi_0-\widetilde{\xi}_0|^s\exp\left(r|\xi_0|^2\right).
\end{aligned} \tag{3.116}$$

Thus, from (3.54), (3.113), (3.116) and since $\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0) = |\xi_0-\widetilde{\xi}_0|^s/\varepsilon$, it follows that

$$\begin{aligned}
&\mathscr{W}_{\varepsilon,s}(P_n^{N,\delta}(\xi_0,\cdot),P_n^{N,\delta}(\widetilde{\xi}_0,\cdot))\\
&\qquad\leq\left[\frac{\tilde{c}^s}{(1+\beta\delta)^{3ns/8}}+\varepsilon 2^{1-s}\left(\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2\right)^{s/2}\right]\exp\left(r|\xi_0|^2\right)\rho_{\varepsilon,s}(\xi_0,\widetilde{\xi}_0)
\end{aligned} \tag{3.117}$$

for every $n \in \mathbb{N}$.

In particular, by estimating $1/(1+\beta\delta)^{3ns/8} \leq 1$, we deduce that (3.58) holds with

$$C(\varepsilon,s) = \tilde{c}^s + \varepsilon 2^{1-s}\left(\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2\right)^{s/2}.$$

Moreover, proceeding as in (3.110) and choosing $T_2 = T_2(\kappa_2, r) > 0$ and $\varepsilon = \varepsilon(\kappa_2, r) > 0$ such that

$$\tilde{c}^s \exp\left(-\frac{3s}{8}\frac{\beta}{1+\beta\delta_0}T_2\right) + \varepsilon 2^{1-s}\left(\tilde{c}\beta\left(1+\beta\delta_0\right)\|\sigma^{-1}\|^2\right)^{s/2} \leq \kappa_2,$$

we conclude from (3.117) that (3.59) holds for every $n \in \mathbb{N}$ with $n\delta \geq T_2$, as desired. □

We conclude this section by combining the above results to deduce a proof of Theorem 3.9.

Proof of Theorem 3.9. Fix $N \in \mathbb{N}$ and $\delta \leq \delta_0$. Following the notation from Theorem 2.1, we take $(X, \|\cdot\|) = (\dot{L}^2, |\cdot|)$, $\mathscr{I} = \delta\mathbb{Z}^+$, $\{P_t\}_{t\in\mathscr{I}}$ given by $P_n^{N,\delta}$, $n \in \mathbb{Z}^+$ and Λ as the class of distance functions defined in (3.26). It follows from Proposition 3.11, Proposition 3.12 and Proposition 3.13 that assumptions (A1), (A2) and (A3) of Theorem 2.1 are satisfied in this setting. Thus, from (2.9) we obtain that for every $m > 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m)$ there exist $\varepsilon > 0$, $s \in (0, 1]$, $T > 0$, and constants $C_1, C_2 > 0$ for which the following holds

$$\mathscr{W}_{\varepsilon,s,\alpha}\left(\mu P_n^{N,\delta}, \tilde{\mu} P_n^{N,\delta}\right) \leq C_1\, e^{-n\delta C_2}\mathscr{W}_{\varepsilon,s,\alpha/m}(\mu, \tilde{\mu}) \tag{3.118}$$

for every $\mu, \tilde{\mu} \in \Pr(\dot{L}^2)$ and all $n \in \mathbb{Z}^+$ such that $n\delta \geq T$.

Now take any $t \in \mathbb{R}^+$ with $t \geq \delta_0 + T$, and let $n_0 := \inf_{n\in\mathbb{Z}^+}\{n\delta_0 \geq T\}$. It follows that $(n_0 - 1)\delta_0 < T \leq t - \delta_0$, and hence $t \geq n_0\delta_0 \geq n_0\delta$ for all $\delta \leq \delta_0$. Thus, there exists $n \in \mathbb{Z}^+$ for which $t \in [n\delta, (n+1)\delta)$, so that from definition (3.28) we have $\mathscr{P}_t^{N,\delta} = P_n^{N,\delta}$. Therefore,

$$\begin{aligned}\mathscr{W}_{\varepsilon,s,\alpha}\left(\mu \mathscr{P}_t^{N,\delta}, \tilde{\mu}\mathscr{P}_t^{N,\delta}\right) = \mathscr{W}_{\varepsilon,s,\alpha}\left(\mu P_n^{N,\delta}, \tilde{\mu} P_n^{N,\delta}\right) &\leq C_1\, e^{-n\delta C_2}\mathscr{W}_{\varepsilon,s,\alpha/m}(\mu, \tilde{\mu}) \\ &\leq C_1 e^{\delta_0 C_2}\, e^{-tC_2}\mathscr{W}_{\varepsilon,s,\alpha/m}(\mu, \tilde{\mu})\end{aligned} \tag{3.119}$$

for every $\mu, \tilde{\mu} \in \Pr(\dot{L}^2)$. Moreover, according to the dependence of the constants C_1, C_2 made explicit in the statement of Theorem 2.1, it follows from Proposition 3.11 that C_1 and C_2 depend *only* on m, α, T, ν, $|\sigma|$, δ_0. Hence, we may take the supremum in (3.119) with respect to $N \in \mathbb{N}$ and δ in $(0, \delta_0]$ to conclude that (3.30) holds for every $t \geq \delta_0 + T$. □

3.3 *Finite-time strong error estimates for the numerical scheme*

In this section, we present an estimate of the error between a solution $\xi(t)$, $t \geq 0$, of (3.1), and a solution $\xi_{N,\delta}^n$, $n \in \mathbb{Z}^+$, of the numerical scheme (3.20), in a suitable strong sense. This will be used later in Section 3.4 to show a uniform weak convergence result for the family of Markov semigroups $\{P_n^{N,\delta}\}_{n\in\mathbb{Z}^+}$, defined in (3.25), as an application of Theorem 2.8. Specifically, it will be used to verify assumption (H3) in Theorem 2.5.

For this purpose, we split the error $|\xi(n\delta) - \xi_{N,\delta}^n|$ into the spatial discretization error $|\xi(n\delta) - \xi_N(n\delta)|$ and the time discretization error $|\xi_N(n\delta) - \xi_{N,\delta}^n|$. Here we recall that $\xi_N(t)$, $t \geq 0$, denotes a solution of the spectral Galerkin discretization scheme (3.12). Concretely, we obtain a strong $L^2(\Omega)$ estimate of the spatial discretization error with respect to the topology in $L^\infty_{\text{loc},t}L^2_x$. For the time discretization error, due to limitations associated to the nonlinear terms in (3.12) and (3.20), we are only able to obtain strong convergence in $L^p(\Omega; L^\infty_{\text{loc},t}L^2_x)$ for sufficiently small $p > 0$. As we show later in Proposition 3.19, this is however compensated in the Wasserstein error estimate by the presence of the Lyapunov function

$\xi \mapsto \exp(\alpha|\xi|^2)$ in the definition of $\rho_{\varepsilon,s,\alpha}$, thanks to the associated Lyapunov inequalities from Proposition 3.11 and Proposition 3.6.

We start by providing an estimate of the spatial discretization error.

PROPOSITION 3.17 Fix any $N \in \mathbb{N}$, $\sigma \in \dot{\boldsymbol{H}}^1$ and $\xi_0 \in \dot{H}^1$. Let $\xi = \xi(t)$ and $\xi_N = \xi_N(t)$ be the solutions of (3.1) and (3.12), satisfying $\xi(0) = \xi_0$ and $\xi_N(0) = \Pi_N \xi_0$ almost surely, respectively. Then, for every α satisfying

$$0 < \alpha \leq \frac{\nu}{2|\sigma|^2},$$

it follows that, for every $T > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} |\xi(t) - \xi_N(t)|^2\right] \leq \frac{C}{N}\left[\exp\left(c\alpha|\xi_0|^2\right) + |\nabla\xi_0|^2\right], \tag{3.120}$$

for some positive constant C depending only on $\nu, |\sigma|, |\nabla\sigma|, \alpha, T$.

Proof. We start by estimating the spatial discretization error by its low mode and high mode components, namely

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]} |\xi(t) - \xi_N(t)|^2\right] &= \mathbb{E}\left[\sup_{t\in[0,T]} |(\Pi_N\xi(t) - \xi_N(t)) + (I - \Pi_N)\xi(t)|^2\right] \\ &\leq 2\mathbb{E}\left[\sup_{t\in[0,T]} |\Pi_N\xi(t) - \xi_N(t)|^2\right] + 2\mathbb{E}\left[\sup_{t\in[0,T]} |(I - \Pi_N)\xi(t)|^2\right]. \end{aligned} \tag{3.121}$$

For the second term, it follows from (3.11) and the analogous version of the bound (3.13) for $\xi(t)$, $t \geq 0$, that

$$\mathbb{E}\left[\sup_{t\in[0,T]} |(I - \Pi_N)\xi(t)|^2\right] \leq \lambda_{N+1}^{-1}\mathbb{E}\left[\sup_{t\in[0,T]} |\nabla\xi(t)|^2\right] \leq N^{-1}C\Big(1 + |\xi_0|^4 + |\nabla\xi_0|^2\Big), \tag{3.122}$$

for some positive constant $C = C(\nu, T, |\sigma|, |\nabla\sigma|)$, where we have also used that $\lambda_j \sim j$ as recalled in Section 3.1.1.

We proceed to estimate the first term in (3.121). Let us denote $\zeta_N = \Pi_N\xi - \xi_N$ and $\mathbf{v}_N = \mathscr{K} * \zeta_N$. Applying the projection Π_N to (3.1), we have

$$\mathrm{d}\Pi_N\xi + \left[-\nu\Delta\Pi_N\xi + \Pi_N(\mathbf{u}\cdot\nabla\xi)\right]\mathrm{d}t = \sum_{k=1}^{d} \Pi_N\sigma_k\,\mathrm{d}W^k. \tag{3.123}$$

Subtracting (3.12) from (3.123), we obtain that ζ_N satisfies

$$\mathrm{d}\zeta_N + \left[-\nu\Delta\zeta_N + \Pi_N\left(\mathbf{u}\cdot\nabla\xi - \mathbf{u}_N\cdot\nabla\xi_N\right)\right]\mathrm{d}t = 0.$$

Hence, it follows by Itô formula that

$$\mathrm{d}|\zeta_N|^2 + 2\nu|\nabla\zeta_N|^2\,\mathrm{d}t = -2\left(\mathbf{u}\cdot\nabla\xi - \mathbf{u}_N\cdot\nabla\xi_N, \zeta_N\right)\,\mathrm{d}t. \tag{3.124}$$

Denote $Q_N = I - \Pi_N$. Notice that

$$\mathbf{u}\cdot\nabla\xi - \mathbf{u}_N\cdot\nabla\xi_N = \mathbf{u}\cdot\nabla Q_N\xi + Q_N\mathbf{u}\cdot\nabla\Pi_N\xi + \Pi_N\mathbf{u}\cdot\nabla\zeta_N + \mathbf{v}_N\cdot\nabla\Pi_N\xi - \mathbf{v}_N\cdot\nabla\zeta_N.$$

Thus, due to the orthogonality property (3.7),

$$\left(\mathbf{u}\cdot\nabla\xi - \mathbf{u}_N\cdot\nabla\xi_N, \zeta_N\right) = \left(\mathbf{u}\cdot\nabla Q_N\xi, \zeta_N\right) + \left(Q_N\mathbf{u}\cdot\nabla\Pi_N\xi, \zeta_N\right) + \left(\mathbf{v}_N\cdot\nabla\Pi_N\xi, \zeta_N\right). \tag{3.125}$$

We proceed to estimate each term in the right-hand side of (3.125). Invoking (3.6), (3.9) and (3.11), we obtain

$$\begin{aligned}
|\left(\mathbf{u}\cdot\nabla Q_N\xi, \zeta_N\right)| &= |\left(\mathbf{u}\cdot\nabla\zeta_N, Q_N\xi\right)| \le c|\xi|^{1/2}|\nabla\xi|^{1/2}|\nabla\zeta_N||Q_N\xi| \\
&\le \frac{c}{\lambda_{N+1}^{1/2}}|\xi|^{1/2}|\nabla\xi|^{3/2}|\nabla\zeta_N| \le \frac{\nu}{6}|\nabla\zeta_N|^2 + \frac{c}{\nu\lambda_{N+1}}|\xi|\,|\nabla\xi|^3.
\end{aligned} \tag{3.126}$$

Moreover, it follows from (3.8) with $a = 1/2$ that

$$\left|\left(\mathbf{v}_N\cdot\nabla\Pi_N\xi, \zeta_N\right)\right| \le c|\nabla\Pi_N\xi|\,|\zeta_N|^{3/2}\,|\nabla\zeta_N|^{1/2} \le \frac{\nu}{6}|\nabla\zeta_N|^2 + \frac{c}{\nu^{1/3}}|\nabla\xi|^{4/3}|\zeta_N|^2, \tag{3.127}$$

and

$$\begin{aligned}
\left|\left(Q_N\mathbf{u}\cdot\nabla\Pi_N\xi, \zeta_N\right)\right| &= \left|\left(Q_N\mathbf{u}\cdot\nabla\zeta_N, \Pi_N\xi\right)\right| \le c|Q_N\xi|\,|\nabla\zeta_N|\,|\Pi_N\xi|^{1/2}\,|\nabla\Pi_N\xi|^{1/2} \\
&\le \frac{c}{\lambda_{N+1}^{1/2}}|\xi|^{1/2}|\nabla\xi|^{3/2}|\nabla\zeta_N| \le \frac{\nu}{6}|\nabla\zeta_N|^2 + \frac{c}{\nu\lambda_{N+1}}|\xi|\,|\nabla\xi|^3.
\end{aligned} \tag{3.128}$$

Hence, from (3.124), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}|\zeta_N|^2 + \nu|\nabla\zeta_N|^2 \le \frac{c}{\nu^{1/3}}|\nabla\xi|^{4/3}|\zeta_N|^2 + \frac{c}{\nu\lambda_{N+1}}|\xi|\,|\nabla\xi|^3.$$

Ignoring the second term in the right-hand side and applying Gronwall's inequality, recalling that $\zeta_N(0) = 0$, it follows that

$$|\zeta_N(t)|^2 \le \frac{c}{\nu\lambda_{N+1}}\int_0^t |\xi(s)|\,|\nabla\xi(s)|^3 \exp\left(\frac{c}{\nu^{1/3}}\int_s^t |\nabla\xi(\tau)|^{4/3}\,\mathrm{d}\tau\right)\,\mathrm{d}s. \tag{3.129}$$

For some $\alpha > 0$ to be appropriately chosen later, we estimate

$$\frac{c}{\nu^{1/3}} \int_s^t |\nabla \xi(\tau)|^{4/3} \, \mathrm{d}\tau \le \int_s^t \left(\frac{\alpha \nu}{2} |\nabla \xi(\tau)|^2 + C_\alpha \right) \mathrm{d}\tau \le \frac{\alpha \nu}{2} \int_0^T |\nabla \xi(\tau)|^2 \, \mathrm{d}\tau + C_\alpha T, \tag{3.130}$$

where $C_\alpha = c(\alpha^2 \nu^3)^{-1}$ for some positive absolute constant c.

Plugging (3.130) into (3.129), and taking the supremum over $t \in [0, T]$, expected values, and applying Hölder's inequality, it follows that

$$\mathbb{E}\left[\sup_{t \in [0,T]} |\zeta_N(t)|^2 \right] \le c \frac{e^{C_\alpha T}}{\nu \lambda_{N+1}} \left(\mathbb{E} \exp \left(\alpha \nu \int_0^T |\nabla \xi(\tau)|^2 \, \mathrm{d}\tau \right) \right)^{1/2} \left(\mathbb{E}\left[\left(\int_0^T |\xi(s)| \, |\nabla \xi(s)|^3 \, \mathrm{d}s \right)^2 \right] \right)^{1/2}. \tag{3.131}$$

Choosing $0 < \alpha \le \nu/(2|\sigma|^2)$ and invoking Proposition 3.5, we estimate the first term between parentheses above as

$$\mathbb{E} \exp \left(\alpha \nu \int_0^T |\nabla \xi(\tau)|^2 \, \mathrm{d}\tau \right) \le 2 \exp \left(\alpha |\xi_0|^2 \right) \exp \left(\alpha |\sigma|^2 T \right). \tag{3.132}$$

For the last term in (3.131), we estimate

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T |\xi(s)| \, |\nabla \xi(s)|^3 \, \mathrm{d}s \right)^2 \right] &\le \mathbb{E}\left[\sup_{t \in [0,T]} |\xi(t)|^2 |\nabla \xi(t)|^2 \left(\int_0^T |\nabla \xi(s)|^2 \, \mathrm{d}s \right)^2 \right] \\ &\le C \mathbb{E}\left[\left(\sup_{t \in [0,T]} \left(|\xi(t)|^2 + \nu \int_0^t |\nabla \xi(s)|^2 \, \mathrm{d}s \right)^3 \right) \left(\sup_{t \in [0,T]} |\nabla \xi(t)|^2 \right) \right]. \end{aligned}$$

Thus, after applying Hölder's inequality, we obtain from the analogous versions of inequalities (3.13) and (3.14) with $k = 0$ satisfied by $\xi(t)$ that

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T |\xi(s)| \, |\nabla \xi(s)|^3 \, \mathrm{d}s \right)^2 \right] &\le C \left(1 + |\xi_0|^6 \right) \left(1 + |\xi_0|^4 + |\nabla \xi_0|^2 \right) \\ &\le C \exp \left(c\alpha |\xi_0|^2 \right) \left(1 + |\nabla \xi_0|^2 \right), \end{aligned} \tag{3.133}$$

for some positive constant C depending on $\nu, \alpha, T, |\nabla \sigma|$.

Plugging (3.132) and (3.133) into (3.131), we deduce that

$$\mathbb{E}\left[\sup_{t \in [0,T]} |\zeta_N(t)|^2 \right] \le \frac{C}{N} \exp \left(c\alpha |\xi_0|^2 \right) (1 + |\nabla \xi_0|), \tag{3.134}$$

where we again used that $\lambda_j \sim j$.

Now combining (3.134) with (3.122), it follows from (3.121) that

$$\mathbb{E}\left[\sup_{t\in[0,T]}|\xi(t)-\xi_N(t)|^2\right] \leq \frac{2C}{N}\exp\left(c\alpha|\xi_0|^2\right)(1+|\nabla\xi_0|)+\frac{2C}{N}\left(1+|\xi_0|^4+|\nabla\xi_0|^2\right)$$
$$\leq \frac{C}{N}\left[\exp\left(c\alpha|\xi_0|^2\right)+|\nabla\xi_0|^2\right].$$

This finishes the proof. □

We proceed by showing an estimate of the time discretization error as mentioned above. We note that a related result is obtained in Bessaih & Millet (2019) (see also Bessaih & Millet (2021, 2022)), where the authors consider instead the velocity formulation of the 2D stochastic Navier–Stokes equations subject to periodic boundary conditions and either multiplicative or additive noise as in (3.1). In particular, for a semi-implicit Euler time discretization under additive noise analogously as in (3.22), their result yields a strong $L^2(\Omega)$ estimate of the discretization error for the approximating velocity fields under the topology of $L^\infty_{\mathrm{loc},t}L^2_x$ and with order of convergence $1/4$. Our result below provides instead a strong error bound for the approximating vorticity fields in $L^\infty_{\mathrm{loc},t}L^2_x$, which implies an error bound for the corresponding velocity fields in $L^\infty_{\mathrm{loc},t}H^1_x$, to a power p sufficiently small, and with higher order of convergence $1/2$. Notably, according to (3.135) below it follows that p can be almost 2 as $\nu\to\infty$. The main difference in our proof in relation to Bessaih & Millet (2019) concerns the definition of the appropriate localization set in the sample space. Here, we consider a sequence of localization sets that are related to a suitable sequence of discrete stopping times, see (3.146) and (3.154) below.

Proposition 3.18 Fix any $N\in\mathbb{N}$, $\delta,\delta_0>0$ with $\delta\leq\delta_0$, $\sigma\in\dot{\boldsymbol{H}}^1$ and $\xi_0\in\dot{H}^2$. Let $\xi_N(t)$, $t\geq 0$, and $\xi^n_{N,\delta}$, $n\in\mathbb{Z}^+$, be the solutions of (3.12) and (3.20) satisfying $\xi_N(0)=\Pi_N\xi_0$ and $\xi^0_{N,\delta}=\Pi_N\xi_0$, respectively. Then there exist positive absolute constants c_1,c_2 such that if

$$0<\alpha\leq\frac{c_1}{|\sigma|^2}\min\left\{\nu,\frac{1}{\delta_0}\right\}\quad\text{and}\quad 0<p<\frac{2\nu^2\alpha}{c_2+\nu^2\alpha},\tag{3.135}$$

then for every $K\in\mathbb{N}$ and $\tilde{p}\in(0,1/2)$

$$\mathbb{E}\left[\sup_{k\leq K}\left|\xi^k_{N,\delta}(\xi_0)-\xi_N(t_k;\xi_0)\right|^p\right]\leq C\delta^{\tilde{p}p}\left(1+|\nabla\xi_0|^4+|A\xi_0|^2\right)^{p/2}\exp\left(\widetilde{C}\alpha|\xi_0|^2\right),\tag{3.136}$$

where $\widetilde{C}=c(1+\nu\delta_0)$ and $C=C(\tilde{p},p,\nu,\delta_0,T,|\sigma|,|\nabla\sigma|,\alpha)$, with $T\geq(K+1)\delta$. Notably, C and $\widetilde{C}$ are independent of N and δ.

