J. Evol. Equ.

https://doi.org/10.1007/s00028-021-00708-z

Journal of Evolution Equations

Stochastic model reduction: convergence and applications to climate equations

Sigurd Assing, Franco Flandoli and Umberto Pappalettera

Abstract. We study stochastic model reduction for evolution equations in infinite-dimensional Hilbert spaces and show the convergence to the reduced equations via abstract results of Wong–Zakai type for stochastic equations driven by a scaled Ornstein–Uhlenbeck process. Both weak and strong convergence are investigated, depending on the presence of quadratic interactions between reduced variables and driving noise. Finally, we are able to apply our results to a class of equations used in climate modeling.

1. Introduction

In this paper we study stochastic model reduction for a system of nonlinear evolution equations in infinite-dimensional Hilbert spaces which is general enough to cover well-established systems of equations used in climate modeling. The big advantage of such a procedure is the lower complexity of the reduced equations, since complexity is still one of the major issues when predicting the evolution of systems over time spans which are typical for climate rather than meteorology.

Following [9,17], we assume that the *climate variables* of the system, i.e., those more relevant to climate prediction, evolve on longer time scales than the *unresolved variables*, which can be modeled stochastically and have a typical time scale much shorter than the climate variables. To be able to close the equation for the climate variables, the task is to understand the effects of unresolved variables when stretching time to climate time. In what follows, we also refer to climate variables as *resolved variables*.

Climate modeling typically starts with equations containing quadratic nonlinearities which can describe many features of oceanic and atmospheric dynamics at meteorological time—see [18,25]. In abstract mathematical terms, such equations would look like

$$\frac{\mathrm{d}Z_t}{\mathrm{d}t} = f_t + AZ_t + B(Z_t, Z_t), \tag{1}$$

Keywords: Stochastic model reduction, Wong–Zakai approximation theorems, Ornstein–Uhlenbeck process.

Published online: 02 May 2021

where $A : H \to H$ is a linear operator, $B : H \times H \to H$ is a bilinear operator, and f is an external forcing term. Here, the variable Z taking values in H is supposed to be a complex mix of climate and unresolved variables, and hence, the space H has to be 'big enough' to 'host' variables of that type. We therefore choose H to be a separable infinite-dimensional Hilbert space.

Now, there is a variety of procedures to identify climate variables in practice which we will not discuss in this paper. We rather assume that climate variables have been identified spanning a Hilbert-subspace $H_d \subset H$, and we further assume that the orthogonal complement H_∞, $H = H_d \oplus H_\infty$, gives the space of unresolved variables. When projecting Z onto H_d, H_∞ via the projection maps π_d, π_∞, Eq. (1) gives raise to two equations

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = f_t^1 + \tilde{A}_1^1 X_t + A_2^1 Y_t + \tilde{B}_{11}^1(X_t, X_t) + B_{12}^1(X_t, Y_t) + B_{22}^1(Y_t, Y_t) \tag{2}$$

and

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = f_t^2 + A_1^2 X_t + A_2^2 Y_t + B_{11}^2(X_t, X_t) + B_{12}^2(X_t, Y_t) + B_{22}^2(Y_t, Y_t) \tag{3}$$

for the collection of climate variables $X = \pi_d(Z)$ and unresolved variables $Y = \pi_\infty(Z)$, respectively.

The next step, called stochastic climate modeling, consists in replacing the complicated nonlinear self-interaction term in (3) by a linear random term. Such a replacement could be justified by the assumption that quickly varying fluctuations of small-scale unresolved variables are more or less indistinguishable from the combined effect of a large number of weakly coupled factors, usually leading to Gaussian driving forces via central limit theorem. But such effects would only become visible at climate time and not at meteorological time used in (2) and (3), so that we are looking to replace $B_{22}^2(Y_{\varepsilon^{-1}t}, Y_{\varepsilon^{-1}t})$ by a linear random term, stretching meteorological time to $\varepsilon^{-1}t$, using a small parameter $\varepsilon \ll 1$.