Proof. For each $j\in\mathbb{N}$, let $\zeta^j_{N,\delta}=\zeta^j:=\xi^j_{N,\delta}-\xi_N(t_j)=\xi^j-\xi_N(t_j)$ and $\mathbf{v}^j=\mathscr{K}*\zeta^j$. Integrating (3.12) with respect to $t\in[t_j,t_{j+1}]$,

$$\xi_N(t_{j+1})-\xi_N(t_j)+\int_{t_j}^{t_{j+1}}\left[-\nu\Delta\xi_N(s)+\Pi_N(\mathbf{u}_N(s)\cdot\nabla\xi_N(s))\right]\,\mathrm{d}s$$
$$=\Pi_N\sigma(W(t_{j+1})-W(t_j)).\tag{3.137}$$

Thus, subtracting (3.137) from (3.20) with $n = j+1$, we obtain

$$\zeta^{j+1} - \zeta^j - \nu \int_{t_j}^{t_{j+1}} \left(\Delta\xi^{j+1} - \Delta\xi_N(s) \right) \mathrm{d}s + \int_{t_j}^{t_{j+1}} \left[\Pi_N \left(\mathbf{u}^j \cdot \nabla \xi^{j+1} \right) - \Pi_N(\mathbf{u}_N(s) \cdot \nabla \xi_N(s)) \right] \mathrm{d}s = 0.$$

Taking the inner product with ζ^{j+1} in $\dot{L}^2$ yields

$$|\zeta^{j+1}|^2 + |\zeta^{j+1} - \zeta^j|^2 - |\zeta^j|^2 = 2\nu \int_{t_j}^{t_{j+1}} \left(\Delta\xi^{j+1} - \Delta\xi_N(s), \zeta^{j+1} \right) \mathrm{d}s \\ - 2 \int_{t_j}^{t_{j+1}} \left(\left(\mathbf{u}^j \cdot \nabla \xi^{j+1} \right) - (\mathbf{u}_N(s) \cdot \nabla \xi_N(s)), \zeta^{j+1} \right) \mathrm{d}s. \quad (3.138)$$

Notice that

$$2\nu \int_{t_j}^{t_{j+1}} \left(\Delta\xi^{j+1} - \Delta\xi_N(s), \zeta^{j+1} \right) \mathrm{d}s \\ = -2\nu\delta |\nabla \zeta^{j+1}|^2 + 2\nu \int_{t_j}^{t_{j+1}} \left(\Delta\xi_N(t_{j+1}) - \Delta\xi_N(s), \zeta^{j+1} \right) \mathrm{d}s. \quad (3.139)$$

Integrating by parts the second term in the right-hand side of (3.139), then applying Cauchy–Schwarz and Young's inequalities, it follows that

$$2\nu \left| \int_{t_j}^{t_{j+1}} \left(\Delta\xi_N(t_{j+1}) - \Delta\xi_N(s), \zeta^{j+1} \right) \mathrm{d}s \right| \\ \le 2\nu \int_{t_j}^{t_{j+1}} \left| \nabla\xi_N(t_{j+1}) - \nabla\xi_N(s) \right| \left| \nabla\zeta^{j+1} \right| \mathrm{d}s \\ \le \frac{\nu\delta}{4} \left| \nabla\zeta^{j+1} \right|^2 + c\nu \int_{t_j}^{t_{j+1}} \left| \nabla\xi_N(t_{j+1}) - \nabla\xi_N(s) \right|^2 \mathrm{d}s. \quad (3.140)$$

Now for the second term in the right-hand side of (3.138), first notice that

$$\left((\mathbf{u}^j \cdot \nabla\xi^{j+1}) - (\mathbf{u}_N(s) \cdot \nabla\xi_N(s)), \zeta^{j+1} \right) \quad (3.141) \\ = \left(\mathbf{v}^j \cdot \nabla\xi^{j+1}, \zeta^{j+1} \right) + \left((\mathbf{u}_N(t_j) - \mathbf{u}_N(s)) \cdot \nabla\xi_N(t_{j+1}), \zeta^{j+1} \right) - \left(\mathbf{u}_N(s) \cdot \nabla(\xi_N(s) - \xi_N(t_{j+1})), \zeta^{j+1} \right).$$

We proceed to estimate each term in the right-hand side of (3.141). With (3.8) and Young's inequality, we obtain

$$\begin{aligned}\left|\left(\mathbf{v}^j\cdot\nabla\xi^{j+1},\zeta^{j+1}\right)\right| &\le c|\zeta^j|\,|\nabla\xi^{j+1}|\,|\zeta^{j+1}|^{1/2}|\nabla\zeta^{j+1}|^{1/2}\\ &\le c|\zeta^j|\,|\nabla\xi^{j+1}|\,|\nabla\zeta^{j+1}|\\ &\le \frac{\nu}{8}|\nabla\zeta^{j+1}|^2+\frac{c}{\nu}|\zeta^j|^2|\nabla\xi^{j+1}|^2. \end{aligned}\tag{3.142}$$

Similarly,

$$\left|\left(\mathbf{u}_N(s)\cdot\nabla(\xi_N(s)-\xi_N(t_{j+1})),\zeta^{j+1}\right)\right|\le\frac{\nu}{8}|\nabla\zeta^{j+1}|^2+\frac{c}{\nu}|\xi_N(s)|^2|\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2.\tag{3.143}$$

Now from (3.6) and (3.9), we obtain

$$\begin{aligned}\left|\left((\mathbf{u}_N(t_j)-\mathbf{u}_N(s))\cdot\nabla\xi_N(t_{j+1}),\zeta^{j+1}\right)\right| &= \left|\left((\mathbf{u}_N(t_j)-\mathbf{u}_N(s))\cdot\nabla\zeta^{j+1},\xi_N(t_{j+1})\right)\right|\\ &\le c|\nabla\xi_N(t_j)-\nabla\xi_N(s)|\,|\nabla\zeta^{j+1}|\,|\xi_N(t_{j+1})|\\ &\le\frac{\nu}{8}|\nabla\zeta^{j+1}|^2+\frac{c}{\nu}|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2|\xi_N(t_{j+1})|^2.\end{aligned}\tag{3.144}$$

With (3.139)–(3.144), we obtain from (3.138) the following inequality valid for every $j\in\mathbb{N}$

$$\begin{aligned}&|\zeta^{j+1}|^2+|\zeta^{j+1}-\zeta^j|^2-|\zeta^j|^2+\nu\delta|\nabla\zeta^{j+1}|^2\\ &\qquad\le c\nu\int_{t_j}^{t_{j+1}}|\nabla\xi_N(t_{j+1})-\nabla\xi_N(s)|^2\,\mathrm{d}s+\frac{c\delta}{\nu}|\zeta^j|^2|\nabla\xi^{j+1}|^2\\ &+\frac{c}{\nu}|\xi_N(t_{j+1})|^2\int_{t_j}^{t_{j+1}}|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2\,\mathrm{d}s+\frac{c}{\nu}\int_{t_j}^{t_{j+1}}|\xi_N(s)|^2|\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2\,\mathrm{d}s.\end{aligned}\tag{3.145}$$

Fix $K\in\mathbb{N}$. For each $l\in\mathbb{N}$, we define the following discrete stopping time

$$\kappa_l:=\min\left\{k\ge1\;:\;\sup_{t\in[0,t_{k+2}]}\frac{|\xi_N(t)|^2}{\nu^2}+\frac{3\delta}{\nu}\sum_{j=1}^{k+2}|\nabla\xi^j|^2\ge l\right\}\wedge K.\tag{3.146}$$

We also define a corresponding family of discrete stopping times κ_l^i, $i=0,1,\dots,L$, for some $L=L(l)\in\mathbb{N}$ to be suitably chosen, given by

$$\kappa_l^0:=0;\quad \kappa_l^i:=\min\left\{k\ge\kappa_l^{i-1}+1\;:\;\frac{\delta}{\nu}\sum_{j=\kappa_l^{i-1}+1}^{k+2}|\nabla\xi^j|^2\ge\frac{l}{L}\right\}\wedge\kappa_l,\quad i=1,\dots,L.$$

It is not difficult to show that $\kappa_l^{i-1}<\kappa_l^i$ for all $i=1,\dots,L$, and $\kappa_l^L=\kappa_l$.

Ignoring the second and fourth terms in the left-hand side of (3.145) and summing over $j = \kappa_l^{i-1}, \dots, k$, for $\kappa_l^{i-1} \leq k \leq \kappa_l^i$ and $i \in \{1, \dots, L\}$, it follows that

$$|\zeta^{k+1}|^2 - |\zeta^{\kappa_l^{i-1}}|^2 \leq c\nu \sum_{j=\kappa_l^{i-1}}^{\kappa_l^i} \int_{t_j}^{t_{j+1}} |\nabla\xi_N(t_{j+1}) - \nabla\xi_N(s)|^2 \, \mathrm{d}s + \frac{c\delta}{\nu} \sup_{\kappa_l^{i-1} \leq j \leq \kappa_l^i} |\zeta^j|^2 \sum_{j=\kappa_l^{i-1}+1}^{\kappa_l^i+1} |\nabla\xi^j|^2$$
$$+ \frac{c}{\nu} \sup_{t\in[0,t_{\kappa_l^i+1}]} |\xi_N(t)|^2 \sum_{j=\kappa_l^{i-1}}^{\kappa_l^i} \int_{t_j}^{t_{j+1}} \left[|\nabla\xi_N(t_j) - \nabla\xi_N(s)|^2 + |\nabla\xi_N(s) - \nabla\xi_N(t_{j+1}))|^2\right] \mathrm{d}s. \quad (3.147)$$

Notice that, by the definition of κ_l^i, $i = 1, \dots, L$, we have

$$\frac{\delta}{\nu} \sum_{j=\kappa_l^{i-1}+1}^{\kappa_l^i+1} |\nabla\xi^j|^2 \leq \frac{l}{L} \quad \text{for all } i = 1, \dots, L.$$

Choose $L \in \mathbb{N}$ as

$$L = \min\{j \in \mathbb{N} \,:\, cl/j \leq 1/2\}, \tag{3.148}$$

with $c > 0$ as in the second term in the right-hand side of (3.147). We then estimate this term as

$$\frac{c\delta}{\nu} \sup_{\kappa_l^{i-1} \leq j \leq \kappa_l^i} |\zeta^j|^2 \sum_{j=\kappa_l^{i-1}+1}^{\kappa_l^i+1} |\nabla\xi^j|^2 \leq c\frac{l}{L} \sup_{\kappa_l^{i-1} \leq j \leq \kappa_l^i} |\zeta^j|^2 \leq \frac{1}{2} \sup_{\kappa_l^{i-1} \leq j \leq \kappa_l^i+1} |\zeta^j|^2. \tag{3.149}$$

Notice that (3.147) is in fact valid for all $k = (\kappa_l^{i-1} - 1), \kappa_l^{i-1}, \dots, \kappa_l^i$. Thus, taking in (3.147) the supremum over $k = (\kappa_l^{i-1} - 1), \kappa_l^{i-1}, \dots, \kappa_l^i$ and invoking (3.149), we obtain

$$\sup_{\kappa_l^{i-1} \leq j \leq \kappa_l^i+1} |\zeta^j|^2 \leq 2\left|\zeta^{\kappa_l^{i-1}}\right|^2 + c\nu \sum_{j=\kappa_l^{i-1}}^{\kappa_l^i} \int_{t_j}^{t_{j+1}} \left|\nabla\xi_N(t_{j+1}) - \nabla\xi_N(s)\right|^2 \mathrm{d}s$$
$$+ \frac{c}{\nu} \sup_{t\in\left[0,t_{\kappa_l^i+1}\right]} |\xi_N(t)|^2 \sum_{j=\kappa_l^{i-1}}^{\kappa_l^i} \int_{t_j}^{t_{j+1}} \left[\left|\nabla\xi_N(t_j) - \nabla\xi_N(s)\right|^2 + \left|\nabla\xi_N(s) - \nabla\xi_N(t_{j+1})\right|^2\right] \mathrm{d}s. \quad (3.150)$$

From the definition of κ_l and since $\kappa_l^i \leq \kappa_l^L = \kappa_l$ for all $i = 1, \dots, L$, we have

$$\sup_{t\in\left[0,t_{\kappa_l^i+1}\right]} \frac{|\xi_N(t)|^2}{\nu^2} \leq \sup_{t\in[0,t_{\kappa_l+1}]} \frac{|\xi_N(t)|^2}{\nu^2} \leq \sup_{t\in[0,t_{\kappa_l+1}]} \frac{|\xi_N(t)|^2}{\nu^2} + \frac{3\delta}{\nu} \sum_{j=1}^{\kappa_l+1} |\nabla\xi^j|^2 \leq l. \tag{3.151}$$

Using (3.151) to estimate the third term in the right-hand side of (3.150), it follows that

$$\sup_{\kappa_l^{i-1}\le j\le \kappa_l^i+1} |\zeta^j|^2 \le 2|\zeta^{\kappa_l^{i-1}}|^2 + \mathscr{A}_i,$$

where

$$\mathscr{A}_i := c\nu l \sum_{j=\kappa_l^{i-1}}^{\kappa_l^i} \int_{t_j}^{t_{j+1}} \Big[|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2 + |\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2\Big]\,\mathrm{d}s. \tag{3.152}$$

Hence, for every $i=1,\dots,L$,

$$\begin{aligned}
\sup_{j\le\kappa_l^i}|\zeta^j|^2 &\le \sup_{j\le \kappa_l^{i-1}}|\zeta^j|^2 + \sup_{\kappa_l^{i-1}\le j\le\kappa_l^i+1}|\zeta^j|^2\\
&\le \sup_{j\le \kappa_l^{i-1}}|\zeta^j|^2 + 2|\zeta^{\kappa_l^{i-1}}|^2 + \mathscr{A}_i\\
&\le 3\sup_{j\le \kappa_l^{i-1}}|\zeta^j|^2 + \mathscr{A}_i.
\end{aligned}$$

By induction, it follows that for every $l\in\mathbb{N}$ and $L=L(l)\in\mathbb{N}$ as in (3.148),

$$\sup_{j\le\kappa_l}|\zeta^j|^2 = \sup_{j\le\kappa_l^L}|\zeta^j|^2 \le 3^L|\zeta^0|^2 + \sum_{i=1}^L 3^{L-i}\mathscr{A}_i = \sum_{i=1}^L 3^{L-i}\mathscr{A}_i, \tag{3.153}$$

since $\zeta^0=0$.

Now, for $K\in\mathbb{N}$ as in (3.146), we define for every $l\in\mathbb{N}$

$$\Omega_l^K := \left\{\omega\in\Omega \,:\, l-1 \le \sup_{t\in[0,t_{K+2}]}\frac{|\xi_N(t)|^2}{\nu^2} + \frac{3\delta}{\nu}\sum_{j=1}^{K+2}|\nabla\xi^j|^2 < l\right\}. \tag{3.154}$$

Since $\Omega=\bigcup_{l\in\mathbb{N}}\Omega_l^K$, we obtain for any $p\in(0,2)$

$$\begin{aligned}
\mathbb{E}[\sup_{j\le K}|\zeta^j|^p] = \mathbb{E}\left(\sup_{j\le K}|\zeta^j|^p\sum_{l=1}^\infty \mathbb{1}_{\Omega_l^K}\right) &= \sum_{l=1}^\infty \mathbb{E}\left(\sup_{j\le K}|\zeta^j|^p\mathbb{1}_{\Omega_l^K}\right)\\
&\le \sum_{l=1}^\infty\left(\mathbb{E}[\sup_{j\le K}|\zeta^j|^2\mathbb{1}_{\Omega_l^K}]\right)^{p/2}\mathbb{P}(\Omega_l^K)^{\frac{2-p}{2}},
\end{aligned} \tag{3.155}$$

where the last step follows from Hölder's inequality. Moreover, from the definition of κ_l and Ω_l^K, it follows that if $\omega \in \Omega_l^K$ then $\kappa_l(\omega) = K$. Thus, from (3.153) and (3.152),

$$\begin{aligned}
\mathbb{E}\left[\sup_{j\le K}|\zeta^j|^2 \mathbb{1}_{\Omega_l^K}\right] &= \mathbb{E}\left[\sup_{j\le \kappa_l}|\zeta^j|^2 \mathbb{1}_{\Omega_l^K}\right] \le \sum_{i=1}^{L} 3^{L-i}\mathbb{E}(\mathbb{1}_{\Omega_l^K}\mathscr{A}_i) \\
&\le c\nu l 3^{L-1}\mathbb{E}\left(\mathbb{1}_{\Omega_l^K}\sum_{i=1}^{L}\sum_{j=\kappa_l^{i-1}}^{\kappa_l^i}\int_{t_j}^{t_{j+1}}\left[|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2+|\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2\right]\mathrm{d}s\right) \\
&\le c\nu l 3^{L-1}\mathbb{E}\sum_{j=0}^{K}\int_{t_j}^{t_{j+1}}\left[|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2+|\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2\right]\mathrm{d}s \\
&= c\nu l 3^{L-1}\sum_{j=0}^{K}\int_{t_j}^{t_{j+1}}\left(\mathbb{E}\left[|\nabla\xi_N(t_j)-\nabla\xi_N(s)|^2\right]+\mathbb{E}\left[|\nabla\xi_N(s)-\nabla\xi_N(t_{j+1})|^2\right]\right)\mathrm{d}s.
\end{aligned}$$

Fix any $T \ge (K+1)\delta$. From Theorem 3.7 with $m = 2$, we have that for every $\tilde{p} \in (0, 1/2)$ and $s, t \in [0, T]$

$$\mathbb{E}[|\nabla\xi_N(t) - \nabla\xi_N(s)|^2] \le CR_0|t-s|^{2\tilde{p}}, \tag{3.156}$$

where $R_0 = (1 + |\xi_0|^8 + |\nabla\xi_0|^4 + |A\xi_0|^2)$, and C is a positive constant depending on $\tilde{p}, T, \nu, |\sigma|, |\nabla\sigma|$. Hence,

$$\mathbb{E}\left[\sup_{j\le K}|\zeta^j|^2 \mathbb{1}_{\Omega_l^K}\right] \le c\nu l 3^{L-1}\sum_{j=0}^{K} CR_0\delta^{2\tilde{p}+1} \le CR_0\delta^{2\tilde{p}} l 3^{L-1},$$

where $C = C(\tilde{p}, T, \nu, |\sigma|, |\nabla\sigma|)$.

Moreover, from the definition of L in (3.148) it follows that $2cl \le L \le 2cl + 1$, so that $l3^{L-1} \le 3^{cl}$ and

$$\mathbb{E}\left[\sup_{j\le K}|\zeta^j|^2 \mathbb{1}_{\Omega_l^K}\right] \le CR_0\delta^{2\tilde{p}}3^{cl}. \tag{3.157}$$

From the definition of the set Ω_l^K in (3.154) and invoking Markov's inequality, it follows that for any $\tilde{\alpha} > 0$

$$\begin{aligned}\mathbb{P}(\Omega_l^K) &\leq \mathbb{P}\left(\sup_{t\in[0,t_{K+2}]}\frac{|\xi_N(t)|^2}{\nu^2} + \frac{3\delta}{\nu}\sum_{j=1}^{K+2}|\nabla\xi^j|^2 \geq l-1\right)\\ &= \mathbb{P}\left(\exp\left(\tilde{\alpha}\sup_{t\in[0,t_{K+2}]}\frac{|\xi_N(t)|^2}{\nu^2} + \tilde{\alpha}\frac{3\delta}{\nu}\sum_{j=1}^{K+2}|\nabla\xi^j|^2\right) \geq e^{\tilde{\alpha}(l-1)}\right)\\ &\leq e^{-\tilde{\alpha}(l-1)}\mathbb{E}\left(\exp\left(\tilde{\alpha}\sup_{t\in[0,t_{K+2}]}\frac{|\xi_N(t)|^2}{\nu^2} + \tilde{\alpha}\frac{3\delta}{\nu}\sum_{j=1}^{K+2}|\nabla\xi^j|^2\right)\right)\\ &\leq e^{-\tilde{\alpha}(l-1)}\left(\mathbb{E}\exp\left(2\tilde{\alpha}\sup_{t\in[0,t_{K+2}]}\frac{|\xi_N(t)|^2}{\nu^2}\right)\right)^{1/2}\left(\mathbb{E}\exp\left(c\tilde{\alpha}\frac{\delta}{\nu}\sum_{j=1}^{K+2}|\nabla\xi^j|^2\right)\right)^{1/2}.\end{aligned}$$

Now we assume that $0 < \tilde{\alpha} \leq \tilde{c}\nu^2|\sigma|^{-2}\min\{\nu, \delta_0^{-1}\}$ for some absolute constant $\tilde{c} > 0$ that is small enough so we can invoke the bounds (3.62) from Proposition 3.15 and (3.16) from Proposition 3.5 to obtain that

$$\begin{aligned}\mathbb{P}\Big(\Omega_l^K\Big) &\leq C\,e^{-\tilde{\alpha}(l-1)}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right)\exp\left(c\frac{\tilde{\alpha}}{\nu^2}|\sigma|^2(K+2)\delta\right)\\ &\leq C\,e^{-\tilde{\alpha}(l-1)}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right),\end{aligned}\tag{3.158}$$

where $C = C(\nu, \delta_0, T, |\sigma|)$ and $\widetilde{C} = c(1+\nu\delta_0)$.

Therefore, it follows from from (3.155), (3.157) and (3.158) that

$$\begin{aligned}\mathbb{E}\left[\sup_{j\leq K}|\zeta^j|^p\right] &\leq \sum_{l=1}^{\infty}\Big(CR_0\delta^{2\tilde{p}}3^{cl}\Big)^{p/2}\left(C\,e^{-\tilde{\alpha}(l-1)}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right)\right)^{\frac{2-p}{2}}\\ &\leq CR_0^{p/2}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right)\delta^{\tilde{p}p}\left(\sum_{l=1}^{\infty}3^{cpl}e^{-\frac{\tilde{\alpha}(2-p)}{2}l}\right)e^{\tilde{\alpha}\frac{(2-p)}{2}}.\end{aligned}\tag{3.159}$$

For the terms depending on the initial datum ξ_0, we have by the definition of R_0 in (3.156) that

$$\begin{aligned}R_0^{p/2}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right) &= \Big(1+|\xi_0|^8+|\nabla\xi_0|^4+|A\xi_0|^2\Big)^{p/2}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right)\\ &\leq C\Big(1+|\nabla\xi_0|^4+|A\xi_0|^2\Big)^{p/2}\exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right),\end{aligned}\tag{3.160}$$

for $C = C(p, \nu, \delta_0, |\sigma|, \tilde{\alpha})$.

Moreover, for the term involving the sum in (3.159), notice that

$$\sum_{l=1}^{\infty} 3^{cpl} e^{-\tilde{\alpha}\frac{(2-p)}{2}l} = \sum_{l=1}^{\infty} e^{\ln(3)cpl} e^{-\tilde{\alpha}\frac{(2-p)}{2}l} = \sum_{l=1}^{\infty} e^{-\gamma l}, \tag{3.161}$$

where $\gamma = [\tilde{\alpha}(2-p)/2] - cp$. We choose $p < 2\tilde{\alpha}/(c+\tilde{\alpha})$, so that $\gamma > 0$ and, consequently, the last sum in (3.161) is finite. We thus conclude from (3.159) and (3.160) that

$$\mathbb{E}\left[\sup_{j\leq K} |\zeta^j|^p\right] \leq C\delta^{\tilde{p}p}\Big(1 + |\nabla\xi_0|^4 + |A\xi_0|^2\Big)^{p/2} \exp\left(\widetilde{C}\frac{\tilde{\alpha}}{\nu^2}|\xi_0|^2\right),$$

for some positive constant $C = C(\tilde{p}, p, \nu, \delta_0, T, |\sigma|, |\nabla\sigma|, \tilde{\alpha})$. This shows (3.136) by denoting $\alpha := \tilde{\alpha}\nu^{-2}$, and finishes the proof. □

3.4 *Uniform in time weak convergence of the numerical scheme*

This section focuses on the application of Theorem 2.5 and Theorem 2.8 to the space-time discretization of the 2D stochastic Navier–Stokes equations introduced in (3.20). Following the notation from Theorem 2.5, similarly as in Section 3.2 we take $(X, \|\cdot\|) = (\dot{L}^2, |\cdot|)$ and again consider Λ to be the class of distance functions $\rho_{\varepsilon,s}$, $\varepsilon > 0$, $s \in (0,1]$, defined in (3.26). We let Θ be the set of pairs $\{(N,\delta) : N \in \mathbb{N},\ \delta > 0\} \cup \{(\infty, 0)\}$. Then, for every $\theta = (N,\delta)$ with $N \in \mathbb{N}$ and $\delta > 0$, we let $\{P_t^\theta\}_{t\geq 0}$ be the family of Markov kernels $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ associated to the numerical scheme (3.22), defined in (3.28) and (3.25). For $\theta_0 := (\infty, 0)$, we let $\{P_t^{\theta_0}\}_{t\geq 0}$ be the Markov semigroup $\{P_t\}_{t\geq 0}$ associated to the 2D SNSE (3.1), and defined in (3.5).

Regarding assumptions (H1)–(H4) of Theorem 2.5, only (H3) requires extra work to be verified. This is done in the following proposition, whose proof follows crucially from the strong error estimates obtained in Proposition 3.17 and Proposition 3.18 above, combined with the exponential Lyapunov inequalities from Proposition 3.11 and Proposition 3.6.

PROPOSITION 3.19 Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \leq \delta_0$, and $\sigma \in \dot{\boldsymbol{H}}^1$. Let $\{P_t\}_{t\geq 0}$ and $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ be the corresponding family of Markov kernels associated to systems (3.1) and (3.22), respectively, as defined in (3.5) and (3.28). Then, there exists a positive absolute constant c_1 such that if

$$\alpha' = \frac{c_1}{|\sigma|^2} \min\left\{\nu, \frac{1}{\delta_0}\right\} \tag{3.162}$$

then for every $\alpha \in (0, \alpha']$, $s \in (0,1]$, $\varepsilon > 0$ and $\tilde{p} \in (0, 1/2)$, it holds that

$$\begin{aligned} &\mathscr{W}_{\varepsilon,s,\alpha}\Big(\mathscr{P}_t^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\Big) \\ &\quad\leq \frac{C\,e^{C't}}{\varepsilon^{1/2}}\Big[\max\{\delta^s, \delta^{p/2}\}^{\tilde{p}/2} + N^{-s/4}\Big] \exp\Big(\widetilde{C}\alpha'|\xi_0|^2\Big)\Big(1 + |\nabla\xi_0| + |A\xi_0|^{1/2}\Big)^s, \end{aligned} \tag{3.163}$$

for every $t \geq 0$ and $\xi_0 \in \dot{H}^2$. Here, $0 < p < \frac{2\nu^2\alpha'}{c_2+\nu^2\alpha'}$ for some absolute constant c_2, and $\widetilde{C} = c(1+\nu\delta_0)$, $C' = C'(\nu, |\sigma|)$, $C = C(\tilde{p}, p, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \alpha)$.

Proof. Fix any $\alpha \in (0, \alpha']$, $s \in (0,1]$, $\varepsilon > 0$, $\tilde{p} \in (0, 1/2)$, $\xi_0 \in \dot{H}^2$ and $t \geq 0$. Let $n \in \mathbb{Z}^+$ such that $t \in [n\delta, (n+1)\delta)$. It follows immediately from the definitions of $\{P_t\}_{t\geq 0}$, $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ and $\mathscr{W}_{\varepsilon,s,\alpha}$ that

$$\begin{aligned}\mathscr{W}_{\varepsilon,s,\alpha}\left(\mathscr{P}_t^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\right) &\leq \mathscr{W}_{\varepsilon,s,\alpha'}\left(\mathscr{P}_t^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\right) = \mathscr{W}_{\varepsilon,s,\alpha'}\left(P_n^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\right) \\ &\leq \frac{1}{\varepsilon^{1/2}}\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi(t;\xi_0)\right|^{s/2} \exp\left(\alpha'\left|\xi^n(\xi_0)\right|^2 + \alpha'|\xi(t;\xi_0)|^2\right)\right].\end{aligned}$$

By Hölder's inequality,

$$\begin{aligned}&\mathscr{W}_{\varepsilon,s,\alpha'}\left(\mathscr{P}_t^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\right) \\ &\leq \frac{1}{\varepsilon^{1/2}}\left(\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi(t;\xi_0)\right|^s\right]\right)^{1/2}\left(\mathbb{E}\exp\left(4\alpha'\left|\xi^n(\xi_0)\right|^2\right)\right)^{1/4}\left(\mathbb{E}\exp\left(4\alpha'|\xi(t;\xi_0)|^2\right)\right)^{1/4}.\end{aligned} \tag{3.164}$$

We assume that the constants c_1 in (3.162) and c_2 in the definition of p are sufficiently small so that the results of Proposition 3.11, Proposition 3.6, Proposition 3.17 and Proposition 3.18 can be applied in the estimates to follow. In particular, invoking the exponential Lyapunov inequalities (3.32) from Proposition 3.11 and the analogous version of (3.17) for $\xi(t)$, $t \geq 0$, from Proposition 3.6, respectively, we estimate the last two terms between parentheses in (3.164) as

$$\begin{aligned}\left(\mathbb{E}\exp\left(4\alpha'\left|\xi^n(\xi_0)\right|^2\right)\right)^{1/4}\left(\mathbb{E}\exp\left(4\alpha'\left|\xi(t;\xi_0)\right|^2\right)\right)^{1/4} &\leq C\exp\left(\frac{2\alpha'|\xi_0|^2}{(1+\nu\lambda_1\delta)^n}\right)\exp\left(\alpha' e^{-\nu t}|\xi_0|^2\right) \\ &\leq C\exp\left(3\alpha'|\xi_0|^2\right),\end{aligned} \tag{3.165}$$

where $C = C(\nu, \delta_0, |\sigma|)$.

Regarding the first term between parentheses in (3.164), we first estimate as

$$\begin{aligned}&\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi(t;\xi_0)\right|^s\right] \\ &\leq \mathbb{E}\left[\left|\xi^n(\xi_0) - \xi_N(n\delta;\xi_0)\right|^s\right] + \mathbb{E}\left[\left|\xi_N(n\delta;\xi_0) - \xi(n\delta;\xi_0)\right|^s\right] + \mathbb{E}\left[\left|\xi(n\delta;\xi_0) - \xi(t;\xi_0)\right|^s\right],\end{aligned} \tag{3.166}$$

where we recall that $s \in (0,1]$. We proceed to estimate the terms in the right-hand side of (3.166) by invoking Theorem 3.7, Proposition 3.17 and Proposition 3.18 with $T = t$. Here we will write the t-dependence of the constant C from (3.18), (3.120) and (3.136) explicitly as $e^{C't}$, for some constant $C' = C'(\nu, |\sigma|)$, as can easily be seen from the corresponding proofs.

In particular, invoking inequality (3.120) from Proposition 3.17 and Hölder's inequality, we estimate the second term in the right-hand side of (3.166) as

$$\begin{aligned}\mathbb{E}\Big[\Big|\xi_N(n\delta;\xi_0)-\xi(n\delta;\xi_0)\Big|^s\Big] &\le \left(\mathbb{E}\Big[\Big|\xi_N(n\delta;\xi_0)-\xi(n\delta;\xi_0)\Big|^2\Big]\right)^{s/2}\\ &\le \left(\frac{C\,e^{C't}}{N}\Big[\exp\left(c\alpha'|\xi_0|^2\right)+|\nabla\xi_0|^2\Big]\right)^{s/2}\\ &\le \frac{C\,e^{C't}}{N^{s/2}}\exp\left(c\alpha'|\xi_0|^2\right)\left(1+|\nabla\xi_0|^2\right)^{s/2},\end{aligned} \tag{3.167}$$

where $C=C(\nu,|\sigma|,|\nabla\sigma|,\alpha')$.

For the third term in the right-hand side of (3.166), we invoke the analogous version of inequality (3.18) from Theorem 3.7 with $\xi(t)$, and obtain

$$\mathbb{E}\Big[|\xi(n\delta;\xi_0)-\xi(t;\xi_0)|^s\Big] \le \left(\mathbb{E}|\xi(n\delta;\xi_0)-\xi(t;\xi_0)|\right)^s \le C\,e^{C't}\delta^{s\tilde{p}}\left(1+|\xi_0|^4+|\nabla\xi_0|^2\right)^s, \tag{3.168}$$

where $C=C(\tilde{p},\nu,|\sigma|,|\nabla\sigma|)$.

Finally, to estimate the first term in the right-hand side of (3.166), let us first assume that $s<p$, with p as in (3.135). In this case, it follows by Hölder's inequality and (3.136) from Proposition 3.18 that

$$\begin{aligned}\mathbb{E}\Big[\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^s\Big] &\le \left(\mathbb{E}\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^p\right)^{s/p}\\ &\le \left(C\,e^{C't}\delta^{\tilde{p}p}\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1+|\nabla\xi_0|^4+|A\xi_0|^2\right)^{p/2}\right)^{s/p}\\ &\le C\,e^{C't}\delta^{\tilde{p}s}\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1+|\nabla\xi_0|^4+|A\xi_0|^2\right)^{s/2},\end{aligned} \tag{3.169}$$

where $\widetilde{C}=c(1+\nu\delta_0)$ and $C=C(\tilde{p},p,\nu,\delta_0,|\sigma|,|\nabla\sigma|)$. On the other hand, if $s\ge p$ we proceed as follows

$$\begin{aligned}\mathbb{E}\Big[\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^s\Big] &= \mathbb{E}\left[\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^{p/2}\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^{s-\frac{p}{2}}\right]\\ &\le \mathbb{E}\left[\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^{p/2}\left(|\xi^n(\xi_0)|^{s-\frac{p}{2}}+|\xi_N(n\delta;\xi_0)|^{s-\frac{p}{2}}\right)\right]\\ &\le C\,\mathbb{E}\left[\Big|\xi^n(\xi_0)-\xi_N(n\delta;\xi_0)\Big|^{p/2}\left(\exp\left(\alpha'\Big|\xi^n(\xi_0)\Big|^2\right)+\exp\left(\alpha'\Big|\xi_N(n\delta;\xi_0)\Big|^2\right)\right)\right],\end{aligned}$$

where $C = C(\alpha')$. Thus, by Hölder's inequality and inequalities (3.32), (3.17) and (3.136) from Proposition 3.11, Proposition 3.6 and Proposition 3.18, respectively, we obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi_N(n\delta;\xi_0)\right|^s\right] \\
&\quad \leq C\left(\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi_N(n\delta;\xi_0)\right|^p\right]\right)^{1/2}\left[\mathbb{E}\exp\left(2\alpha'\left|\xi^n(\xi_0)\right|^2\right) + \mathbb{E}\exp\left(2\alpha'\left|\xi_N(n\delta;\xi_0)\right|^2\right)\right]^{1/2} \\
&\quad \leq C\left(e^{C't}\delta^{\tilde{p}p}\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0|^4 + |A\xi_0|^2\right)^{p/2}\right)^{1/2} \\
&\qquad \cdot\left[\exp\left(\frac{4\alpha'|\xi_0|^2}{(1+\nu\lambda_1\delta)^n}\right) + \exp\left(2\alpha' e^{-\nu n\delta}|\xi_0|^2\right)\right]^{1/2} \\
&\quad \leq C\, e^{C't}\delta^{\tilde{p}p/2}\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0|^4 + |A\xi_0|^2\right)^{s/4},
\end{aligned} \tag{3.170}
$$

where $C = C(\tilde{p}, p, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \alpha')$. Combining (3.169) and (3.170), it thus follows that for every $s \in (0, 1]$

$$
\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi_N(n\delta;\xi_0)\right|^s\right] \leq C\, e^{C't}\max\{\delta^s, \delta^{p/2}\}^{\tilde{p}}\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0|^4 + |A\xi_0|^2\right)^{s/2}. \tag{3.171}
$$

Plugging (3.167), (3.168) and (3.171) into (3.166), we thus have

$$
\mathbb{E}\left[\left|\xi^n(\xi_0) - \xi(n\delta;\xi_0)\right|^s\right] \leq C\, e^{C't}\left[\max\{\delta^s, \delta^{p/2}\}^{\tilde{p}} + N^{-s/2}\right]\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0|^4 + |A\xi_0|^2\right)^{s/2}. \tag{3.172}
$$

Now, plugging (3.165) and (3.172) into (3.164), we deduce that

$$
\begin{aligned}
&\mathscr{W}_{\varepsilon,s,\alpha'}\left(\mathscr{P}_t^{N,\delta}(\xi_0,\cdot), P_t(\xi_0,\cdot)\right) \\
&\qquad\qquad \leq \frac{C\, e^{C't}}{\varepsilon^{1/2}}\left[\max\{\delta^s, \delta^{p/2}\}^{\tilde{p}/2} + N^{-s/4}\right]\exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0| + |A\xi_0|^{1/2}\right)^s,
\end{aligned}
$$

with $C = C(\tilde{p}, p, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \alpha')$, and we recall that $\widetilde{C} = c(1 + \nu\delta_0)$, $C' = C'(\nu, |\sigma|)$. This shows (3.163) and concludes the proof. □

Before proceeding with the application of Theorem 2.5 and Theorem 2.8 within this setting, we present the following lemma showing finiteness of certain moments for the invariant measure of P_t, $t \geq 0$. This will ensure that the terms in (2.27) and (2.39) concerning $\mu_{\theta_0} = \mu_*$ are finite.

LEMMA 3.20 Fix any $\sigma \in \dot{\boldsymbol{L}}^2$, and let μ_* be an invariant measure of the corresponding Markov semigroup P_t, $t \geq 0$, defined in (3.5). Then, the following statements hold:

1. For every $\alpha > 0$ satisfying condition (3.15),

$$\int_{\dot{L}^2} \exp\left(\alpha|\xi_0|^2\right)\mu_*(\mathrm{d}\xi_0) \leq C, \tag{3.173}$$

2. for some constant $C = C(\nu, |\sigma|)$.
3. Suppose additionally that $\sigma \in \dot{\boldsymbol{H}}^k$, for some fixed $k \in \mathbb{N}$. Then, for every $m \in \mathbb{N}$,

$$\int_{\dot{L}^2} \|\xi_0\|_{\dot{H}^k}^{2m}\mu_*(\mathrm{d}\xi_0) \leq C, \tag{3.174}$$

4. for some constant $C = C(\nu, m, k, \|\sigma\|_{\dot{H}^k})$. Consequently, μ_* is supported in $\dot{H}^k$, i.e.

$$\mu_*(\dot{L}^2 \setminus \dot{H}^k) = 0.$$

Proof. The proof of (3.173) follows as a consequence of the analogous version of the exponential Lyapunov inequality (3.17) for $\xi(t)$, $t \geq 0$, similarly as in the proof of Kuksin & Shirikyan (2012, Theorem 2.5.3). The subsequent bound (3.174) then follows by combining (3.173) with inequality (3.14) for $\xi(t)$, $t \geq 0$, by estimating $1 + |\xi_0|^{2mp} \leq C(m, p, \alpha)\exp(\alpha|\xi_0|^2)$. We omit further details. □

We are now ready to apply Theorem 2.5 and derive a bias estimate between invariant measures of $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ and $\{P_t\}_{t\geq 0}$.

THEOREM 3.21 (long time bias estimate). Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \leq \delta_0$. Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \dot{\boldsymbol{H}}^2$ such that

$$\Pi_K \dot{L}^2 \subset range(\sigma) \quad \text{and} \quad \lambda_{K+1} \geq \frac{c}{\nu}\max\left\{1, \frac{1}{\delta_0}, \frac{\delta_0^2|\sigma|^4}{\nu^3}, \frac{|\sigma|^4}{\nu^5}\right\} \tag{3.175}$$

for some absolute constant $c > 0$. Let $\{P_t\}_{t\geq 0}$ and $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ be the corresponding family of Markov kernels associated to systems (3.1) and (3.22), respectively, as defined in (3.5) and (3.28). Let μ_* and $\mu_*^{N,\delta}$ be invariant measures for $\{P_t\}_{t\geq 0}$ and $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$, respectively.

Then, there exists $\alpha_* > 0$ such that for each fixed $\alpha \in (0, \alpha_*]$ there exists $\varepsilon > 0$ and $s \in (0, 1]$ for which the following inequality holds for every $\tilde{p} \in (0, 1/2)$

$$\mathscr{W}_{\varepsilon,s,\alpha}\left(\mu_*^{N,\delta}, \mu_*\right) \leq Cg(N, \delta), \tag{3.176}$$

where

$$g(N, \delta) = \max\{\delta^s, \delta^{p/2}\}^{\tilde{p}/2} + N^{-s/4}, \tag{3.177}$$

with

$$0 < p < \frac{2\nu^2\alpha'}{c_2 + \nu^2\alpha'}, \quad \alpha' = \frac{c_1}{|\sigma|^2} \min\left\{\nu, \frac{1}{\delta_0}\right\}, \tag{3.178}$$

for some absolute constants c_1, c_2, and where $C = C(\varepsilon, s, \alpha, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \|\sigma\|_{\dot{H}^2}, \tilde{p})$.

Proof. We verify that all assumptions from Theorem 2.5 are satisfied with the choices of $(X, \|\cdot\|)$, Λ, Θ, $\{P_t^\theta\}_{t\geq 0}$ and $\{P_t^{\theta_0}\}_{t\geq 0}$ taken in the introduction to this section.

Assumption 2.5 follows as a consequence of Proposition A.1 applied to the metric $\rho_{\varepsilon,s}$. Indeed, items (i) and (ii) of Proposition A.1 hold with $M = K = 1$. Moreover, taking $c = 1$ in item (iii), we notice that if $\rho_{\varepsilon,s}(\xi_1, \xi_2) < 1$ then $|\xi_1 - \xi_2|^s < \varepsilon$. Thus, by triangle inequality, $|\xi_1|^2 < 2|\xi_2|^2 + 2\varepsilon^{2/s}$, so that item (iii) holds with $\gamma = 2$ and $C = 2\varepsilon^{2/s}$. It thus follows from (A2) and the definition of $\tilde{K}$ inside the proof that

$$\rho_{\varepsilon,s,\alpha}(\xi_1, \xi_2) \leq \exp(2\alpha\varepsilon^{2/s}) \left[\rho_{\varepsilon,s,2\alpha}(\xi_1, \xi_3) + \rho_{\varepsilon,s,2\alpha}(\xi_3, \xi_2)\right] \quad \text{for all } \xi_1, \xi_2, \xi_3 \in \dot{L}^2,$$

as desired.

Regarding assumption 2.5, the existence of an invariant measure for $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ is shown in Corollary 3.10. Whereas the existence of an invariant measure for $\{P_t\}_{t\geq 0}$, as mentioned in Section 3.1.1, is a well-known result that is valid for even more general noise structures than we consider here, see e.g. Flandoli (1994).

Assumption 2.5 follows as a direct consequence of Theorem 3.9. Finally, assumption 2.5 follows from Proposition 3.19, with $g(N, \delta)$, $R(t)$ and $f(\xi_0)$ given for any fixed $\varepsilon > 0$ and $s \in (0, 1]$ as

$$g(N, \delta) = \max\{\delta^s, \delta^{p/2}\}^{\tilde{p}/2} + N^{-s/4}, \quad \text{for all } N \in \mathbb{N},\ 0 < \delta \leq \delta_0, \tag{3.179}$$

$$R(t) = \frac{C\, e^{C't}}{\varepsilon^{1/2}}, \quad \text{for all } t \geq 0, \tag{3.180}$$

$$f(\xi_0) = \begin{cases} \exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)(1 + |\nabla\xi_0| + |A\xi_0|^{1/2})^s & \text{for } \xi_0 \in \dot{H}^2, \\ \infty & \text{for } \xi_0 \in \dot{L}^2 \backslash \dot{H}^2, \end{cases} \tag{3.181}$$

with $p, \tilde{p}, \alpha', C, C', \widetilde{C}$ as in (3.163).