In this work, following [17,22], we suppose that

$$B_{22}^2(Y_{\varepsilon^{-1}t}, Y_{\varepsilon^{-1}t}) \text{ is replaced by } -\mu\varepsilon^{-1}Y_{\varepsilon^{-1}t} + \sigma\dot{W}_t,$$

where μ, σ are positive constants, and $\dot{W}$ is Gaussian noise, white in time, and colored in space. This way, the parameter ε is used to scale time, but also to adjust for the size of the involved variables when scaling time.

Another assumption made in [17] is that climate variables at climate time have small forcing and self-interaction, and hence, we also suppose that

$$f_{\varepsilon^{-1}t}^1 + \tilde{A}_1^1 X_{\varepsilon^{-1}t} + \tilde{B}_{11}^1(X_{\varepsilon^{-1}t}, X_{\varepsilon^{-1}t}) \text{ is replaced by } \varepsilon F_t^1 + \varepsilon A_1^1 X_{\varepsilon^{-1}t} + \varepsilon B_{11}^1(X_{\varepsilon^{-1}t}, X_{\varepsilon^{-1}t}),$$

avoiding so-called fast forcing and fast waves.

All in all, when introducing the notation $X_t^\varepsilon = X_{\varepsilon^{-1}t}$ for climate variables at climate time, and $Y_t^\varepsilon = \varepsilon^{-1}Y_{\varepsilon^{-1}t}$ for the effect of unresolved variables at climate time, Eqs. (2) and (3) translate into

$$\frac{\mathrm{d}X_t^\varepsilon}{\mathrm{d}t} = F_t^1 + A_1^1 X_t^\varepsilon + A_2^1 Y_t^\varepsilon + B_{11}^1(X_t^\varepsilon, X_t^\varepsilon) + B_{12}^1(X_t^\varepsilon, Y_t^\varepsilon) + \varepsilon B_{22}^1(Y_t^\varepsilon, Y_t^\varepsilon), \quad (4)$$

$$\frac{\mathrm{d}Y_t^\varepsilon}{\mathrm{d}t} = \varepsilon^{-2} f^2_{\varepsilon^{-1}t} + \varepsilon^{-2} A_1^2 X_t^\varepsilon + \varepsilon^{-1} A_2^2 Y_t^\varepsilon + \varepsilon^{-2} B_{11}^2(X_t^\varepsilon, X_t^\varepsilon) + \varepsilon^{-1} B_{12}^2(X_t^\varepsilon, Y_t^\varepsilon) - \varepsilon^{-2} Y_t^\varepsilon + \varepsilon^{-2}\dot{W}_t, \quad (5)$$

where we have set $\mu = \sigma = 1$ for the sake of simplicity.

The hope is now that, when ε tends to zero, climate variables at climate time can be approximated by a random variable $\bar{X}$ which solves a closed stochastic equation with new coefficients not depending on unresolved variables any more. Of course, these new coefficients will be functions of the coefficients of Eqs. (4) and (5), and the process of finding these new coefficients is called *stochastic model reduction*.

Stochastic model reduction of finite-dimensional systems similar to (4), (5) was extensively discussed in [17]. However, one of the key steps, i.e., proving the convergence $X^\varepsilon \to \bar{X}$, $\varepsilon \downarrow 0$, was kept rather short. Indeed, the authors first sketch a perturbation method based on a theorem by T.G. Kurtz, [16], which is their general method, and they then briefly describe a so-called *direct averaging method* for special cases based on limits of solutions to stochastic differential equations. In particular, the latter method lacks a certain amount of rigor because the convergence of the involved stochastic processes is not shown, and this gap has not been closed in follow-up papers—see [6,7,13] for example.

In this paper we are not only closing this gap, but also develop a new method of proof.