Therefore, it follows from Theorem 2.5 that

$$\mathscr{W}_{\varepsilon,s,\alpha}\left(\mu_*^{N,\delta}, \mu_*\right) \leq C g(N, \delta) \int_{\dot{L}^2} f(\xi_0)\mu_*(\mathrm{d}\xi_0), \tag{3.182}$$

where $C = C(\varepsilon, s, \alpha, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \tilde{p})$. Moreover, from the definitions of $\widetilde{C}$ and α' given in Proposition 3.19 it is not difficult to show that $\widetilde{C}\alpha' \leq c\nu|\sigma|^{-2}$. With this, we may apply Hölder's inequality and Lemma 3.20 to obtain that $\int_{\dot{L}^2} f(\xi_0)\mu_*(\mathrm{d}\xi_0) \leq C$, for $C = C(\nu, \|\sigma\|_{\dot{H}^2})$. Plugging this into (3.182) we conclude (3.176). ☐

We conclude this section by applying Theorem 2.8 to show convergence in Wasserstein distance for $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ towards $\{P_t\}_{t\geq 0}$, and consequently weak convergence for $\xi_{N,\delta}^n$, $n \in \mathbb{Z}^+$, towards $\xi(t)$, $t \geq 0$, as a result of Corollary 2.11.

THEOREM 3.22 (Uniform in time weak convergence). Fix any $N \in \mathbb{N}$, $\delta, \delta_0 > 0$ with $\delta \leq \delta_0$. Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \dot{\boldsymbol{H}}^2$ satisfying (3.175). Let $\{P_t\}_{t\geq 0}$ and $\{\mathscr{P}_t^{N,\delta}\}_{t\geq 0}$ be the corresponding family of Markov kernels associated to systems (3.1) and (3.22), respectively, as defined in (3.5) and (3.28).

Then, there exists $\hat{\alpha} > 0$ such that for each fixed $\alpha \in (0, \hat{\alpha}]$ there exists $\varepsilon > 0$ and $s \in (0, 1]$ for which the following inequality holds

$$\sup_{t\geq 0} \mathscr{W}_{\varepsilon,s,\alpha}\left(\mu \mathscr{P}_t^{N,\delta}, \mu P_t\right) \leq C \max\left\{g(N,\delta)^{qC'}, g(N,\delta), g(N,\delta)^{1-q}\right\}, \tag{3.183}$$

with g as in (3.177), for every $\tilde{p} \in (0, 1/2)$, $q \in (0, 1)$ and $\mu \in \Pr(\dot{L}^2)$ satisfying

$$\int_{\dot{L}^2} \left[\exp\left(c\alpha|\xi_0|^2\right) + \exp\left(\widetilde{C}\alpha'|\xi_0|^2\right)\left(1 + |\nabla\xi_0| + |A\xi_0|^{1/2}\right)^s\right] \mu(\mathrm{d}\xi_0) < \infty \tag{3.184}$$

for some absolute constant c, with $\widetilde{C}$ and α' being the same as in (3.163) and (3.178), respectively. Moreover, $C = C(\varepsilon, s, \alpha, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \|\sigma\|_{\dot{H}^2}, \tilde{p}, \mu)$ and $C' = C'(\varepsilon, s, \alpha, \nu, |\sigma|)$.

Consequently, under these same assumptions it follows that for all $\xi_0 \in \dot{H}^2$ and $\rho_{\varepsilon,s,\alpha}$-Lipschitz function $\varphi : \dot{L}^2 \to \mathbb{R}$ with Lipschitz constant L_φ,

$$\sup_{n\in\mathbb{N}} \left|\mathbb{E}\left[\varphi(\xi(n\delta;\xi_0)) - \varphi\left(\xi_{N,\delta}^n(\xi_0)\right)\right]\right| \leq L_\varphi C \max\left\{g(N,\delta)^{qC'}, g(N,\delta), g(N,\delta)^{1-q}\right\}, \tag{3.185}$$

with C and C' as in (3.183). Here, $\xi(t;\xi_0)$, $t \geq 0$, and $\xi_{N,\delta}^n(\xi_0)$, $n \in \mathbb{N}$, denote the unique solutions of (3.1) and (3.22), respectively, such that $\xi(0;\xi_0) = \xi_0$ and $\xi_{N,\delta}^0(\xi_0) = \xi_0$.

REMARK 3.23 Regarding the Lipschitz condition on the test function φ in (3.185) above, we note that, by following similar arguments as in Glatt-Holtz & Mondaini (2022, Proposition 46), it is not difficult to show that the class of $\rho_{\varepsilon,s,\alpha}$-Lipschitz functions includes all C^1 functions $\varphi : \dot{L}^2 \to \mathbb{R}$ such that

$$L_\varphi := \sup_{\xi\in\dot{L}^2} \max\left\{\frac{2|\varphi(\xi)|}{\exp(\alpha|\xi|^2)}, \frac{\varepsilon^{1/s}\|D\varphi(\xi)\|}{\exp\left(\frac{\alpha}{2}|\xi|^2\right)}\right\} < \infty. \tag{3.186}$$

Here, $D\varphi(\xi)$ denotes the Fréchet derivative of φ at $\xi \in \dot{L}^2$, and $\|D\varphi(\xi)\|$ its corresponding standard operator norm. In this case, L_φ yields a suitable Lipschitz constant with respect to $\rho_{\varepsilon,s,\alpha}$. In particular, observe that (3.186) allows for exponentially growing φ and $D\varphi$.

Proof. Let us verify that the assumptions of Theorem 2.8 hold. The verification of assumptions (H1)–(H4) follows as in the proof of Theorem 3.21. Moreover, from the definitions of $g_{\theta_0}(N,\delta)$ and $R(t)$ in (3.179) and (3.180), respectively, it is clear that R is continuous and strictly increasing in t, and g_{θ_0} is bounded with respect to $(N,\delta) \in \mathbb{N} \times (0, \delta_0]$. Further, as argued at the end of the proof of Theorem 3.21, denoting by μ_* an invariant measure of $\{P_t\}_{t\geq 0}$ it follows from the definition of f in (3.181) and Lemma 3.20 that $\int_{\dot{L}^2} f(\xi_0)\mu_*(\mathrm{d}\xi_0) \leq C < \infty$, for $C = C(\nu, \|\sigma\|_{\dot{H}^2})$.

Thus, from Theorem 2.8 and Remark 2.10 we deduce that

$$\sup_{t\geq 0} \mathscr{W}_{\varepsilon,s,\alpha}\left(\mu \mathscr{P}_t^{N,\delta}, \mu P_t\right) \leq \tilde{g}(N,\delta)\left[\mathscr{W}_{\varepsilon,s,2\alpha}(\mu,\mu_*) + \int_{\dot{L}^2} f(\xi_0)\mu(\mathrm{d}\xi_0) + \int_{\dot{L}^2} f(\xi_0)\mu_*(\mathrm{d}\xi_0)\right], \quad (3.187)$$

for every $\mu \in \Pr(\dot{L}^2)$ satisfying (3.184), where

$$\tilde{g}(N,\delta) = C \max\left\{g(N,\delta)^{qC'}, g(N,\delta), g(N,\delta)^{1-q}\right\},$$

for any fixed $q < 1$. Here, as seen from the proof of Theorem 2.8 and the invoked results, it follows that C and C' are positive constants with $C = C(\varepsilon, s, \alpha, \nu, \delta_0, |\sigma|, |\nabla\sigma|, \tilde{p})$ and $C' = C'(\varepsilon, s, \alpha, \nu, |\sigma|)$.

Moreover, from the definition of $\mathscr{W}_{\varepsilon,s,\alpha}$, together with Lemma 3.20 and under condition (3.184), it is not difficult to see that for $\alpha > 0$ sufficiently small $\mathscr{W}_{\varepsilon,s,2\alpha}(\mu,\mu_*) \leq C < \infty$, with $C = C(\nu, |\sigma|, \mu)$. This concludes the proof of (3.183). The final inequality (3.185) is clearly a direct consequence of Corollary 2.11. □

REMARK 3.24 As mentioned in Section 1.3, a useful consequence of the Wasserstein contraction result (3.30), together with the long time bias estimate (3.176) established in Theorem 3.21, are error estimates between the stationary average $\int \varphi(\xi')\mu_*(\mathrm{d}\xi')$ and its estimator $\frac{1}{n}\sum_{k=1}^n \varphi(\xi_{N,\delta}^k(\xi_0))$ for given $n \in \mathbb{N}$, $\xi_0 \in \Pi_N \dot{L}^2$, and suitable observable $\varphi : \dot{L}^2 \to \mathbb{R}$. Here, μ_* denotes the invariant measure of the Markov semigroup P_t, $t \geq 0$, associated to the 2D SNSE (3.1) and defined in (3.4). Commonly, estimates are sought for the estimator bias

$$\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n \varphi\left(\xi_{N,\delta}^k(\xi_0)\right) - \int \varphi(\xi')\mu_*(\mathrm{d}\xi')\right) = \frac{1}{n}\sum_{k=1}^n P_k^{N,\delta}\varphi(\xi_0) - \int \varphi(\xi')\mu_*(\mathrm{d}\xi'), \quad (3.188)$$

and the mean-squared error

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{k=1}^n \varphi\left(\xi_{N,\delta}^k(\xi_0)\right) - \int \varphi(\xi')\mu_*(\mathrm{d}\xi')\right)^2\right]. \quad (3.189)$$

Let us briefly sketch some of the steps that lead to these estimates. We assume φ is a $\rho_{\varepsilon,s,\alpha}$-Lipschitz function, with ε, s, α fixed so that (3.30) holds, and denote its Lipschitz constant by L_φ.

To estimate the bias (3.188), we first decompose it as

$$\left(\frac{1}{n}\sum_{k=1}^n P_k^{N,\delta}\varphi(\xi_0) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi')\right) + \left(\int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi') - \int \varphi(\xi')\mu_*(\mathrm{d}\xi')\right). \quad (3.190)$$

From the Lipschitzianity of φ and the contraction inequality (3.30), the first term can be bounded as

$$\left|\frac{1}{n}\sum_{k=1}^{n} P_k^{N,\delta}\varphi(\xi_0) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi')\right| \leq \frac{L_\varphi}{n}\sum_{k=1}^{n} \mathscr{W}_{\varepsilon,s,\alpha}\left(P_k^{N,\delta}(\xi_0,\cdot),\mu_*^{N,\delta}\right).$$

Take $T > 0$ as in Theorem 3.9. Then fix $0 < \delta \leq \delta_0$ and consider $k_0 \in \mathbb{N}$ sufficiently large such that $k_0\delta \geq T$. From (3.30), we obtain

$$\begin{aligned}&\left|\frac{1}{n}\sum_{k=1}^{n} P_k^{N,\delta}\varphi(\xi_0) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi')\right| \\ &\quad\leq \frac{L_\varphi}{n} k_0 \sup_{1\leq k\leq k_0} \mathscr{W}_{\varepsilon,s,\alpha}\left(P_k^{N,\delta}(\xi_0,\cdot),\mu_*^{N,\delta}\right) + \frac{L_\varphi}{n}\sum_{k=k_0+1}^{n} C_1\, e^{-k\delta C_2}\mathscr{W}_{\varepsilon,s,\alpha}\left(\delta_{\xi_0},\mu_*^{N,\delta}\right). \end{aligned} \tag{3.191}$$

From (3.27) and (3.33) above, one can show that

$$\sup_{1\leq k\leq k_0} \mathscr{W}_{\varepsilon,s,\alpha}\left(P_k^{N,\delta}(\xi_0,\cdot),\mu_*^{N,\delta}\right) < \infty \quad \text{and also} \quad \mathscr{W}_{\varepsilon,s,\alpha}\left(\delta_{\xi_0},\mu_*^{N,\delta}\right) < \infty,$$

both with bounds independent of $0 < \delta \leq \delta_0$ so that from (3.191) we deduce

$$\left|\frac{1}{n}\sum_{k=1}^{n} P_k^{N,\delta}\varphi(\xi_0) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi')\right| = O\left(\frac{1}{n\delta}\right) \quad \text{as } n\to\infty. \tag{3.192}$$

Estimating the second term in (3.190) as $L_\varphi \mathscr{W}_{\varepsilon,s,\alpha}(\mu_*^{N,\delta},\mu_*)$ and invoking the bias estimate (3.176) above thus yields (1.19).

Regarding the mean-squared error (3.189), we may proceed similarly as in e.g. Glatt-Holtz & Mondaini (2022, Appendix) (see also references therein) and write

$$\frac{1}{n}\sum_{k=1}^{n}\varphi\left(\xi_{N,\delta}^k(\xi_0)\right) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi') = \frac{1}{n}\sum_{k=1}^{\infty}\left(P_k^{N,\delta}\bar{\varphi}(\xi_0) - P_k^{N,\delta}\bar{\varphi}\left(\xi_{N,\delta}^n(\xi_0)\right)\right) + \frac{M_n^\varphi}{n} =: T_1^{(n)} + T_2^{(n)},$$

where $\bar{\varphi}(\xi_0) := \varphi(\xi_0) - \int \varphi(\xi')\mu_*^{N,\delta}(\mathrm{d}\xi')$, and

$$\begin{aligned} M_n^\varphi :&= \sum_{k=1}^{\infty}\left[\mathbb{E}\left(\bar{\varphi}\left(\xi_{N,\delta}^k(\xi_0)\right)|\mathscr{F}_{n\delta}\right) - \mathbb{E}\left(\bar{\varphi}\left(\xi_{N,\delta}^k(\xi_0)\right)\right)\right] \\ &= \sum_{k=1}^{n}\bar{\varphi}\left(\xi_{N,\delta}^k(\xi_0)\right) + \sum_{k=1}^{\infty}\left(P_{k+1}^{N,\delta}\bar{\varphi}\left(\xi_{N,\delta}^n(\xi_0)\right) - P_k^{N,\delta}\bar{\varphi}(\xi_0)\right), \end{aligned}$$

$n \in \mathbb{N}$, is a martingale (relative to the filtration given by the noise increments). In view of (3.176), it suffices to estimate $\mathbb{E}(T_1^{(n)})^2$ and $\mathbb{E}(T_2^{(n)})^2$.

For the first term, we have

$$\mathbb{E}\Big[\Big(T_1^{(n)}\Big)^2\Big] \le \frac{L_\varphi^2}{n^2}\left(\sum_{k=1}^{\infty}\mathscr{W}_{\varepsilon,s,\alpha}\Big(P_k^{N,\delta}(\xi_0,\cdot),\mu_*^{N,\delta}\Big)+\mathscr{W}_{\varepsilon,s,\alpha}\Big(P_k^{N,\delta}\Big(\xi_{N,\delta}^n(\xi_0),\cdot\Big),\mu_*^{N,\delta}\Big)\right)^2,$$

so that by proceeding analogously as in (3.191)–(3.192) we obtain $\mathbb{E}[(T_1^{(n)})^2] = O((n\delta)^{-2})$ as $n \to \infty$.

For the second term, invoking standard martingale properties it follows that

$$\mathbb{E}\left[\left(\frac{M_n^\varphi}{n}\right)^2\right] = \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\Big[\Big(M_k^\varphi - M_{k-1}^\varphi\Big)^2\Big] \le 2\left(T_{2,1}^{(n)} + T_{2,2}^{(n)}\right),$$

where

$$T_{2,1}^{(n)} := \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left[\Big(\bar\varphi\Big(\xi_{N,\delta}^k(\xi_0)\Big)\Big)^2\right],$$

$$T_{2,2}^{(n)} := \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left[\left(\sum_{l=1}^{\infty}P_l^{N,\delta}\bar\varphi\Big(\xi_{N,\delta}^k(\xi_0)\Big) - P_l^{N,\delta}\bar\varphi\Big(\xi_{N,\delta}^{k-1}(\xi_0)\Big)\right)^2\right].$$

Now for $T_{2,1}^{(n)}$ we fix $\bar\xi \in \dot L^2$ and estimate

$$\begin{aligned}\mathbb{E}\left[\Big(\bar\varphi\Big(\xi_{N,\delta}^k(\xi_0)\Big)\Big)^2\right] &\le 2\left\{\mathbb{E}\left[\Big(\bar\varphi\Big(\xi_{N,\delta}^k(\xi_0)\Big) - \bar\varphi(\bar\xi)\Big)^2\right] + \bar\varphi(\bar\xi)^2\right\}\\ &\le 2\left\{L_\varphi\mathbb{E}\Big[\rho_{\varepsilon,s,\alpha}\Big(\xi_{N,\delta}^k(\xi_0),\bar\xi\Big)^2\Big] + \bar\varphi(\bar\xi)^2\right\}.\end{aligned}$$

Again from (3.27) and (3.33), we obtain $\sup_k \mathbb{E}\left[\left(\bar\varphi(\xi_{N,\delta}^k(\xi_0))\right)^2\right] < \infty$, which yields $T_{2,1}^{(n)} \le C/n$ for some constant C.

Lastly, we bound $T_{2,2}^{(n)}$ as

$$T_{2,2}^{(n)} \le \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left[\left(\sum_{l=1}^{\infty}L_\varphi\mathscr{W}_{\varepsilon,s,\alpha}\Big(P_l^{N,\delta}\Big(\xi_{N,\delta}^k(\xi_0),\cdot\Big),P_l^{N,\delta}\Big(\xi_{N,\delta}^{k-1}(\xi_0),\cdot\Big)\Big)\right)^2\right].$$

Taking k_0 as in (3.191), we obtain after further estimates that

$$T_{2,2}^{(n)} \le \frac{C}{n} + \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{E}\left[\left(\sum_{l=1}^{\infty}C_1\,e^{-l\delta C_2}\rho_{\varepsilon,s,\alpha}\Big(\xi_{N,\delta}^k(\xi_0),\xi_{N,\delta}^{k-1}(\xi_0)\Big)\right)^2\right].$$

By estimating the difference $|\xi^k_{N,\delta} - \xi^{k-1}_{N,\delta}|$ according to (3.20) and choosing s appropriately, one can show that $\rho_{\varepsilon,s,\alpha}(\xi^k_{N,\delta}(\xi_0), \xi^{k-1}_{N,\delta}(\xi_0)) \lesssim \delta^{1/2}$, which ultimately yields $T^{(n)}_{2,2} = O((n\delta)^{-1})$ as $n \to \infty$. Such considerations together with the bias estimate (3.176) thus imply (1.20).

4. Wasserstein contraction in the case of a bounded domain

In this section, we apply Theorem 2.1 to show Wasserstein contraction for the Markov kernel associated to the 2D stochastic Navier–Stokes equations on a bounded domain. As discussed in the introduction we include this domain example to illustrate the full scope and significance of Theorem 2.1. Indeed such a suitable form of contraction does not appear to follow from the approaches taken in previous relevant works in this direction (Hairer & Mattingly, 2008; Hairer *et al.*, 2011; Butkovsky *et al.*, 2020), as we describe in technical detail in Remark 4.6 below.

4.1 *Mathematical setting*

We start by briefly recalling the associated mathematical setting in Section 4.1. For further details, we refer to e.g. Constantin & Foias (1988); Foias *et al.* (2001); Temam (2001); Albeverio *et al.* (2008). Let $\mathscr{D} \subset \mathbb{R}^2$ be an open and bounded domain with smooth boundary $\partial\mathscr{D}$. Similarly as in Section 3.1.1, we fix a stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}, \{W^k\}_{k=1}^d)$ equipped with a finite family $\{W^k\}_{k=1}^d$ of standard independent real-valued Brownian motions on Ω that are adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$. We then consider the following stochastically forced 2D Navier–Stokes equations in $\mathscr{D}$

$$\mathrm{d}\mathbf{u} + (-\nu\Delta\mathbf{u} + \mathbf{u}\cdot\nabla\mathbf{u} + \nabla p)\,\mathrm{d}t = \mathbf{f}\,\mathrm{d}t + \sum_{k=1}^d \sigma_k\,\mathrm{d}W^k, \quad \nabla\cdot\mathbf{u} = 0, \tag{4.1}$$

subject to the no-slip (Dirichlet) boundary condition

$$\mathbf{u}|_{\partial\mathscr{D}} = 0, \tag{4.2}$$

where $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ and $p = p(\mathbf{x}, t)$, $(\mathbf{x}, t) \in \mathscr{D} \times [0, \infty)$, are the unknowns, and denote the velocity vector field and the scalar pressure field, respectively; whereas $\nu > 0$ and $\mathbf{f} = \mathbf{f}(\mathbf{x})$, $\mathbf{x} \in \mathscr{D}$, are given and represent the kinematic viscosity parameter and a deterministic body force, respectively. Moreover, $\{\sigma_k\}_{k=1}^d$ are given vector fields in $\mathscr{D}$. We assume that $\mathbf{f}, \sigma_1, \ldots, \sigma_d \in L^2(\mathscr{D})^2$.

Consider the following functional spaces

$$H = \Big\{\mathbf{u} \in L^2(\mathscr{D})^2 \,:\, \nabla\cdot\mathbf{u} = 0,\ \mathbf{u}\cdot\mathbf{n}|_{\partial\mathscr{D}} = 0\Big\},$$
$$V = \Big\{\mathbf{u} \in H^1(\mathscr{D})^2 \,:\, \nabla\cdot\mathbf{u} = 0,\ \mathbf{u}|_{\partial\mathscr{D}} = 0\Big\},$$

where $\mathbf{n}$ denotes the outward unit normal vector to $\partial\mathscr{D}$. See e.g. Constantin & Foias (1988); Temam (2001). We endow H with the standard inner product and associated norm from $L^2(\mathscr{D})^2$, which we denote as $(\cdot,\cdot)$ and $|\cdot|$, respectively. For V, we consider the inner product $((\mathbf{u}, \mathbf{v})) := (\nabla\mathbf{u}, \nabla\mathbf{v})$, with associated norm $\|\mathbf{u}\| := ((\mathbf{u}, \mathbf{u}))^{1/2} = |\nabla\mathbf{u}|$, which is well-defined due to Poincaré inequality (4.4) below. We identify H with its dual H', so that $V \subseteq H \equiv H' \subseteq V'$, with continuous injections, where V' denotes the dual space of V.

Denoting by P_L the Leray projection of $L^2(\mathscr{D})^2$ onto H, and applying P_L to (4.1) yields the following functional formulation

$$\mathrm{d}\mathbf{u} + (\nu A\mathbf{u} + B(\mathbf{u}, \mathbf{u}))\,\mathrm{d}t = \mathbf{f}\,\mathrm{d}t + \sigma\,\mathrm{d}W, \tag{4.3}$$

where we assume without loss of generality that $P_L\mathbf{f} = \mathbf{f}$ and $P_L\sigma_k = \sigma_k$,[3] and use the abbreviated notation $\sigma\,\mathrm{d}W := \sum_{k=1}^{d} \sigma_k\,\mathrm{d}W^k$. Here, $A : V \cap H^2(\mathscr{D})^2 \to H$, $A\mathbf{u} = -P_L\Delta\mathbf{u}$, is the Stokes operator and $B : V \times V \to V'$ is the bilinear mapping $B(\mathbf{u}, \mathbf{v}) = P_L(\mathbf{u}\cdot\nabla)\mathbf{v}$. Similarly as in the periodic case, we have that A is a positive and self-adjoint operator with compact inverse. Therefore, it admits a nondecreasing sequence of positive eigenvalues $\{\lambda_k\}_{k\in\mathbb{N}}$ with $\lambda_k \to \infty$ as $k \to \infty$, which is associated to a sequence of eigenfunctions $\{e_k\}_{k\in\mathbb{N}}$ forming an orthonormal basis of H. For each $K \in \mathbb{N}$, we denote by $\Pi_K : H \to H$ the projection operator onto the subspace $\Pi_K H$ of H consisting of the span of the first K eigenfunctions of A.

We recall Poincaré inequality

$$|\mathbf{u}| \le \lambda_1^{-1/2}\|\mathbf{u}\| \quad \text{for all } \mathbf{u} \in V, \tag{4.4}$$

where λ_1 denotes the smallest eigenvalue of the Stokes operator A. Moreover, for each $K \in \mathbb{N}$ we have

$$|(I - \Pi_K)\mathbf{u}| \le \lambda_{K+1}^{-1/2}\|(I - \Pi_K)\mathbf{u}\| \quad \text{for all } \mathbf{u} \in V. \tag{4.5}$$

Recall also the following property of the bilinear term

$$(B(\mathbf{u}, \mathbf{v}), \mathbf{w}) = -(B(\mathbf{u}, \mathbf{w}), \mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v}, \mathbf{w} \in V,$$

which implies

$$(B(\mathbf{u}, \mathbf{v}), \mathbf{v}) = 0 \quad \text{for all } \mathbf{u}, \mathbf{v} \in V. \tag{4.6}$$

We adopt similar notation from Section 3.1.1 regarding the noise term $\sigma\,\mathrm{d}W$. Specifically, let $\mathbf{H}$ denote the d-fold product of H and define, for each $\sigma = (\sigma_1, \ldots, \sigma_d) \in \mathbf{H}$, $|\sigma|^2 := \sum_{k=1}^{d} |\sigma_k|^2$. We also abuse notation and see any $\sigma \in \mathbf{H}$ as a mapping $\sigma : \mathbb{R}^d \to H$, with $\sigma(w_1, \ldots, w_d) = \sum_{k=1}^{d} \sigma_k w_k$, and denote by $\sigma^{-1} : range(\sigma) \to \mathbb{R}^d$ its corresponding pseudo-inverse. Clearly, both σ and σ^{-1} are bounded operators.

The existence and uniqueness of pathwise solutions of (4.1)–(4.2) satisfying a given initial condition follows analogously as in Proposition 3.1, with appropriate modifications in the functional spaces. Namely, it holds by replacing $\dot{L}^2$ and $\dot{H}^1$ with H and V, respectively. We thus define the associated transition function for all $\mathbf{u}_0 \in H$ and Borel set $\mathscr{O} \in \mathscr{B}(H)$ as

$$P_t(\mathbf{u}_0, \mathscr{O}) := \mathbb{P}(\mathbf{u}(t; \mathbf{u}_0) \in \mathscr{O}),$$

[3] Indeed, we have more generally that $\mathbf{f} = \nabla q + P_L\mathbf{f}$ for some $q \in H^1(\mathscr{D})$ (see e.g. Constantin & Foias, 1988, Proposition 1.8 and Proposition 1.9), so that we may redefine the pressure term p in (4.1) as $\tilde{p} = p + q$, and similarly for σ_k. In particular, (4.3) remains the same.

where $\mathbf{u}(t;\mathbf{u}_0)$, $t \geq 0$, is the unique pathwise solution of (4.1)–(4.2) satisfying $\mathbf{u}(0;\mathbf{u}_0) = \mathbf{u}_0$ almost surely. The associated Feller Markov semigroup P_t, $t \geq 0$, is defined as

$$P_t\varphi(\mathbf{u}_0) = \mathbb{E}\varphi(\mathbf{u}(t;\mathbf{u}_0)), \quad \mathbf{u}_0 \in H, \tag{4.7}$$

for every bounded and measurable function $\varphi : H \to \mathbb{R}$.

4.2 *Wasserstein contraction estimates*

We proceed to verify that assumptions (A1)–(A3) of Theorem 2.1 hold in this setting. Specifically, following the notation from Theorem 2.1, we take $(X, \|\cdot\|) = (H, |\cdot|)$, $\mathscr{I} = \mathbb{R}^+$ and $\{P_t\}_{t\in\mathscr{I}}$ to be the Markov semigroup defined in (4.7). Moreover, we take Λ to be the class of distances

$$\Lambda = \left\{\rho_{\varepsilon,s} \,:\, \varepsilon > 0,\, 0 < s \leq \min\left\{1, c\frac{\nu^3\lambda_1}{|\sigma|^2}\right\}\right\} \tag{4.8}$$

for some positive absolute constant c, with $\rho_{\varepsilon,s}$ defined analogously as in (3.26), namely

$$\rho_{\varepsilon,s}(\mathbf{u},\tilde{\mathbf{u}}) = 1 \wedge \frac{|\mathbf{u}-\tilde{\mathbf{u}}|^s}{\varepsilon} \qquad \text{for all } \mathbf{u},\tilde{\mathbf{u}} \in H. \tag{4.9}$$

We notice that in (4.8) we impose a different assumption on s than in Section 3.2, where we simply consider $s \in (0,1]$. See Remark 4.6 below for more details.

In the next proposition, we show with (4.12) that 2.1 is satisfied under this setting. We also provide the energy-type inequality (4.11) to be used later in the verification of 2.1–2.1.

Proposition 4.1 Fix any $\sigma \in \mathbf{H}$ and $\mathbf{u}_0 \in H$. Let $\mathbf{u}(t)$, $t \geq 0$, be the solution of (4.3) satisfying $\mathbf{u}(0) = \mathbf{u}_0$ almost surely. Then, for all $\alpha \in \mathbb{R}$ satisfying

$$0 < \alpha \leq \frac{\nu\lambda_1}{4|\sigma|^2} \tag{4.10}$$

the following inequalities hold:

$$\mathbb{E}\sup_{t\geq 0}\exp\left(\alpha|\mathbf{u}(t)|^2 + \alpha\nu\int_0^t \|\mathbf{u}(s)\|^2\,\mathrm{d}s - \alpha t\left(|\sigma|^2 + \frac{2}{\nu\lambda_1}|\mathbf{f}|^2\right)\right) \leq 2\exp(\alpha|\mathbf{u}_0|^2), \tag{4.11}$$

and

$$P_t\exp(\alpha|\mathbf{u}_0|^2) = \mathbb{E}\exp(\alpha|\mathbf{u}(t)|^2) \leq C\exp\left(\alpha\, e^{-\nu\lambda_1 t}|\mathbf{u}_0|^2\right) \qquad \text{for all } t \geq 0, \tag{4.12}$$

where $C = C(\nu, \lambda_1, |\mathbf{f}|, |\sigma|)$.

Proof. The proof of (4.11) follows by applying Itô's formula to the mapping $\mathbf{u} \mapsto |\mathbf{u}|^2$ and invoking standard exponential martingale arguments. We refer to Hairer & Mattingly (2008); Kuksin & Shirikyan (2012); Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020) for further details.

The proof of (4.12) is essentially an analogous continuous version of Proposition 3.11. Indeed, fixing $T > 0$, applying Itô's formula to the mapping $(\tau, \mathbf{u}) \mapsto e^{-\nu\lambda_1(T-\tau)}|\mathbf{u}(\tau)|^2$ and invoking (4.6), it follows that for all $t \in [0, T]$

$$\begin{aligned} &e^{-\nu\lambda_1(T-t)}|\mathbf{u}(t)|^2 + 2\nu \int_0^t e^{-\nu\lambda_1(T-\tau)}\|\mathbf{u}(\tau)\|\,\mathrm{d}\tau \\ &\quad = e^{-\nu\lambda_1 T}|\mathbf{u}_0|^2 + \nu\lambda_1 \int_0^t e^{-\nu\lambda_1(T-\tau)}|\mathbf{u}(\tau)|\,\mathrm{d}\tau + \int_0^t e^{-\nu\lambda_1(T-\tau)}\left[2(\mathbf{u}(\tau), \mathbf{f}) + |\sigma|^2\right]\,\mathrm{d}\tau + M_t, \end{aligned} \tag{4.13}$$

where $M_t := 2\int_0^t e^{-\nu\lambda_1(T-\tau)}(\mathbf{u}, \sigma)\,\mathrm{d}W(\tau)$, $t \in [0, T]$, is a martingale. We estimate its quadratic variation $\langle M\rangle_t$ as

$$\langle M\rangle_t = 4\int_0^t e^{-2\nu\lambda_1(T-\tau)}(\mathbf{u}, \sigma)\,\mathrm{d}\tau \leq 4\frac{|\sigma|^2}{\lambda_1}\int_0^t e^{-\nu\lambda_1(T-\tau)}\|\mathbf{u}(\tau)\|^2\,\mathrm{d}\tau, \tag{4.14}$$

where we applied Cauchy–Schwarz and Poincaré inequality (4.4). Moreover, again from (4.4) and Young's inequality, we have

$$(\mathbf{u}(\tau), \mathbf{f}) \leq \frac{1}{\lambda_1^{1/2}}\|\mathbf{u}(\tau)\|\,|\mathbf{f}| \leq \frac{\nu}{4}\|\mathbf{u}(\tau)\|^2 + \frac{1}{\nu\lambda_1}|\mathbf{f}|^2. \tag{4.15}$$

Now we add and subtract $\alpha\langle M\rangle_t/2$ in (4.13), for α satisfying (4.10), and invoke (4.4) once again to estimate the second term in the right-hand side of (4.13). Plugging the estimates (4.14)–(4.15), it follows after rearranging terms that

$$e^{-\nu\lambda_1(T-t)}|\mathbf{u}(t)|^2 \leq e^{-\nu\lambda_1 T}|\mathbf{u}_0|^2 + \left(\frac{2}{\nu\lambda_1}|\mathbf{f}|^2 + |\sigma|^2\right)\frac{e^{-\nu\lambda_1 T}(e^{\nu\lambda_1 t} - 1)}{\nu\lambda_1} + M_t - \frac{\alpha}{2}\langle M\rangle_t.$$

Multiplying by α and taking expected values on both sides,

$$\begin{aligned} &\mathbb{E}\exp\left(\alpha\, e^{-\nu\lambda_1(T-t)}|\mathbf{u}(t)|^2\right) \\ &\quad \leq \exp\left(\alpha\, e^{-\nu\lambda_1 T}|\mathbf{u}_0|^2\right)\exp\left(\alpha\left(\frac{2}{\nu\lambda_1}|\mathbf{f}|^2 + |\sigma|^2\right)\frac{e^{-\nu\lambda_1 T}(e^{\nu\lambda_1 t} - 1)}{\nu\lambda_1}\right), \end{aligned} \tag{4.16}$$

where we used that $\{N_t\}_{t\geq 0} = \{\exp(\alpha M_t - (\alpha^2/2)\langle M\rangle_t)\}_{t\geq 0}$ is a supermartingale (see e.g. Kuksin & Shirikyan, 2012, Appendix A.11), and hence $\mathbb{E}N_t \leq \mathbb{E}N_0 = 1$ for all $t \geq 0$. Taking in particular $T = t$ in (4.16), we deduce (4.12) with $C = \exp\left(\frac{\alpha}{\nu\lambda_1}\left(\frac{2}{\nu\lambda_1}|\mathbf{f}|^2 + |\sigma|^2\right)\right)$. □

To verify the remaining assumptions (A2)–(A3) of Theorem 2.1, we proceed similarly as in Section 3.2 and consider the following modified system

$$\mathrm{d}\tilde{\mathbf{u}} + \left[\nu A\tilde{\mathbf{u}} + B(\tilde{\mathbf{u}}, \tilde{\mathbf{u}}) + \frac{\nu\lambda_{K+1}}{2}\Pi_K\left(\tilde{\mathbf{u}} - \mathbf{u}(\mathbf{u}_0)\right)\right]\mathrm{d}t = \mathbf{f}\,\mathrm{d}t + \sigma\,\mathrm{d}W \tag{4.17}$$

for each fixed $\mathbf{u}_0 \in H$ and corresponding pathwise solution $\mathbf{u}(t;\mathbf{u}_0)$, $t \geq 0$, of (4.3) satisfying $\mathbf{u}(0;\mathbf{u}_0) = \mathbf{u}_0$ almost surely. Here, $K \in \mathbb{N}$ is a parameter to be appropriately chosen in (4.23) below.

With similar arguments as in Proposition 3.1, we can show (4.17) to be well-posed in the pathwise sense. This allows us to define, for any fixed $\mathbf{u}_0 \in H$, the mapping

$$\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0, \mathscr{O}) = \mathbb{P}(\tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0, \mathbf{u}_0) \in \mathscr{O}) \quad \text{for all } t \geq 0,\ \tilde{\mathbf{u}}_0 \in H \text{ and } \mathscr{O} \in \mathscr{B}(H), \tag{4.18}$$

where $\tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0, \mathbf{u}_0)$, $t \geq 0$, is the unique pathwise solution of (4.17) satisfying $\tilde{\mathbf{u}}(0;\tilde{\mathbf{u}}_0, \mathbf{u}_0) = \tilde{\mathbf{u}}_0$ almost surely.

The next proposition presents a pathwise contraction estimate for the difference between a solution $\tilde{\mathbf{u}}(\tilde{\mathbf{u}}_0, \mathbf{u}_0)$ of (4.17) and the corresponding solution $\mathbf{u}(\mathbf{u}_0)$ of (4.3). The proof is given in Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020), but we present the main ideas here for completeness.

PROPOSITION 4.2 Fix any $\sigma \in \mathbf{H}$, $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ and $K \in \mathbb{N}$. Let $\tilde{\mathbf{u}}(t) = \tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0, \mathbf{u}_0)$, $t \geq 0$, be the solution of (4.17) corresponding to this data and satisfying $\tilde{\mathbf{u}}(0) = \tilde{\mathbf{u}}_0$ almost surely. Then the following inequality holds almost surely for all $t \geq 0$

$$|\tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0;\mathbf{u}_0) - \mathbf{u}(t;\mathbf{u}_0)|^2 \leq |\tilde{\mathbf{u}}_0 - \mathbf{u}_0|^2 \exp\left(-\nu\lambda_{K+1}t + \frac{c}{\nu}\int_0^t \|\mathbf{u}(\tau;\mathbf{u}_0)\|^2\,\mathrm{d}\tau\right). \tag{4.19}$$

Proof. Denote $\mathbf{v}(t) = \tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0, \mathbf{u}_0) - \mathbf{u}(t;\mathbf{u}_0)$. Subtracting (4.3) from (4.17), it follows that

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \nu A\mathbf{v} + B(\mathbf{v}, \mathbf{v}) + B(\mathbf{v}, \mathbf{u}) + B(\mathbf{u}, \mathbf{v}) + \frac{\nu\lambda_{K+1}}{2}\Pi_K\mathbf{v} = 0.$$

Taking the inner product in H with $\mathbf{v}$ and invoking (4.6),

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{v}|^2 + \nu\|\mathbf{v}\|^2 + \frac{\nu\lambda_{K+1}}{2}|\Pi_K\mathbf{v}|^2 = -(B(\mathbf{v}, \mathbf{u}), \mathbf{v}).$$

By classical Hölder, interpolation and Young's inequalities, we estimate the nonlinear term as

$$|(B(\mathbf{v}, \mathbf{u}), \mathbf{v})| \leq |\mathbf{v}|\,\|\mathbf{v}\|\,\|\mathbf{u}\| \leq \frac{\nu}{2}\|\mathbf{v}\|^2 + \frac{c}{\nu}\|\mathbf{u}\|^2|\mathbf{v}|^2, \tag{4.20}$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{v}|^2 + \nu\|\mathbf{v}\|^2 + \nu\lambda_{K+1}|\Pi_K\mathbf{v}|^2 \leq \frac{c}{\nu}\|\mathbf{u}\|^2|\mathbf{v}|^2. \tag{4.21}$$

Moreover, from (4.5),

$$\nu\|\mathbf{v}\|^2 = \nu\|\Pi_K \mathbf{v}\|^2 + \nu\|(I-\Pi_K)\mathbf{v}\|^2 \geq \nu\|\Pi_K \mathbf{v}\|^2 + \nu\lambda_{K+1}|(I-\Pi_K)\mathbf{v}|^2.$$

Plugging back into (4.21), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{v}|^2 + \left(\nu\lambda_{K+1} - \frac{c}{\nu}\|\mathbf{u}\|^2\right)|\mathbf{v}|^2 \leq 0,$$

from which (4.19) follows after integrating on $[0,t]$. □

In the next two propositions, we establish the validity of assumptions (A2) and (A3) from Theorem 2.1, with the help of the pathwise contraction (4.19). In view of the crucial differences between (4.19) and the analogous pathwise contraction estimate (3.86) from Section 3 that are pointed out in detail in Remark 4.6 below, to verify the smallness property from (A2) we must resort here to the following estimate for the total variation distance between the laws of a Wiener process W in $\mathbb{R}^d$ and the corresponding shifted process $\widehat{W}$ as in (3.92):

$$\left\|\mathscr{L}(\widehat{W}) - \mathscr{L}(W)\right\|_{\mathrm{TV}} \leq 1 - \frac{1}{6}\min\left\{\frac{1}{8}, \exp\left[-\left(2^{2-a}\,\mathbb{E}\left[\left(\int_0^\infty |\varphi(t)|\,\mathrm{d}t\right)^a\right]\right)^{\frac{1}{a}}\right]\right\} \tag{4.22}$$

for any $a \in (0,1]$,[4] see Butkovsky *et al.* (2020, Theorem A.5, (A.14)).

Proposition 4.3 Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \mathbf{H}$ such that

$$\Pi_K H \subset \mathit{range}(\sigma) \quad \text{and} \quad \lambda_{K+1} \geq \frac{c}{\nu^3}\left(|\sigma|^2 + \frac{2}{\nu\lambda_1}|\mathbf{f}|^2\right) \tag{4.23}$$

for some absolute constant $c > 0$. Then, for every $M > 0$, and ε, s as in (4.8), there exist a time $T_1 = T_1(M,\varepsilon,s) > 0$ and a coefficient $\kappa_1 = \kappa_1(M) \in (0,1)$, which is independent of ε, s, for which the following inequality holds

$$\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot), P_t(\tilde{\mathbf{u}}_0,\cdot)) \leq 1 - \kappa_1$$

for all $t \geq T_1$ and for every $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ with $|\mathbf{u}_0| \leq M$ and $|\tilde{\mathbf{u}}_0| \leq M$.

Proof. Fix $M > 0$, and ε, s as in (4.8). Let $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ such that $|\mathbf{u}_0| \leq M$ and $|\tilde{\mathbf{u}}_0| \leq M$. Recalling the definition of $\mathscr{W}_{\varepsilon,s}$ in (2.2), (3.26) and of $\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot)$ in (4.18), we obtain by invoking Proposition A.2 that

$$\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot), P_t(\tilde{\mathbf{u}}_0,\cdot)) \leq \mathscr{W}_{\varepsilon,s}\left(P_t(\mathbf{u}_0,\cdot), \widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot)\right) + \mathscr{W}_{\varepsilon,s}\left(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot), P_t(\tilde{\mathbf{u}}_0,\cdot)\right). \tag{4.24}$$

[4] Similarly as for (3.94), here we notice that in Butkovsky *et al.* (2020, Theorem A.5, (A.14)) it is assumed instead $a \in (0,1)$. In fact, (3.95) and (3.93) above imply that (4.22) also holds with $a = 1$, although with an even sharper bound.

For the first term in the right-hand side of (4.24), we have

$$\mathscr{W}_{\varepsilon,s}\Big(P_t(\mathbf{u}_0,\cdot),\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot)\Big) \leq \frac{1}{\varepsilon}\mathbb{E}\Big[\Big|\mathbf{u}(t;\mathbf{u}_0)-\tilde{\mathbf{u}}(t;\tilde{\mathbf{u}}_0,\mathbf{u}_0)\Big|^s\Big], \tag{4.25}$$

so that by invoking the pathwise estimate (4.19) it follows that

$$\mathscr{W}_{\varepsilon,s}\Big(P_t(\mathbf{u}_0,\cdot),\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot)\Big) \leq \frac{1}{\varepsilon}|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^s \exp\left(-\frac{s}{2}\nu\lambda_{K+1}t\right)\mathbb{E}\exp\left(\frac{cs}{\nu}\int_0^t \|\mathbf{u}(\tau)\|^2\,\mathrm{d}\tau\right). \tag{4.26}$$

Let $\alpha = cs/\nu^2$. Since $0 < s \leq c\nu^3\lambda_1/|\sigma|^2$ then α satisfies (4.10). We may thus invoke (4.11) to further estimate (4.26) as

$$\begin{aligned}
&\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot))\\
&\qquad\leq \frac{c}{\varepsilon}|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^s\exp\left(-\frac{s}{2}\nu\lambda_{K+1}t\right)\exp\left(\frac{cs}{\nu^2}|\mathbf{u}_0|^2\right)\exp\left(\frac{cst}{\nu^2}\left(|\sigma|^2+\frac{2}{\nu\lambda_1}|\mathbf{f}|^2\right)\right)\\
&\qquad\leq \frac{c}{\varepsilon}|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^s\exp\left(-\frac{s}{4}\nu\lambda_{K+1}t\right)\exp\left(\frac{cs}{\nu^2}|\mathbf{u}_0|^2\right)
\end{aligned} \tag{4.27}$$

$$\leq \frac{c}{\varepsilon}(2M)^s\exp\left(\frac{cs}{\nu^2}M^2\right)\exp\left(-\frac{s}{4}\nu\lambda_{K+1}t\right), \tag{4.28}$$

where the second inequality follows from assumption (4.23) on K.