We at first identify the limit process $\bar{X}$, and then study the convergence $X^\varepsilon \to \bar{X}$ as $\varepsilon \downarrow 0$, when X^ε solves a general evolution equation of type

$$\frac{\mathrm{d}X_t^\varepsilon}{\mathrm{d}t} = F(t, X_t^\varepsilon) + \sigma(t, X_t^\varepsilon) Y_t^\varepsilon + \varepsilon\beta(Y_t^\varepsilon, Y_t^\varepsilon), \quad (6)$$

where Y^ε is a decoupled infinite-dimensional Ornstein–Uhlenbeck process satisfying

$$\frac{\mathrm{d}Y_t^\varepsilon}{\mathrm{d}t} = -\varepsilon^{-2} Y_t^\varepsilon + \varepsilon^{-2}\dot{W}_t. \quad (7)$$

Since Eq. (6) is more general than (4), once stochastic model reduction is established for the system (6), (7) with decoupled unresolved variables, it also follows for an interesting subclass of systems of type (4), (5) with coupled unresolved variables—basically those systems for which $B_{12}^2 = 0$, see Theorem 5.3. Part (ii) of this theorem deals with the case of linear scattering, that is $B_{22}^1 = 0$, and in this case we achieve showing 'strong' convergence in probability:

$$\lim_{\varepsilon\to 0} \mathbb{P}\left\{\sup_{t\le T} \|X_t^\varepsilon - \bar{X}_t\|_{H_d} > \delta\right\} = 0, \quad \forall \delta > 0, \quad (8)$$

on a given climate time interval $[0, T]$. When the quadratic interaction term B_{22}^1 is non-trivial, we can only show convergence in law, as stated in Theorem 5.3(i). We

refer to Remark 4.3(ii) for an argument which suggests that one cannot expect much more than a weak-type convergence in the general case. This insight of course sheds new light on the results given in [17] and follow-up papers.

At this point it should be mentioned that throughout this paper we assume that H_d is finite-dimensional which seems to be a natural choice when it comes to climate modeling. However, our arguments are general and can be adapted to infinite-dimensional subspaces, see [5].

In the case of the more abstract system (6), (7), the process Y^ε will eventually behave like white noise, as $\varepsilon \downarrow 0$. This limiting behavior is fundamental for finding the limit of Eq. (6) because it opens the door for using arguments similar to those of Wong and Zakai in [26]. Of course, Wong and Zakai formulated their results in a finite-dimensional setting. There have been earlier attempts of proving similar results in infinite dimensions; we refer to [2,23,24], for example. However, we would like to emphasize that these earlier attempts dealt with piecewise linear approximations of noise rather than an infinite-dimensional Ornstein–Uhlenbeck process. Note that it is typical for Wong–Zakai results that stochastic integral terms of limiting equations are interpreted in the sense of Stratonovich.

Finally, it is worth comparing our results with those in the literature concerning averaging principles, see, for instance, [8, Sect. 7.9], [20,21] and references therein. Roughly speaking, in those results the unresolved variables satisfy the equation $\mathrm{d}Y_t^\varepsilon = -\varepsilon^2 Y_t^\varepsilon \mathrm{d}t + \varepsilon^{-1}\mathrm{d}W_t$, with a weaker noise intensity compared to our, and therefore, the resolved variables only undergo a change of drift in the limit $\varepsilon \downarrow 0$. On the contrary, in our setting a diffusion term also appears in the limit, see (13) below.

The paper is structured as follows.

In Sect. 2, we formulate our main results on the convergence of solutions to (6), (7). First, the limiting equation for $\bar{X}$ is identified, and then conditions for weak convergence $X^\varepsilon \to \bar{X}$ are stated in Theorem 2.2(i). However, when (6) is a simpler equation, i.e., $\beta = 0$, even the stronger convergence (8) can be shown under the same conditions—see Theorem 2.2(ii).

In Sect. 3, we give the proof of Theorem 2.2(ii). The proof relies on preliminary localization and discretization arguments which allow to consider, instead of (8), its discrete version

$$\lim_{\varepsilon\to 0} \mathbb{P}\left\{\sup_k \|X_{t_k}^\varepsilon - \bar{X}_{t_k}\|_{H_d} > \delta\right\} = 0, \quad \forall \delta > 0,$$

for only finitely many $t_k \in [0, T]$.

In Sect. 4, we give the proof of Theorem 2.2(i) which, at the beginning, requires a careful analysis of the quadratic term $\beta(Y_t^\varepsilon, Y_t^\varepsilon)$, but otherwise is an adaptation of the proof given in the previous section.

In Sect. 5, we eventually use the results of Sect. 2 to prove Theorem 5.3 under quite natural conditions, thus making the connection to our main applications in climate modeling.

2. Notation and main result

Let H_d, H_∞ be real separable Hilbert spaces. Assume that H_d is finite-dimensional, $\dim H_d = d$, with given orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_d$, and that H_∞ is infinite-dimensional with given orthonormal basis $\mathbf{f}_1, \mathbf{f}_2, \ldots$

Given two Banach spaces U, V, let $\mathcal{L}(U, V)$ denote the Banach space of continuous linear operators mapping U to V, endowed with the operator norm.

For each $\varepsilon > 0$, consider the pair of stochastic processes $(X^\varepsilon, Y^\varepsilon)$, taking values in $H_d \times H_\infty$, where X^ε satisfies (6) over a fixed finite time interval $[0, T]$, and Y^ε is given by

$$Y_t^\varepsilon = \int_{-\infty}^t \varepsilon^{-2} e^{-\varepsilon^{-2}(t-s)} \mathrm{d}W_s, \quad t \geq 0,$$

where W is a Wiener process in H_∞, with real-valued time parameter and self-adjoint trace class covariance operator $Q \in \mathcal{L}(H_\infty, H_\infty)$.

Remark 2.1. (i) A Wiener process with real-valued time parameter can be obtained in the following way: given two independent Wiener processes $(W_t^+)_{t\geq 0}$ and $(W_t^-)_{t\geq 0}$ defined on filtered probability spaces $(\Omega^+, (\mathcal{F}_t^+), \mathbb{P}^+)$ and $(\Omega^-, (\mathcal{F}_t^-), \mathbb{P}^-)$, respectively, set $W_t = W_t^+$, for $t \geq 0$, and $W_t = W_{-t}^-$, for $t < 0$.

(ii) Using such a representation of W, we can also write

$$Y_t^\varepsilon = -\int_0^\infty \varepsilon^{-2} e^{-\varepsilon^{-2}(t+s)} \mathrm{d}W_s^- + \int_0^t \varepsilon^{-2} e^{-\varepsilon^{-2}(t-s)} \mathrm{d}W_s^+, \quad t \geq 0,$$

which clearly is a stationary Ornstein–Uhlenbeck process on $(\Omega, \mathcal{F}_\infty^- \otimes \mathcal{F}_\infty^+, \mathbb{P})$, where $\Omega = \Omega^- \times \Omega^+$ and $\mathbb{P} = \mathbb{P}^- \otimes \mathbb{P}^+$, see [3]. Furthermore, setting up the stochastic basis for our processes $(X^\varepsilon, Y^\varepsilon)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be the completion of $(\Omega, \mathcal{F}_\infty^- \otimes \mathcal{F}_\infty^+, \mathbb{P})$, and $(\mathcal{F}_t)_{t\geq 0}$ be the augmentation of the filtration $(\mathcal{F}_\infty^- \otimes \mathcal{F}_t^+)_{t\geq 0}$. Note that this filtration would satisfy the usual conditions.

(iii) Since Q is trace class, both W and Y^ε take values in H_∞. Without loss of generality, we can assume that Q is diagonal with respect to the chosen basis $\{\mathbf{f}_m\}_{m\in\mathbb{N}}$ of H_∞, that the eigenvalues of Q form a sequence $\{q_m\}_{m\in\mathbb{N}}$ satisfying $\sum_m q_m < \infty$, and that $\mathbb{E}\left[\langle W_t, \mathbf{f}_m\rangle_{H_\infty}^2\right] = |t| q_m$, for every $t \geq 0$ and $m \in \mathbb{N}$. Moreover, since

$$\langle Y_t^\varepsilon, \mathbf{f}_m\rangle_{H_\infty} = \int_{-\infty}^t \varepsilon^{-2} e^{-\varepsilon^{-2}(t-s)} d\langle W_s, \mathbf{f}_m\rangle_{H_\infty}$$

we also have $\mathbb{E}\left[\langle Y_t^\varepsilon, \mathbf{f}_m\rangle_{H_\infty}^2\right] = \frac{\varepsilon^{-2}}{2} q_m$ for every $t \geq 0$ and $m \in \mathbb{N}$.