Regarding the second term in the right-hand side of (4.24), we proceed analogously as in (3.100)–(3.102) and obtain

$$\mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)\Big) \leq \big\|\mathscr{L}(\widehat{W})-\mathscr{L}(W)\big\|_{\mathrm{TV}}, \tag{4.29}$$

where

$$\widehat{W}(t) := W(t) - \int_0^t \frac{\nu\lambda_{K+1}}{2}\sigma^{-1}\Pi_K(\tilde{\mathbf{u}}-\mathbf{u})(\tau)\,\mathrm{d}\tau,\quad t\geq 0.$$

From (4.22), it follows that for any $a\in(0,1]$

$$\begin{aligned}
&\mathscr{W}_{\varepsilon,s}\Big(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)\Big)\\
&\quad\leq 1-\frac{1}{6}\min\left\{\frac{1}{8},\exp\left[-\left(2^{2-a}\,\mathbb{E}\left[\left(\int_0^\infty\left|\frac{\nu\lambda_{K+1}}{2}\sigma^{-1}\Pi_K(\tilde{\mathbf{u}}-\mathbf{u})(t)\right|^2\mathrm{d}t\right)^a\right]\right)^{\frac{1}{a}}\right]\right\}.
\end{aligned} \tag{4.30}$$

Invoking (4.19) once again, we deduce

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_0^\infty \left|\frac{\nu\lambda_{K+1}}{2}\sigma^{-1}\Pi_K(\tilde{\mathbf{u}}-\mathbf{u})(t)\right|^2 \mathrm{d}t\right)^a\right] &\le \left(\frac{\nu\lambda_{K+1}}{2}\|\sigma^{-1}\|\right)^{2a} \mathbb{E}\left[\left(\int_0^\infty |\tilde{\mathbf{u}}(t)-\mathbf{u}(t)|^2 \,\mathrm{d}t\right)^a\right] \\
&\le \left(\frac{\nu\lambda_{K+1}}{2}\|\sigma^{-1}\|\right)^{2a} \mathbb{E}\left[\left(\int_0^\infty |\tilde{\mathbf{u}}_0-\mathbf{u}_0|^2 \exp\left(-\nu\lambda_{K+1}t+\frac{c}{\nu}\int_0^t \|\mathbf{u}(\tau)\|^2 \,\mathrm{d}\tau\right) \mathrm{d}t\right)^a\right] \\
&\le \left(\frac{\nu\lambda_{K+1}}{2}\|\sigma^{-1}\|\right)^{2a} |\tilde{\mathbf{u}}_0-\mathbf{u}_0|^{2a} \\
&\qquad \cdot \mathbb{E}\left[\left(\int_0^\infty \exp\left(-\nu\lambda_{K+1}t+\frac{c}{\nu^2}|\sigma|^2 t\right) \mathrm{d}t\right)^a \sup_{t\ge 0}\exp\left(\frac{ca}{\nu}\int_0^t \|\mathbf{u}(\tau)\|^2\,\mathrm{d}\tau - \frac{ca}{\nu^2}|\sigma|^2 t\right)\right] \\
&\le \left(\frac{\nu\lambda_{K+1}}{2}\|\sigma^{-1}\|\right)^{2a} |\tilde{\mathbf{u}}_0-\mathbf{u}_0|^{2a} \left(\int_0^\infty e^{-\frac{\nu\lambda_{K+1}t}{2}}\,\mathrm{d}t\right)^a \mathbb{E}\left[\sup_{t\ge 0}\exp\left(\frac{ca}{\nu}\int_0^t \|\mathbf{u}(\tau)\|^2\,\mathrm{d}\tau - \frac{ca}{\nu^2}|\sigma|^2 t\right)\right],
\end{aligned} \tag{4.31}
$$

where the last inequality follows from assumption (4.23) on K. Now assuming $a \le c\nu^3\lambda_1/|\sigma|^2$ so that we can resort to (4.11), we further obtain

$$
\begin{aligned}
\mathbb{E}\left[\left(\int_0^\infty \left|\frac{\nu\lambda_{K+1}}{2}\sigma^{-1}\Pi_K(\tilde{\mathbf{u}}-\mathbf{u})(t)\right|^2 \mathrm{d}t\right)^a\right] &\le c(\nu\lambda_{K+1})^a\|\sigma^{-1}\|^{2a}|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^{2a}\exp\left(\frac{ca}{\nu^2}|\mathbf{u}_0|^2\right) \\
&\le c(\nu\lambda_{K+1})^a\|\sigma^{-1}\|^{2a}M^{2a}\exp\left(\frac{ca}{\nu^2}M^2\right).
\end{aligned} \tag{4.32}
$$

Plugging back into (4.30),

$$
\mathscr{W}_{\varepsilon,s}(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)) \le 1-\frac{1}{6}\min\left\{\frac{1}{8},\exp\left[-c\nu\lambda_{K+1}\|\sigma^{-1}\|^2M^2\exp\left(c\frac{M^2}{\nu^2}\right)\right]\right\}. \tag{4.33}
$$

Thus, from (4.24), (4.28) and (4.33),

$$
\begin{aligned}
\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)) &\le \frac{c}{\varepsilon}(2M)^s\exp\left(\frac{cs}{\nu^2}M^2\right)\exp\left(-\frac{s}{4}\nu\lambda_{K+1}t\right) \\
&\quad + 1-\frac{1}{6}\min\left\{\frac{1}{8},\exp\left[-c\nu\lambda_{K+1}\|\sigma^{-1}\|^2M^2\exp\left(c\frac{M^2}{\nu^2}\right)\right]\right\}.
\end{aligned}
$$

Therefore, we may choose $T_1 = T_1(\nu, K, \|\sigma^{-1}\|, M, \varepsilon, s) > 0$ sufficiently large such that for all $t \ge T_1$

$$
\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)) \le 1-\frac{1}{12}\min\left\{\frac{1}{8},\exp\left[-c\nu\lambda_{K+1}\|\sigma^{-1}\|^2M^2\exp\left(c\frac{M^2}{\nu^2}\right)\right]\right\}.
$$

This concludes the proof. □

PROPOSITION 4.4 Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \mathbf{H}$ such that (4.23) holds. Then, for every $\kappa_2 \in (0,1)$ and for every $r > 0$ there exists $s > 0$ for which the following holds:

(i) For every $\varepsilon > 0$, there exists a constant $C = C(\varepsilon, s) > 0$ such that

$$\sup_{t\geq 0} \mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)) \leq C(\varepsilon,s)\exp\left(r|\mathbf{u}_0|^2 + r|\tilde{\mathbf{u}}_0|^2\right)\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) \tag{4.34}$$

for every $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ with $\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) < 1$.

(ii) There exist a parameter $\varepsilon > 0$ and a time $T_2 > 0$ such that

$$\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)) \leq \kappa_2\exp\left(r|\mathbf{u}_0|^2 + r|\tilde{\mathbf{u}}_0|^2\right)\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) \tag{4.35}$$

for all $t \geq T_2$ and for every $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ with $\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) < 1$.

Proof. Fix any $\kappa_2 \in (0,1)$ and $r > 0$. Let $\mathbf{u}_0, \tilde{\mathbf{u}}_0 \in H$ such that $\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) < 1$. Now choose s such that

$$0 < s \leq c\min\left\{r\nu^2, \frac{\nu^3\lambda_1}{|\sigma|^2 + \nu^3\lambda_1}\right\}, \tag{4.36}$$

for some absolute constant c. We estimate the Wasserstein distance $\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot))$ as in (4.24), and then the first term on the right-hand side as in (4.27). Since $\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0) < 1$, then $|\tilde{\mathbf{u}}_0 - \mathbf{u}_0|^s\varepsilon^{-1} = \rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0)$. Thus, from (4.27),

$$\mathscr{W}_{\varepsilon,s}\left(P_t(\mathbf{u}_0,\cdot),\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot)\right) \leq c\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0)\exp\left(-\frac{s}{4}\nu\lambda_{K+1}t\right)\exp\left(\frac{cs}{\nu^2}|\mathbf{u}_0|^2\right). \tag{4.37}$$

For the second term, we first estimate as in (4.29), and then invoke (3.94) to obtain for any $a \in (0,1]$

$$\mathscr{W}_{\varepsilon,s}\left(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)\right) \leq 2^{\frac{1-a}{1+a}}\left\{\mathbb{E}\left[\left(\int_0^\infty\left|\frac{\nu\lambda_{K+1}}{2}\sigma^{-1}\varPi_K(\tilde{\mathbf{u}}-\mathbf{u})(t)\right|^2 \mathrm{d}t\right)^a\right]\right\}^{\frac{1}{1+a}}. \tag{4.38}$$

Choose $a \in (0,1]$ such that $\frac{2a}{1+a} = s$. From the choice of s in (4.36), it follows in particular that $a \leq c\nu^3\lambda_1/|\sigma|^2$. We may thus proceed as in (4.31)–(4.32) and obtain that

$$\begin{aligned}\mathscr{W}_{\varepsilon,s}\left(\widetilde{P}_{t,\mathbf{u}_0}(\tilde{\mathbf{u}}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot)\right) &\leq c(\nu\lambda_{K+1})^{\frac{a}{1+a}}\|\sigma^{-1}\|^{\frac{2a}{1+a}}|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^{\frac{2a}{1+a}}\exp\left(\frac{c}{\nu^2}\frac{a}{1+a}|\mathbf{u}_0|^2\right)\\ &\leq c(\nu\lambda_{K+1})^{s/2}\|\sigma^{-1}\|^s|\tilde{\mathbf{u}}_0-\mathbf{u}_0|^s\exp\left(r|\mathbf{u}_0|^2\right)\\ &= c\varepsilon(\nu\lambda_{K+1})^{s/2}\|\sigma^{-1}\|^s\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0)\exp\left(r|\mathbf{u}_0|^2\right),\end{aligned} \tag{4.39}$$

where in the second inequality we used that $s \leq cr\nu^2$. From (4.24), (4.37) and (4.39), we thus have

$$\begin{aligned}&\mathscr{W}_{\varepsilon,s}(P_t(\mathbf{u}_0,\cdot),P_t(\tilde{\mathbf{u}}_0,\cdot))\\ &\quad\leq c\left[\exp\left(-\frac{s}{4}\nu\lambda_{K+1}t\right) + \varepsilon(\nu\lambda_{K+1})^{s/2}\|\sigma^{-1}\|^s\right]\exp\left(r|\mathbf{u}_0|^2\right)\rho_{\varepsilon,s}(\mathbf{u}_0,\tilde{\mathbf{u}}_0)\end{aligned} \tag{4.40}$$

for all $t \geq 0$. This shows (4.34).

For (4.35), we choose $T_2 = T_2(s, K, \kappa_2) > 0$ and $\varepsilon = \varepsilon(s, K, \|\sigma^{-1}\|, \kappa_2) > 0$ such that the expression between brackets in (4.40) is less than κ_2 for all $t \geq T_2$. This concludes the proof. □

From (4.12), Proposition 4.3 and Proposition 4.4, we now obtain the following Wasserstein contraction result as an immediate consequence of Theorem 2.1.

THEOREM 4.5 Suppose there exists $K \in \mathbb{N}$ and $\sigma \in \mathbf{H}$ such that (4.23) holds. Let P_t, $t \geq 0$, be the Markov semigroup defined in (4.7). Then, for every $m > 1$ there exists $\alpha_m > 0$ such that for each $\alpha \in (0, \alpha_m]$ there exist $\varepsilon, s, T > 0$ and constants $C_1, C_2 > 0$ for which the following inequality holds

$$\mathscr{W}_{\varepsilon,s,\alpha}(\mu P_t, \tilde{\mu} P_t) \leq C_1\, e^{-tC_2} \mathscr{W}_{\varepsilon,s,\alpha/m}(\mu, \tilde{\mu}) \tag{4.41}$$

for every $\mu, \tilde{\mu} \in \Pr(H)$ and all $t \geq T$. Here we recall that, for every $a > 0$, $\mathscr{W}_{\varepsilon,s,a}$ denotes the Wasserstein-like extension to $\Pr(H)$ of the distance-like function $\rho_{\varepsilon,s,a}$ defined as in (3.27), with $\dot{L}^2$ replaced by H.

REMARK 4.6 In the proofs of Proposition 4.3 and Proposition 4.4, namely in pursuit of the conditions (A.2) and (A.3) in Theorem 2.1, we employed the pathwise estimate (4.19). It is precisely in making use of this challenging form of the 'Foias-Prodi' estimate where Theorem 2.1 improves upon the previous formulations of the weak Harris theorem in Hairer & Mattingly (2008); Hairer *et al.* (2011); Butkovsky *et al.* (2020). In effect, it does not seem possible to establish a suitable contraction bound à la (4.41) for the system (4.1), (4.2) with an obvious or direct application of these previous results.

As throughout the extant SPDE literature, (4.19) represents a crucial structural property of the dynamics which leads to the type of 'irreducibility' and 'smoothing' conditions embodied in (A.2), (A.3), respectively. However, the form that (4.19) takes here illustrates a paradigmatic challenge in regards to establishing a suitable form of contractivity in the Markovian dynamic, one that has not been fully addressed previously in the literature as far as we can tell. This is due to the coefficient in front of the integral term in (4.19), which is expected to be large in general. In turn, the size of this coefficient presents difficulties in terms of the available exponential moments on the law of the solution; namely the quadratic exponential moment bound (4.11) in (4.1) degenerates as $O(\nu^2)$ while the demands of the integral term in (4.19) increases as $O(\nu^{-1})$ for small ν. Here it is notable that a similar issue does not occur in the case of periodic boundary conditions from Section 3, thanks to better properties regarding the nonlinear term in (3.1) in this periodic case, cf. (3.82) and (4.20). On the other hand, analogously difficult (or even more difficult) forms of (4.19) appear in other models considered in e.g. Glatt-Holtz *et al.* (2017); Butkovsky *et al.* (2020); Glatt-Holtz *et al.* (2021).

This issue makes itself evident in establishing smoothing estimates for P_t as follows. In the weak Harris approach developed in Hairer *et al.* (2011) and in the subsequent literature, the appropriate smoothing condition is the ρ-contractivity condition. In contrast to (A.3) in our formulation from Theorem 2.1, the ρ-contractivity is assumed to hold uniformly over the phase space. To be specific, Hairer *et al.* (2011) requires that the bounded distance ρ that they use to eventually build their contraction distance (in a fashion closely analogous to (2.8)) maintains, for some $\alpha \in (0, 1)$,

$$\mathscr{W}_\rho(P_{t_*}(\mathbf{u}_0, \cdot), P_{t_*}(\mathbf{v}_0, \cdot)) \leq \alpha \rho(\mathbf{u}_0, \mathbf{v}_0), \qquad \text{whenever } \rho(\mathbf{u}_0, \mathbf{v}_0) < 1. \tag{4.42}$$

We observe that, even in the absence of boundaries, using a distance of the form $\rho(\mathbf{u}, \mathbf{v}) = \rho_{\varepsilon,s}(\mathbf{u}, \mathbf{v})$ given in (4.9) for appropriately tuned $s, \epsilon > 0$ only leads to a local form of this contraction estimate (4.42); cf. (3.59) above. To circumvent this difficulty, it is suggested in Hairer *et al.* (2011, Proposition 5.4) that one use a certain geodesic distance developed in Hairer & Mattingly (2008, Section 4) adapted to the ‘Lyapunov’ structure in the available form $V(\mathbf{u}) = \exp(\alpha|\mathbf{u}|^2)$. There they show that a time asymptotic smoothing estimate for the gradient estimate on the Markovian semigroup then provides the necessary global form of (4.42). This is an infinitesimal approach in the sense that we are trying to bound the distance between $P_{t_*}(\mathbf{u}_0, \cdot)$ and $P_{t_*}(\mathbf{v}_0, \cdot)$ with $\nabla P_{t_*}(\mathbf{u}_0, \cdot)(\mathbf{u}_0 - \mathbf{v}_0)$.

In our context, we could try to repeat this strategy from Hairer & Mattingly (2008, Section 4) and Hairer *et al.* (2011, Proposition 5.4) as follows. Observe that, for any C^1 observable ϕ and any $\xi \in H$, we have from (4.7)

$$\nabla P_t\phi(\mathbf{u}_0)\xi = \mathbb{E}\left(\nabla\phi(\mathbf{u}(t;\mathbf{u}_0))\mathbf{v}\right) + \mathbb{E}\left(\phi(\mathbf{u}(t;\mathbf{u}_0))\int_0^t \left(\sigma^{-1}\frac{\nu\lambda_{K+1}}{2}\Pi_K\mathbf{v}\right)\cdot\, \mathrm{d}W\right), \tag{4.43}$$

where

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \nu A\mathbf{v} + B(\mathbf{u},\mathbf{v}) + B(\mathbf{v},\mathbf{u}) + \frac{\nu\lambda_{K+1}}{2}\Pi_K\mathbf{v} = 0, \quad \mathbf{v}(0) = \xi.$$

The identity (4.43) follows by adding and subtracting the gradient of $\mathbf{u}(t;\mathbf{u}_0)$ in its noise variable, taken in the direction $\mathbf{w} = \sigma^{-1}\frac{\nu\lambda_{K+1}}{2}\Pi_K\mathbf{v}$. One then performs a Malliavin integration by parts—really just an application of the Girsanov theorem here since $\mathbf{w}$ is adapted—to obtain the second term in (4.43). One may now view (4.43) as means of estimating the possibility of coupling of two nearby solutions in two terms analogous to (4.24).

This analogy is precisely what is operationalized in the proof of Hairer *et al.* (2011, Proposition 5.4). To follow this approach, one would treat the first term as

$$\left(P_t|\nabla\phi|^2(\mathbf{u}_0)\right)^{1/2}(\mathbb{E}|\mathbf{v}|^2)^{1/2}, \tag{4.44}$$

which we compare to our bound (4.37). The second term is estimated as

$$\sup_{\mathbf{u}}|\phi(\mathbf{u})|\left(\mathbb{E}\int_0^t\left|\sigma^{-1}\frac{\nu\lambda_{K+1}}{2}\Pi_K\mathbf{v}\right|^2\,\mathrm{d}t\right)^{1/2}, \tag{4.45}$$

which we may compare to (4.38). However, in our setting neither of these terms can be shown to be finite. It is not clear how to introduce a suitable localization to the above arguments to avoid these issues with moments.

Here we notice that choosing s sufficiently small, specifically as in (4.8), is what allows us to make use of the exponential moment bound (4.11) to proceed with the estimates in (4.26), (4.39). However, these estimates in themselves are insufficient as they do not lead to the global form (4.42). On the other hand, Butkovsky *et al.* (2020) analogously employs the pseudo-metric

$$\rho(\mathbf{u},\mathbf{v}) = 1 \wedge \left(\frac{|\mathbf{u}-\mathbf{v}|^s}{\varepsilon}\exp(\alpha|\mathbf{u}|^2)\right) \wedge \left(\frac{|\mathbf{u}-\mathbf{v}|^s}{\varepsilon}\exp(\alpha|\mathbf{v}|^2)\right)$$

to achieve (4.42). While this approach from Butkovsky *et al.* (2020) would lead to a contraction in a related pseudo-metric as a direct consequence of Hairer *et al.* (2011, Theorem 4.8), it is not clear that this pseudo-metric satisfies any usable form of the generalized triangle inequality. Obviously, having such a generalized triangle inequality is indispensable for establishing continuous parameter dependence in the long time statistics of certain stochastic systems using the strategies we overviewed in Section 1.1 and used throughout Section 2.3 and Section 3.

Funding

Our efforts are supported under the grants DMS-1816551, DMS-2108790 (NEGH) and DMS-2009859, DMS-2239325 (CFM).

References

Abdulle, A., Vilmart, G. & Zygalakis, K. C. (2014) High order numerical approximation of the invariant measure of ergodic SDEs. *SIAM J. Numer. Anal.*, **52**, 1600–1622.

Albeverio, S., Flandoli, F. & Sinai, Y. G. (2008) *SPDE in Hydrodynamics: Recent Progress and Prospects: Lectures Given at the CIME Summer School Held in Cetraro, Italy, August 29–September 3, 2005*. Springer.

Ambrosio, L., Gigli, N. & Savaré, G. (2005) *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Basel, Switzerland: Birkhäuser.

Andersson, A., Kruse, R. & Larsson, S. (2016) Duality in refined Sobolev–Malliavin spaces and weak approximation of SPDE. *Stoch. Partial Differ. Equ. Anal. Comput.*, **4**, 113–149.

Andersson, A. & Larsson, S. (2016) Weak convergence for a spatial approximation of the nonlinear stochastic heat equation. *Math. Comp.*, **85**, 1335–1358.

Bally, V. & Talay, D. (1996) The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function. *Probab. Theory Related Fields*, **104**, 43–60.

Bessaih, H., Brzeźniak, Z. & Millet, A. (2014) Splitting up method for the 2D stochastic Navier–Stokes equations. *Stoch. Partial Differ. Equ. Anal. Comput.*, **2**, 433–470.

Bessaih, H. & Millet, A. (2019) Strong convergence of time numerical schemes for the stochastic two-dimensional Navier–Stokes equations. *IMA J. Numer. Anal.*, **39**, 2135–2167.

Bessaih, H. & Millet, A. (2021) Space-time Euler discretization schemes for the stochastic 2D Navier–Stokes equations. *Stoch. Partial Differ. Equ. Anal. Comput.*, **10**, 1515–1558.