(iv) Let Z be an ε-independent stationary Ornstein–Uhlenbeck process solving $\mathrm{d}Z_t = -Z_t \mathrm{d}t + \mathrm{d}W_t$, which is explicitly given by the formula

$$Z_t = \int_{-\infty}^t e^{-(t-s)} \mathrm{d}W_s, \quad t \geq 0. \tag{9}$$

Due to the self-similarity of W, it is easy to check that the process $(Y_t^\varepsilon)_{t\geq 0}$ equals in law the process $(\varepsilon^{-1}Z_{t\varepsilon^{-2}})_{t\geq 0}$, thus making more transparent why we expect the process Y^ε to behave like a white noise as $\varepsilon \downarrow 0$, see, for instance, [1].

Adopting the useful notation $W_t^\varepsilon = \int_0^t Y_s^\varepsilon \mathrm{d}s$, we can write (6) in integral form as

$$X_t^\varepsilon = x_0 + \int_0^t F(s, X_s^\varepsilon)\mathrm{d}s + \int_0^t \sigma(s, X_s^\varepsilon)\mathrm{d}W_s^\varepsilon + \int_0^t \varepsilon\beta(Y_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s, \quad t \in [0, T], \tag{10}$$

where $x_0 \in H_d$ is a deterministic initial condition, as well as $F : [0, T] \times H_d \to H_d$, $\sigma : [0, T] \times H_d \to \mathcal{L}(H_\infty, H_d)$, $\beta : H_\infty \times H_\infty \to H_d$. We make the following assumptions on these coefficients:

(A1) $F \in C([0, T]\times H_d, H_d)$, and $F(t, \cdot) \in Lip_{loc}(H_d, H_d)$, uniformly in $t \in [0, T]$;

(A2) $\sigma \in C^{1,\gamma}([0, T] \times H_d, \mathcal{L}(H_\infty, H_d))$, the space of C^1 functions with γ-Hölder derivative, for some $\gamma \in (0, 1)$ and its space-differential $D\sigma(t, \cdot) \in Lip_{loc}(H_d, \mathcal{L}(H_d, \mathcal{L}(H_\infty, H_d)))$, uniformly in $t \in [0, T]$;

(A3) $\beta : H_\infty \times H_\infty \to H_d$ is a continuous bilinear map.

Of course, by standard theory (see [3], for example), Eq. (10) admits a unique local strong solution, for each $\varepsilon > 0$.

Next, we introduce the limiting equation for the wanted limit $\bar{X}$ of the processes X^ε, when $\varepsilon \downarrow 0$. First, define the so-called *Stratonovich correction* term $C : [0, T]\times H_d \to H_d$ by

$$C^i(s, x) = \langle C(s, x), \mathbf{e}_i\rangle_{H_d} = \frac{1}{2}\sum_{m\in\mathbb{N}} q_m \sum_{j=1}^d D_j\sigma^{i,m}(s, x)\sigma^{j,m}(s, x), \quad i = 1, \ldots, d, \tag{11}$$

where

$$\sigma^{i,m}(s, x) = \langle \sigma(s, x)\mathbf{f}_m, \mathbf{e}_i\rangle_{H_d}, \quad i = 1, \ldots, d, \ m \in \mathbb{N},$$

is matrix notation for the linear map $\sigma(s, x) \in \mathcal{L}(H_\infty, H_d)$ with respect to our chosen basis vectors; second, let

$$b^i_{\ell,m} = \langle \beta(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d}\sqrt{\frac{q_\ell q_m}{2}}, \quad i = 1, \ldots, d, \ \ell, m \in \mathbb{N}. \tag{12}$$

Then, our limiting equation would read

$$\begin{aligned}\bar{X}_t = x_0 &+ \int_0^t \big(F(s, \bar{X}_s) + C(s, \bar{X}_s)\big)\,\mathrm{d}s \\ &+ \int_0^t \sigma(s, \bar{X}_s)dW_s + \sum_{\ell,m\in\mathbb{N}} b_{\ell,m}\bar{W}_t^{\ell,m}, \quad t \in [0, T],\end{aligned} \tag{13}$$

where W is the same Wiener process used to define Y^ε in Remark 2.1, while $\{\bar{W}^{\ell,m}\}_{\ell,m\in\mathbb{N}}$ is a family of independent one-dimensional standard Wiener processes, which are also independent of W.