Bessaih, H. & Millet, A. (2022) Strong rates of convergence of space-time discretization schemes for the 2D Navier–Stokes equations with additive noise. *Stoch. Dyn.*, **22**, 2240005.

Bréhier, C.-E. (2014) Approximation of the invariant measure with an Euler scheme for stochastic PDEs driven by space-time white noise. *Potential Anal.*, **40**, 1–40.

Bréhier, C.-E. (2022) Approximation of the invariant distribution for a class of ergodic SPDEs using an explicit tamed exponential Euler scheme. *ESAIM Math. Model. Numer. Anal.*, **56**, 151–175.

Bréhier, C.-E. & Debussche, A. (2018) Kolmogorov equations and weak order analysis for SPDEs with nonlinear diffusion coefficient. *J. Math. Pures Appl. (9)*, **119**, 193–254.

Bréhier, C.-E. & Goudenège, L. (2020) Weak convergence rates of splitting schemes for the stochastic Allen–Cahn equation. *BIT Numer. Math.*, **60**, 543–582.

Bréhier, C.-E. & Kopec, M. (2017) Approximation of the invariant law of SPDEs: error analysis using a Poisson equation for a full-discretization scheme. *IMA J. Numer. Anal.*, **37**, 1375–1410.

Breit, D. & Dodgson, A. (2021) Convergence rates for the numerical approximation of the 2D stochastic Navier–Stokes equations. *Numer. Math.*, **147**, 553–578.

Breit, D. & Prohl, A. (2022) Error analysis for 2D stochastic Navier–Stokes equations in bounded domains with Dirichlet data. arXiv preprint arXiv:2109.06495.

Bricmont, J., Kupiainen, A. & Lefevere, R. (2001) Ergodicity of the 2D Navier–Stokes equations with random forcing. *Comm. Math. Phys.*, **224**, 65–81.

Bricmont, J., Kupiainen, A. & Lefevere, R. (2002) Exponential mixing of the 2D stochastic Navier–Stokes dynamics. *Comm. Math. Phys.*, **230**, 87–132.

Brzeźniak, Z., Carelli, E. & Prohl, A. (2013) Finite-element-based discretizations of the incompressible Navier–Stokes equations with multiplicative random forcing. *IMA J. Numer. Anal.*, **33**, 771–824.

BUCKWAR, E. & SHARDLOW, T. (2005) Weak approximation of stochastic differential delay equations. *IMA J. Numer. Anal.*, **25**, 57–86.

BUTKOVSKY, O., KULIK, A. & SCHEUTZOW, M. (2020) Generalized couplings and ergodic rates for spdes and other markov models. *Ann. Appl. Probab.*, **30**, 1–39.

CAI, M., GAN, S. & WANG, X. (2021) Weak convergence rates for an explicit full-discretization of stochastic Allen–Cahn equation with additive noise. *J. Sci. Comput.*, **86**, 1–30.

CARELLI, E. & PROHL, A. (2012) Rates of convergence for discretizations of the stochastic incompressible Navier–Stokes equations. *SIAM J. Numer. Anal.*, **50**, 2467–2496.

CERRAI, S. & GLATT-HOLTZ, N. E. (2020) On the convergence of stationary solutions in the Smoluchowski–Kramers approximation of infinite dimensional systems. *J. Funct. Anal.*, **278**, 108421.

CLÉMENT, E., KOHATSU-HIGA, A. & LAMBERTON, D. (2006) A duality approach for the weak approximation of stochastic differential equations. *Ann. Appl. Probab.*, **16**, 1124–1154.

CONSTANTIN, P. & FOIAS, C. (1988) *Navier–Stokes Equations*. Chicago Lectures in Mathematics. Chicago: The University of Chicago Press.

CONUS, D., JENTZEN, A. & KURNIAWAN, R. (2019) Weak convergence rates of spectral Galerkin approximations for SPDEs with nonlinear diffusion coefficients. *Ann. Appl. Probab.*, **29**, 653–716.

CUI, J. & HONG, J. (2018) Analysis of a splitting scheme for damped stochastic nonlinear Schrödinger equation with multiplicative noise. *SIAM J. Numer. Anal.*, **56**, 2045–2069.

CUI, J. & HONG, J. (2019) Strong and weak convergence rates of a spatial approximation for stochastic partial differential equation with one-sided Lipschitz coefficient. *SIAM J. Numer. Anal.*, **57**, 1815–1841.

CUI, J., HONG, J. & SUN, L. (2021) Weak convergence and invariant measure of a full discretization for parabolic SPDEs with non-globally Lipschitz coefficients. *Stochastic Process. Appl.*, **134**, 55–93.

DA PRATO, G. & ZABCZYK, J. (1996) *Ergodicity for Infinite-Dimensional Systems*, vol. 229. Cambridge, UK: Cambridge University Press.

DA PRATO, G. & ZABCZYK, J. (2014) *Stochastic Equations in Infinite Dimensions*, 2nd edn. Cambridge, UK: Cambridge University Press.

DAVIE, A. & GAINES, J. (2001) Convergence of numerical schemes for the solution of parabolic stochastic partial differential equations. *Math. Comp.*, **70**, 121–134.

DE BOUARD, A. & DEBUSSCHE, A. (2006) Weak and strong order of convergence of a semidiscrete scheme for the stochastic nonlinear Schrödinger equation. *Appl. Math. Optim.*, **54**, 369–399.

DEBUSSCHE, A. (2011) Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comp.*, **80**, 89–117.

DEBUSSCHE, A. (2013) Ergodicity results for the stochastic Navier–Stokes equations: an introduction. *Topics in Mathematical Fluid Mechanics*. Lecture Notes in Mathematics, vol. 2073. Berlin Heidelberg: Springer, pp. 23–108.

DEBUSSCHE, A. & FAOU, E. (2012) Weak backward error analysis for SDEs. *SIAM J. Numer. Anal.*, **50**, 1735–1752.

DEBUSSCHE, A. & ODASSO, C. (2005) Ergodicity for a weakly damped stochastic non-linear Schrödinger equation. *J. Evol. Equ.*, **5**, 317–356.

DEBUSSCHE, A. & PRINTEMS, J. (2009) Weak order for the discretization of the stochastic heat equation. *Math. Comp.*, **78**, 845–863.

DÖRSEK, P. (2012) Semigroup splitting and cubature approximations for the stochastic Navier–Stokes equations. *SIAM J. Numer. Anal.*, **50**, 729–746.

DOUC, R., MOULINES, E., PRIOURET, P. & SOULIER, P. (2018) *Markov Chains*. Springer Series in Operations Research and Financial Engineering. Cham, Switzerland: Springer, pp. 3–25.

DUDLEY, R. M. (2002) *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics, vol. 74. Cambridge, UK: Cambridge University Press.

E, W. & LIU, D. (2002) Gibbsian dynamics and invariant measures for stochastic dissipative PDEs. *J. Statist. Phys.*, **108**, 1125–1156.

E, W., MATTINGLY, J. C. & SINAI, Y. (2001) Gibbsian dynamics and ergodicity for the stochastically forced Navier–Stokes equation. *Comm. Math. Phys.*, **224**, 83–106.

FLANDOLI, F. (1994) Dissipativity and invariant measures for stochastic Navier–Stokes equations. *Nonlin. Diff. Equ. Appl. NoDEA*, **1**, 403–423.

FLANDOLI, F. & MASLOWSKI, B. (1995) Ergodicity of the 2-D Navier–Stokes equation under random perturbations. *Comm. Math. Phys.*, **172**, 119–141.

FOIAS, C., JOLLY, M. S., KEVREKIDIS, I. G. & TITI, E. S. (1991) Dissipativity of numerical schemes. *Nonlinearity*, **4**, 591–613.

FOIAS, C., MANLEY, O., ROSA, R. & TEMAM, R. (2001) *Navier–Stokes Equations and Turbulence*. Encyclopedia of Mathematics and its Applications, vol. 83. Cambridge, UK: Cambridge University Press.

FOLDES, J., FRIEDLANDER, S., GLATT-HOLTZ, N. E. & RICHARDS, G. (2017) Asymptotic analysis for randomly forced MHD. *SIAM J. Math. Anal.*, **49**, 4440–4469.

FÖLDES, J., GLATT-HOLTZ, N. E. & RICHARDS, G. (2019) Large Prandtl number asymptotics in randomly forced turbulent convection. *Nonlin. Diff. Equ. Appl. NoDEA*, **26**, 1–43.

FÖLDES, J., GLATT-HOLTZ, N. E., RICHARDS, G. & THOMANN, E. (2015) Ergodic and mixing properties of the Boussinesq equations with a degenerate random forcing. *J. Funct. Anal.*, **269**, 2427–2504.

GLATT-HOLTZ, N. E. (2014) Notes on statistically invariant states in stochastically driven fluid flows. arXiv preprint arXiv:1410.8622.

GLATT-HOLTZ, N. E., HERZOG, D. P. & MATTINGLY, J. C. (2018) Scaling and saturation in infinite-dimensional control problems with applications to stochastic partial differential equations. *Ann. PDE*, **4**, 1–103.

GLATT-HOLTZ, N. E., HOLBROOK, A. J., KROMETIS, J. A. & MONDAINI, C. F. (2022a) A rigorous framework for bias estimation in metropolis-Hastings type sampling algorithms. In press.

GLATT-HOLTZ, N. E., MARTINEZ, V. R. & NGUYEN, H. D. (2022b) The short memory limit for long time statistics in a stochastic Coleman–Gurtin model of heat conduction. arXiv preprint arXiv:2212.05646.

GLATT-HOLTZ, N. E., MARTINEZ, V. R. & RICHARDS, G. H. (2021) On the long-time statistical behavior of smooth solutions of the weakly damped, stochastically-driven KdV equation. arXiv preprint arXiv:2103.12942.

GLATT-HOLTZ, N. E., MATTINGLY, J. C. & RICHARDS, G. (2017) On unique ergodicity in nonlinear stochastic partial differential equations. *J. Statist. Phys.*, **166**, 618–649.

GLATT-HOLTZ, N. E. & MONDAINI, C. F. (2022) Mixing rates for Hamiltonian Monte Carlo algorithms in finite and infinite dimensions. *Stoch. Partial Differ. Equ. Anal. Comput.*, **10**, 1318–1391.

GLATT-HOLTZ, N. E. & ZIANE, M. (2009) Strong pathwise solutions of the stochastic Navier–Stokes system. *Adv. Diff. Equ.*, **14**, 567–600.

GOTTLIEB, S., TONE, F., WANG, C., WANG, X. & WIROSOETISNO, D. (2012) Long time stability of a classical efficient scheme for two-dimensional Navier–Stokes equations. *SIAM J. Numer. Anal.*, **50**, 126–150.

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

HAIRER, M. & MAJDA, A. J. (2010) A simple framework to justify linear response theory. *Nonlinearity*, **23**, 909–922.

HAIRER, M. & MATTINGLY, J. C. (2006) Ergodicity of the 2D Navier–Stokes equations with degenerate stochastic forcing. *Ann. Math. (2)*, **164**, 993–1032.

HAIRER, M. & MATTINGLY, J. C. (2008) Spectral gaps in Wasserstein distances and the 2D stochastic Navier–Stokes equations. *Ann. Probab.*, **36**, 2050–2091.

HAIRER, M. & MATTINGLY, J. C. (2011a) A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.*, **16**, 658–738.

HAIRER, M. & MATTINGLY, J. C. (2011b) Yet another look at Harris' ergodic theorem for Markov chains. *Seminar on Stochastic Analysis, Random Fields and Applications VI*. Progress in Probability, vol. 63. Basel: Springer, pp. 109–117.

HAIRER, M., MATTINGLY, J. C. & 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.

HAUSENBLAS, E. (2010) Weak approximation of the stochastic wave equation. *J. Comput. Appl. Math.*, **235**, 33–58.

HAUSENBLAS, E. & RANDRIANASOLO, T. A. (2019) Time-discretization of stochastic 2-D Navier–Stokes equations with a penalty-projection method. *Numer. Math.*, **143**, 339–378.

Hong, J. & Wang, X. (2019) *Invariant Measures for Stochastic Nonlinear Schrödinger Equations: Numerical Approximations and Symplectic Structures*. Lecture Notes in Mathematics, vol. 2251. Singapore: Springer.

Jentzen, A. & Kloeden, P. E. (2009) The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, **77**, 205–244.

Jentzen, A. & Kurniawan, R. (2021) Weak convergence rates for Euler-type approximations of semilinear stochastic evolution equations with nonlinear diffusion coefficients. *Found. Comput. Math.*, **21**, 445–536.

Johndrow, J. E. & Mattingly, J. C. (2017) Error bounds for approximations of Markov chains used in Bayesian sampling. arXiv preprint arXiv:1711.05382.

Karatzas, I. & Shreve, S. (1991) *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics, vol. 113. New York: Springer.

Kloeden, P. E. & Platen, E. (1992) *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability, vol. 23. Berlin Heidelberg: Springer, pp. 103–160.

Kohatsu-Higa, A. (2001) Weak approximations. A Malliavin calculus approach. *Math. Comp.*, **70**, 135–172.

Kovács, M., Larsson, S. & Lindgren, F. (2012) Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise. *BIT Numer. Math.*, **52**, 85–108.

Kovács, M., Larsson, S. & Lindgren, F. (2013) Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise II. Fully discrete schemes. *BIT Numer. Math.*, **53**, 497–525.

Kuksin, S., Nersesyan, V. & Shirikyan, A. (2020) Exponential mixing for a class of dissipative PDEs with bounded degenerate noise. *Geom. Funct. Anal.*, **30**, 126–187.

Kuksin, S. & Shirikyan, A. (2001) A coupling approach to randomly forced nonlinear PDE's. i. *Comm. Math. Phys.*, **221**, 351–366.

Kuksin, S. & Shirikyan, A. (2002) Coupling approach to white-forced nonlinear PDEs. *J. Math. Pures Appl. (9)*, **81**, 567–602.

Kuksin, S. & Shirikyan, A. (2003) Some limiting properties of randomly forced two-dimensional Navier–Stokes equations. *Proc. R. Soc. Edinb. Sec. A Math.*, **133**, 875–891.

Kuksin, S. & Shirikyan, A. (2012) *Mathematics of Two-Dimensional Turbulence*. Cambridge Tracts in Mathematics, vol. 194. Cambridge, UK: Cambridge University Press.

Kulik, A. (2018) *Ergodic Behavior of Markov Processes*, vol. 67. Berlin: De Gruyter.

Lamba, H., Mattingly, J. C. & Stuart, A. M. (2007) An adaptive Euler–Maruyama scheme for SDEs: convergence and stability. *IMA J. Numer. Anal.*, **27**, 479–506.

Lord, G. J., Powell, C. E. & Shardlow, T. (2014) *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics, vol. 50. Cambridge, UK: Cambridge University Press.

Majda, A. J. & Bertozzi, A. L. (2001) *Vorticity and Incompressible Low*. Cambridge Texts in Applied Mathematics, vol. 27. Cambridge, UK: Cambridge University Press.

Mattingly, J. (2003) On recent progress for the stochastic Navier–Stokes equations. *J. Equ. Dériv. Part.*, 1–52.

Mattingly, J. C. (1999) Ergodicity of 2D Navier–Stokes equations with random forcing and large viscosity. *Comm. Math. Phys.*, **206**, 273–288.

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

Mattingly, J. C., Stuart, A. M. & Higham, D. J. (2002) Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, **101**, 185–232.

Mattingly, J. C., Stuart, A. M. & Tretyakov, M. V. (2010) Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, **48**, 552–577.

Meyn, S. P. & Tweedie, R. L. (2009) *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge, UK: Cambridge University Press.

Mikulevicius, R. & Rozovskii, B. L. (2004) Stochastic Navier–Stokes equations for turbulent flows. *SIAM J. Math. Anal.*, **35**, 1250–1310.

Milshtein, G. N. (1979) A method of second-order accuracy integration of stochastic differential equations. *Theory Probab. Appl.*, **23**, 396–401.

MILSHTEIN, G. N. (1995) Weak approximation of solutions of systems of stochastic differential equations. *Numerical Integration of Stochastic Differential Equations*. Mathematics and its Applications, vol. 313. Dordrecht: Springer, pp. 101–134.

MILSTEIN, G. N. & TRETYAKOV, M. V. (2004) *Stochastic Numerics for Mathematical Physics*, vol. 456. Springer.

MILSTEIN, G. N. & TRETYAKOV, M. V. (2021) Mean-square approximation of Navier–Stokes equations with additive noise in vorticity-velocity formulation. *Numer. Math. Theor. Meth. Appl.*, **14**, 1–30.

PAZY, A. (1983) *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences, vol. 44. New York, NY: Springer.

PRINTEMS, J. (2001) On the discretization in time of parabolic stochastic partial differential equations. *ESAIM Math. Model. Numer. Anal.*, **35**, 1055–1078.

SHARDLOW, T. & STUART, A. M. (2000) A perturbation theory for ergodic Markov chains and application to numerical approximations. *SIAM J. Numer. Anal.*, **37**, 1120–1137.

SHEFFIELD, R. D. (1956) On pseudo-inverses of linear transformations in Banach spaces. *Technical Report*. Oak Ridge National Lab., Tenn.

SHIRIKYAN, A., KUKSIN, S. & PIATNISKI, A. (2002) A coupling approach to randomly forced nonlinear PDE's. Ii. *Comm. Math. Phys.*, **230**, 81–85.

SZEPESSY, A., TEMPONE, R. & ZOURARIS, G. E. (2001) Adaptive weak approximation of stochastic differential equations. *Comm. Pure Appl. Math.*, **54**, 1169–1214.

TALAY, D. (1984) Efficient numerical schemes for the approximation of expectations of functionals of the solution of a SDE, and applications. *Filtering and Control of Random Processes* (H. Korezlioglu, G. Mazziotto & J. Szpirglas eds.). Berlin Heidelberg: Springer, pp. 294–313.

TALAY, D. (1986) Discrétisation d'une équation différentielle stochastique et calcul approché d'espérances de fonctionnelles de la solution. *ESAIM. Math. Model. Numer. Anal.*, **20**, 141–179.

TALAY, D. (1990) Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics Int. J. Probab. Stoch. Process.*, **29**, 13–36.

TALAY, D. (2002) Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Process. Related Fields*, **8**, 163–198.

TALAY, D. & TUBARO, L. (1990) Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.*, **8**, 483–509.

TEMAM, R. (2001) *Navier–Stokes Equations: Theory and Numerical Analysis*. Providence, RI: AMS Chelsea Publishing, pp. xiv+408. Reprint of the 1984 edition.

TSYBAKOV, A. B. (2009) *Introduction to Nonparametric Estimation*. Springer Series in Statistics. New York, NY: Springer Science+Business Media.

VILLANI, C. (2008) *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, A Series of Comprehensive Studies in Mathematics, vol. 338. Berlin Heidelberg: Springer.

WANG, X. (2010) Approximation of stationary statistical properties of dissipative dynamical systems: time discretization. *Math. Comp.*, **79**, 259–280.

WANG, X. (2016) Weak error estimates of the exponential Euler scheme for semi-linear SPDEs without Malliavin calculus. *Discrete Contin. Dyn. Syst.*, **36**, 481–497.

WANG, X. & GAN, S. (2013) Weak convergence analysis of the linear implicit Euler method for semilinear stochastic partial differential equations with additive noise. *J. Math. Anal. Appl.*, **398**, 151–169.

Appendix A. Distance-like functions

Here we present some simple general results concerning distance-like functions on a Polish space X and their corresponding Wasserstein-like extensions to $\mathrm{Pr}(X)$, as recalled in Section 2.1.

We start by showing that if a given distance-like function ρ satisfies a generalized form of triangle inequality, namely (A.1) below, together with a suitable set of conditions, then the corresponding distance-like function ρ_α for a fixed parameter $\alpha > 0$, defined in (2.8), satisfies an inequality of the form (2.24) from assumption (H1) in Theorem 2.5. The proof below follows similar ideas from Hairer *et al.* (2011, Lemma 4.14).

PROPOSITION A.1 Let $(X, \|\cdot\|)$ be a Banach space and let $\rho : X \times X \to \mathbb{R}^+$ be a distance-like function on X satisfying the following conditions:

i. ρ is bounded, i.e. there exists a constant $M > 0$ such that $\rho(u, v) \leq M$ for all $u, v \in X$.

ii. There exists a constant $K > 0$ such that

$$\rho(u, v) \leq K\left[\rho(u, w) + \rho(w, v)\right] \quad \text{for all } u, v, w \in X. \tag{A.1}$$

iii. There exists a constant $c > 0$ for which the following holds: if $\rho(u, v) < c$ for some $u, v \in X$, then $\|u\|^2 \leq \gamma\|v\|^2 + C$ for some constants $\gamma \geq 2$ and $C > 0$, which are independent of u and v.

Then, for the distance-like function $\rho_\alpha : X \times X \to \mathbb{R}^+$ defined for a fixed parameter $\alpha > 0$ by

$$\rho_\alpha(u, v) = \rho(u, v)^{1/2} \exp\left(\alpha\|u\|^2 + \alpha\|v\|^2\right) \quad \text{for all } u, v \in X,$$

it follows that there exists a constant $\tilde{K} > 0$ such that

$$\rho_\alpha(u, v) \leq \tilde{K}\left[\rho_{\gamma\alpha}(u, w) + \rho_{\gamma\alpha}(w, v)\right] \quad \text{for all } u, v, w \in X, \tag{A.2}$$

where $\gamma \geq 2$ is the constant from assumption (iii).

Proof. Let $u, v, w \in X$. Since, for any $\alpha > 0$, ρ_α is symmetric, we may assume without loss of generality that $\|v\| \leq \|u\|$, with $\gamma \geq 2$ being the constant from assumption (iii).

First, suppose that $\rho(u, w) \geq c$. Then, by invoking assumption (i), we obtain that

$$\begin{aligned}\rho_\alpha(u, v) &\leq M^{1/2} \exp\left(\alpha\|u\|^2 + \alpha\|v\|^2\right) \leq M^{1/2} \frac{\rho(u, w)^{1/2}}{c^{1/2}} \exp\left(2\alpha\|u\|^2\right) \\ &\leq \frac{M^{1/2}}{c^{1/2}} \rho_{\gamma\alpha}(u, w) \leq \frac{M^{1/2}}{c^{1/2}}\left[\rho_{\gamma\alpha}(u, w) + \rho_{\gamma\alpha}(w, v)\right]. \end{aligned}\tag{A.3}$$

On the other hand, if $\rho(u, w) < c$ then by invoking assumptions (ii) and (iii) it follows that

$$\begin{aligned}\rho_\alpha(u, v) &\leq K^{1/2}\left[\rho(u, w)^{1/2} + \rho(w, v)^{1/2}\right] \exp\left(\alpha\|u\|^2 + \alpha\|v\|^2\right) \\ &\leq K^{1/2}\left[\rho(u, w)^{1/2} \exp\left(2\alpha\|u\|^2\right) + \rho(w, v)^{1/2} \exp\left(\alpha\gamma\|w\|^2 + \alpha C + \alpha\|v\|^2\right)\right] \\ &\leq \tilde{C}\left[\rho_{\gamma\alpha}(u, w) + \rho_{\gamma\alpha}(w, v)\right], \end{aligned}\tag{A.4}$$

where $\tilde{C} = K^{1/2} \exp(\alpha C)$.

From (A.3) and (A.4), we conclude that (A.2) holds with $\tilde{K} = \max\{(M/c)^{1/2}, \tilde{C}\}$. □

In the following result, we show that a generalized triangle inequality satisfied by given distance-like functions, namely (A.6) below, induces an analogous inequality for the corresponding Wasserstein-like extensions, (A.7). The proof relies essentially on the Disintegration theorem (see e.g. Ambrosio *et al.*, 2005, Lemma 5.3.2): fixed measures $\mu, \nu, \tilde{\nu} \in \Pr(X)$, and given any couplings $\Gamma \in \mathscr{C}(\mu, \tilde{\nu})$,

$\Gamma' \in \mathscr{C}(\tilde{\nu}, \nu)$, it provides a way of constructing a coupling $\Gamma'' \in \mathscr{C}(\mu, \nu)$, so that one can pass from (A.6) to (A.7). Before we state the result, let us recall a few definitions.

Let $(\mathscr{X}, \Sigma_{\mathscr{X}})$ and $(\mathscr{Y}, \Sigma_{\mathscr{Y}})$ be measurable spaces. Given a measurable function $\phi : \mathscr{X} \to \mathscr{Y}$ and a measure $\mu \in \Pr(\mathscr{X})$, the *pushforward of μ by ϕ*, denoted by $\phi^*\mu$, is defined as the measure on $\mathscr{Y}$ given by

$$\phi^*\mu(A) := \mu(\phi^{-1}(A)) \quad \text{for any } A \in \Sigma_{\mathscr{X}},$$

where $\phi^{-1}(A)$ denotes the preimage of the set A by ϕ. Moreover, given a $(\phi^*\mu)$-integrable function $\psi : \mathscr{Y} \to \mathbb{R}$, it follows that the composition $\psi \circ \phi : \mathscr{X} \to \mathbb{R}$ is μ-integrable and the following change of variables formula holds

$$\int_{\mathscr{Y}} \psi(u)(\phi^*\mu)(\mathrm{d}u) = \int_{\mathscr{X}} \psi(\phi(u))\mu(\mathrm{d}u). \tag{A.5}$$

PROPOSITION A.2 Let X be a Polish space. Suppose there exist distance-like functions $\rho_1, \rho_2, \rho_3 : X \times X \to \mathbb{R}^+$ for which there exists a constant $C > 0$ such that

$$\rho_1(u, v) \leq C\left[\rho_2(u, w) + \rho_3(w, v)\right] \quad \text{for all } u, v, w \in X. \tag{A.6}$$

Let $\mathscr{W}_{\rho_1}$, $\mathscr{W}_{\rho_2}$ and $\mathscr{W}_{\rho_3}$ be the Wasserstein-like extensions of ρ_1, ρ_2 and ρ_3, respectively, to $\Pr(X)$, according to the definition given in (2.2). Then,

$$\mathscr{W}_{\rho_1}(\mu, \mu') \leq C\left[\mathscr{W}_{\rho_2}(\mu, \tilde{\mu}) + \mathscr{W}_{\rho_3}(\tilde{\mu}, \mu')\right] \quad \text{for all } \mu, \mu', \tilde{\mu} \in \Pr(X). \tag{A.7}$$

Proof. Let $\pi_1, \pi_2 : X \times X \to X$ denote the projection functions onto the first and second components, respectively. Namely, $\pi_1(u, v) = u$ and $\pi_2(u, v) = v$ for all $u, v \in X$. Then, recalling the definition of the family of couplings $\mathscr{C}(\mu, \mu')$ of any two measures $\mu, \mu' \in \Pr(X)$, given in Section 2.1, it follows that, for any $\Gamma \in \mathscr{C}(\mu, \mu') \subset \Pr(X \times X)$, $\pi_1^*\Gamma = \mu$ and $\pi_2^*\Gamma = \mu'$.

Fix $\mu, \mu', \tilde{\mu} \in \Pr(X)$. From (2.2), it follows that for any given $\tilde{\varepsilon} > 0$ there exist $\Gamma \in \mathscr{C}(\mu, \tilde{\mu})$ and $\Gamma' \in \mathscr{C}(\tilde{\mu}, \mu')$ such that

$$\int_{X \times X} \rho_2(u, v)\Gamma(\mathrm{d}u, \mathrm{d}v) < \mathscr{W}_{\rho_2}(\mu, \tilde{\mu}) + \tilde{\varepsilon},$$

and

$$\int_{X \times X} \rho_3(u, v)\Gamma'(\mathrm{d}u, \mathrm{d}v) < \mathscr{W}_{\rho_3}(\tilde{\mu}, \mu') + \tilde{\varepsilon}.$$

Further, let us denote by $\pi_{i,j} : X \times X \times X \to X \times X$, $i, j = 1, 2, 3$, the projection functions

$$\pi_{i,j}(u_1, u_2, u_3) = (u_i, u_j), \quad \text{for all } u_1, u_2, u_3 \in X.$$

Since $\pi_2^*\Gamma = \tilde{\mu} = \pi_1^*\Gamma'$, it follows from the Disintegration theorem (see e.g. Ambrosio *et al.*, 2005, Lemma 5.3.2) that there exists $\tilde{\Gamma} \in \Pr(X \times X \times X)$ such that $\pi_{1,2}^*\tilde{\Gamma} = \Gamma$ and $\pi_{2,3}^*\tilde{\Gamma} = \Gamma'$. Consequently,

$\pi_{1,3}^{*}\tilde{\Gamma} \in \mathscr{C}(\mu,\mu')$ and

$$\mathscr{W}_{\rho_1}(\mu,\mu') \leq \int_{X\times X} \rho_1(u,v)\pi_{1,3}^{*}\tilde{\Gamma}(\mathrm{d}u,\mathrm{d}v) = \int_{X\times X\times X} \rho_1(\pi_{1,3}(u,w,v))\tilde{\Gamma}(\mathrm{d}u,\mathrm{d}w,\mathrm{d}v). \tag{A.8}$$

From assumption (A.6), we have that for any $u, v, w \in X$

$$\begin{aligned}\rho_1(\pi_{1,3}(u,w,v)) = \rho_1(u,v) &\leq C\left[\rho_2(u,w) + \rho_3(w,v)\right]\\ &= C\left[\rho_2(\pi_{1,2}(u,w,v)) + \rho_3(\pi_{2,3}(u,w,v))\right].\end{aligned} \tag{A.9}$$

Plugging (A.9) in (A.8) and changing variables as in (A.5), we deduce that

$$\begin{aligned}\mathscr{W}_{\rho_1}(\mu,\mu') &\leq C\left[\int_{X\times X} \rho_2(u,w)\pi_{1,2}^{*}\tilde{\Gamma}(\mathrm{d}u,\mathrm{d}w) + \int_{X\times X} \rho_3(w,v)\pi_{2,3}^{*}\tilde{\Gamma}(\mathrm{d}w,\mathrm{d}v)\right]\\ &= C\left[\int_{X\times X} \rho_2(u,w)\Gamma(\mathrm{d}u,\mathrm{d}w) + \int_{X\times X} \rho_3(w,v)\Gamma'(\mathrm{d}w,\mathrm{d}v)\right]\\ &< C\left[\mathscr{W}_{\rho_2}(\mu,\tilde{\mu}) + \mathscr{W}_{\rho_3}(\tilde{\mu},\mu') + 2\tilde{\varepsilon}\right].\end{aligned} \tag{A.10}$$

Since $\tilde{\varepsilon} > 0$ is arbitrary, taking the limit as $\tilde{\varepsilon}$ goes to 0 in (A.10) we conclude (A.7). □

Appendix B. Proof of Theorem 3.7

With the same notation from (3.2), we write the Galerkin system (3.12) in the following functional form

$$\mathrm{d}\xi_N + \left[\nu A\xi_N + \Pi_N B(\xi_N,\xi_N)\right]\,\mathrm{d}t = \Pi_N\sigma\,\mathrm{d}W. \tag{B.1}$$

The following preliminary lemma provides some suitable bounds for the analytic semigroup $e^{-\nu tA}$, $t \geq 0$, generated by the operator $-\nu A$. For the proof, we refer to Pazy (1983, Theorem 6.13, Chapter 2). The notation $\|\cdot\|_{\mathscr{L}(\dot{L}^2)}$ below refers to the standard operator norm of a linear operator on $\dot{L}^2$.

LEMMA B.1 For every $a \geq 0$ and $b \in (0,1]$, there exist constants $c_a > 0$ and $c_b > 0$ such that

$$\|A^a e^{-\nu tA}\|_{\mathscr{L}(\dot{L}^2)} \leq c_a(\nu t)^{-a}, \tag{B.2}$$

$$\|A^{-b}(I - e^{-\nu tA})\|_{\mathscr{L}(\dot{L}^2)} \leq c_b(\nu t)^{b}, \tag{B.3}$$

for all $t > 0$.

Having fixed the necessary terminology, we proceed to show the desired Hölder regularity for solutions of the Galerkin system (3.12).

Proof of Theorem 3.7. We only show a proof of inequality (3.19), since the proof of (3.18) is simpler and follows entirely analogously. Fix $T > 0$, $m \in \mathbb{N}$ and $\tilde{p} \in (0, 1/2)$. We consider the mild form of the solution ξ_N that follows from the functional formulation (B.1), namely

$$\xi_N(t) = e^{-\nu tA}\Pi_N\xi_0 - \int_0^t e^{-\nu(t-\tau)A}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau + \int_0^t e^{-\nu(t-\tau)A}\Pi_N\sigma\,\mathrm{d}W(\tau),$$

for every $t \geq 0$. Thus, for every $s, t \in [0, T]$,

$$\begin{aligned}\nabla\xi_N(t) - \nabla\xi_N(s) = &\left(e^{-\nu tA} - e^{-\nu sA}\right)\nabla\Pi_N\xi_0 \\ &- \left(\int_0^t e^{-\nu(t-\tau)A}\nabla\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau - \int_0^s e^{-\nu(s-\tau)A}\nabla\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau\right) \\ &+ \left(\int_0^t e^{-\nu(t-\tau)A}\nabla\Pi_N\sigma\,\mathrm{d}W(\tau) - \int_0^s e^{-\nu(s-\tau)A}\nabla\Pi_N\sigma\,\mathrm{d}W(\tau)\right) \\ = &\ (I) + (II) + (III).\end{aligned} \tag{B.4}$$

We proceed to estimate each term in the right-hand side of (B.4). Without loss of generality, let us assume $s < t$. We estimate (I) as

$$\begin{aligned}|(I)| = \left|\left(e^{-\nu tA} - e^{-\nu sA}\right)\nabla\Pi_N\xi_0\right| &= \left|e^{-\nu sA}(e^{-\nu(t-s)A} - I)\nabla\Pi_N\xi_0\right| \\ &\leq \|e^{-\nu sA}\|_{\mathscr{L}(\dot{L}^2)}\|A^{-\tilde{p}}(e^{-\nu(t-s)A} - I)\|_{\mathscr{L}(\dot{L}^2)}|A^{\tilde{p}}\nabla\xi_0| \\ &\leq c|t-s|^{\tilde{p}}|A\xi_0|,\end{aligned}$$

where the last inequality follows from Lemma B.1, and the fact that $A^{\tilde{p}}\nabla\xi_0 = A^{\tilde{p}}A^{1/2}\xi_0 = A^{\tilde{p}+1/2}\xi_0$, so that since $\tilde{p} \in (0, 1/2)$ we have $|A^{\tilde{p}}\nabla\xi_0| = \|\xi_0\|_{\dot{H}^{2\tilde{p}+1}} \leq \|\xi_0\|_{\dot{H}^2}$, see Section 3.1.1. Hence,

$$\mathbb{E}[|(I)|^m] \leq c|t-s|^{m\tilde{p}}|A\xi_0|^m. \tag{B.5}$$

Now for term (II) we have

$$\begin{aligned}|(II)| &= \left|\int_0^t e^{-\nu(t-\tau)A}A^{1/2}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau - \int_0^s e^{-\nu(s-\tau)A}A^{1/2}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau\right| \\ &\leq \left|\int_0^s (e^{-\nu(t-\tau)A} - e^{-\nu(s-\tau)A})A^{1/2}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau\right| + \left|\int_s^t e^{-\nu(t-\tau)A}A^{1/2}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau\right| \\ &= |(II_a)| + |(II_b)|.\end{aligned}$$

Notice that

$$\begin{aligned}\mathbb{E}[|(II_a)|^m] &= \mathbb{E}\left[\left|\int_0^s e^{-\nu(s-\tau)A}(e^{-\nu(t-s)A} - I)A^{1/2}\Pi_N B(\xi_N,\xi_N)\,\mathrm{d}\tau\right|^m\right] \\ &\leq \mathbb{E}\left[\left(\int_0^s \|A^{\tilde{p}+1/2}e^{-\nu(s-\tau)A}\|_{\mathscr{L}(\dot{L}^2)}\|A^{-\tilde{p}}(e^{-\nu(t-s)A} - I)\|_{\mathscr{L}(\dot{L}^2)}\,|\Pi_N B(\xi_N,\xi_N)|\,\mathrm{d}\tau\right)^m\right] \\ &\leq \frac{c}{\nu^{m/2}}\mathbb{E}\left[\sup_{0\leq\tau\leq T}|\Pi_N B(\xi_N,\xi_N)|^m\right]\left(\int_0^s |s-\tau|^{-\tilde{p}-1/2}|t-s|^{\tilde{p}}\,\mathrm{d}\tau\right)^m,\end{aligned} \tag{B.6}$$

where in the last inequality we invoked Lemma B.1 once again.

With inequality (3.9) for the nonlinear term and estimate (3.13) from Proposition 3.4, it follows that

$$\begin{aligned}\frac{c}{\nu^{m/2}}\mathbb{E}\left[\sup_{0\leq\tau\leq T}|\Pi_N B(\xi_N,\xi_N)|^m\right] &\leq \frac{c}{\nu^{m/2}}\mathbb{E}\left[\sup_{0\leq\tau\leq T}|B(\xi_N,\xi_N)|^m\right] \leq \frac{c}{\nu^{m/2}}\mathbb{E}\left[\sup_{0\leq\tau\leq T}|\nabla\xi_N|^{2m}\right]\\ &\leq C\Big(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}\Big),\end{aligned}$$

where C is a constant depending on $m, T, \nu, |\sigma|$ and $|\nabla\sigma|$. Thus, from (B.6) and since $\tilde{p}\in(0,1/2)$

$$\begin{aligned}\mathbb{E}\Big[|(II_a)|^m\Big] &\leq C\Big(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}\Big)|t-s|^{m\tilde{p}}\left(\int_0^s|s-\tau|^{-\tilde{p}-1/2}\,\mathrm{d}\tau\right)^m\\ &\leq C\Big(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}\Big)|t-s|^{m\tilde{p}}.\end{aligned}\tag{B.7}$$

Similarly, we have for (II_b) that

$$\begin{aligned}\mathbb{E}\Big[|(II_b)|^m\Big] &\leq \mathbb{E}\left[\left(\int_s^t\|A^{1/2}\,e^{-\nu(t-\tau)A}\|_{\mathscr{L}(\dot{L}^2)}|\Pi_N B(\xi_N,\xi_N)|\,\mathrm{d}\tau\right)^m\right]\\ &\leq \frac{c}{\nu^{m/2}}\mathbb{E}\left[\sup_{0\leq\tau\leq T}|\Pi_N B(\xi_N,\xi_N)|^m\right]\left(\int_s^t|t-\tau|^{-1/2}\,\mathrm{d}\tau\right)^m\\ &\leq C\left(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}\right)|t-s|^{m/2}\\ &\leq C\Big(1+|\xi_0|^{4m}+|\nabla\xi_0|^{2m}\Big)|t-s|^{m\tilde{p}}T^{(-\tilde{p}+1/2)m}.\end{aligned}\tag{B.8}$$

Lastly, we estimate (III) as

$$\begin{aligned}|(III)| &\leq \left|\int_0^s(e^{-\nu(t-\tau)A}-e^{-\nu(s-\tau)A})\nabla\Pi_N\sigma\,\mathrm{d}W(\tau)\right|+\left|\int_s^t e^{-\nu(t-\tau)A}\nabla\Pi_N\sigma\,\mathrm{d}W(\tau)\right|\\ &= |(III_a)|+|(III_b)|.\end{aligned}$$

For each fixed $s,t\in[0,T]$, we define for every $r\in[0,s]$

$$M_r := \int_0^r(e^{-\nu(t-\tau)A}-e^{-\nu(s-\tau)A})\nabla\Pi_N\sigma\,\mathrm{d}W(\tau).$$

Then, $\{M_r\}_{0\leq r\leq s}$ is a martingale. By Burkholder–Davis–Gundy inequality (Karatzas & Shreve, 1991, Theorem 3.28), for every $p\in(0,\infty)$,

$$\mathbb{E}\Big[|M_s|^p\Big] \leq \mathbb{E}\left[\sup_{0\leq r\leq s}|M_r|^p\right] \leq c\,\mathbb{E}\left(\langle M\rangle_s^{p/2}\right),$$

where

$$\langle M\rangle_s = \int_0^s \left|(e^{-\nu(t-\tau)A} - e^{-\nu(s-\tau)A})\nabla\Pi_N\sigma\right|^2 \mathrm{d}\tau.$$

Hence, invoking Lemma B.1 again,

$$\begin{aligned}
\mathbb{E}\Big[|(III_a)|^m\Big] = \mathbb{E}\Big[|M_s|^m\Big] &\le c\left(\int_0^s |(e^{-\nu(t-\tau)A} - e^{-\nu(s-\tau)A})\nabla\Pi_N\sigma|^2\,\mathrm{d}\tau\right)^{m/2}\\
&\le c\left(\int_0^s |e^{-\nu(s-\tau)A}(e^{-\nu(t-s)A} - I)\nabla\Pi_N\sigma|^2\,\mathrm{d}\tau\right)^{m/2}\\
&\le c\left(\int_0^s \|A^{\tilde p}e^{-\nu(s-\tau)A}\|^2_{\mathscr{L}(\dot L^2)}\|A^{-\tilde p}(e^{-\nu(t-s)A} - I)\|^2_{\mathscr{L}(\dot L^2)}|\nabla\sigma|^2\,\mathrm{d}\tau\right)^{m/2}\\
&\le c\left(\int_0^s |s-\tau|^{-2\tilde p}|t-s|^{2\tilde p}|\nabla\sigma|^2\,\mathrm{d}\tau\right)^{m/2}\\
&= c|t-s|^{m\tilde p}|\nabla\sigma|^m\left(\int_0^s |s-\tau|^{-2\tilde p}\,\mathrm{d}\tau\right)^{m/2}\\
&\le c|t-s|^{m\tilde p}|\nabla\sigma|^m s^{(-\tilde p+1/2)m} \le c|t-s|^{m\tilde p}|\nabla\sigma|^m T^{(-\tilde p+1/2)m},
\end{aligned}\tag{B.9}$$

where the last inequality holds thanks to the assumption that $\tilde p \in (0, 1/2)$.

Analogously, we estimate $\mathbb{E}|(III_b)|^m$ as

$$\begin{aligned}
\mathbb{E}\Big[|(III_b)|^m\Big] \le \left(\int_s^t |e^{-\nu(t-\tau)A}\nabla\Pi_N\sigma|^2\,\mathrm{d}\tau\right)^{m/2} &\le \left(\int_s^t \|e^{-\nu(t-\tau)A}\|^2_{\mathscr{L}(\dot L^2)}|\nabla\sigma|^2\,\mathrm{d}\tau\right)^{m/2}\\
&\le |\nabla\sigma|^m|t-s|^{m/2}\\
&\le |t-s|^{m\tilde p}\,|\nabla\sigma|^m\,T^{(-\tilde p+1/2)m}.
\end{aligned}\tag{B.10}$$

Therefore, it follows from (B.4) and the estimates (B.5), (B.7)–(B.10) above that for all $s, t \in [0, T]$ with $s \le t$

$$\mathbb{E}\Big[|\nabla\xi_N(t) - \nabla\xi_N(s)|^m\Big] \le C|t-s|^{m\tilde p}\Big[1 + |\xi_0|^{4m} + |\nabla\xi_0|^{2m} + |A\xi_0|^m\Big],\tag{B.11}$$

where $C = C(m, \tilde p, T, \nu, |\sigma|, |\nabla\sigma|)$. This concludes the proof of (3.18). Clearly, by following similar steps as above one can show that (3.18) and (3.19) also hold with $\xi_N(t)$ replaced by the solution $\xi(t)$, $t \ge 0$, of (3.1) satisfying $\xi(0) = \xi_0$ almost surely. $\square$