As for (10), also (13) admits a unique local strong solution. However, in view of the interpretation of our results with respect to climate modeling, it is natural to further assume that

(A4) both Eqs. (10) and (13) admit global solutions on $[0, T]$.

Another assumption specific to climate modeling, which has been advocated in [17], is the zero-mean property of $\beta(Y^\varepsilon_s, Y^\varepsilon_s)$, $s \geq 0$. Since all Y^ε are stationary under $\mathbb{P}$, see Remark 2.1(ii), this assumption would translate into

$$\mathbb{E}\left[\langle\beta(Y^\varepsilon_s, Y^\varepsilon_s), \mathbf{e}_i\rangle_{H_d}\right] = \sum_{\ell,m\in\mathbb{N}} \langle\beta(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d}\, \mathbb{E}\left[Y^{\varepsilon,\ell}_s Y^{\varepsilon,m}_s\right] = \sum_{\ell\in\mathbb{N}} \langle\beta(\mathbf{f}_\ell, \mathbf{f}_\ell), \mathbf{e}_i\rangle_{H_d}\, \frac{\varepsilon^{-2}}{2} q_\ell = 0,$$

where $Y^{\varepsilon,\ell}_s$ is short notation for the coordinates $\langle Y^\varepsilon_s, \mathbf{f}_\ell\rangle_{H_\infty}$, $\ell = 1, 2, \ldots,$ $s \in [0, T]$. As a consequence, we also impose the zero-mean condition

(A5) $\sum_{\ell\in\mathbb{N}} \langle\beta(\mathbf{f}_\ell, \mathbf{f}_\ell), \mathbf{e}_i\rangle_{H_d}\, q_\ell = 0$, for all $i = 1, \ldots, d$,

which is usually true for equations from fluid dynamics and can in general be understood as a renormalization procedure for the quadratic term.

The following theorem is the main result of this paper.

Theorem 2.2. *(i) Assume (A1)–(A5). Then, X^ε converges to $\bar{X}$, in law, $\varepsilon \downarrow 0$.*
(ii) However, if (A1)–(A4) and (A5) comes via $\beta = 0$, then the stronger convergence (8) holds true.

In what follows, to keep notation light in proofs, when no confusion may occur, the norms in both spaces H_d and H_∞ will be denoted by $|\cdot|$, and their scalar products by $\langle\cdot,\cdot\rangle$. The symbol $\lesssim$ means inequality up to a multiplicative constant, possibly depending on the parameters of our equations, but not on ε.

3. Strong convergence

In this section we give the proof of Theorem 2.2(ii), which is divided into several steps.

First, by localization, we argue that we can restrict ourselves to $|X^\varepsilon_t|, |\bar{X}_t| \leq R$, for some large R, which is effectively leading to Lipschitz continuity of the coefficients of (10).

Second, we discretize the problem, which allows us to reduce the proof of Theorem 2.2(ii) to its discrete version:

$$\lim_{\varepsilon\to 0} \mathbb{P}\left\{\sup_k |X^\varepsilon_{t_k} - \bar{X}_{t_k}| > \delta\right\} = 0, \quad \forall\delta > 0,$$

for only finitely many $t_k \in [0, T]$. Here, we choose $t_k = k\Delta$, where $\Delta = \Delta_\varepsilon$ is a positive parameter whose ε-dependence has to be carefully chosen in the proof—see Remark 3.9.

Third, we prove the above discretized version.

3.1. Localization

Fix $\varepsilon > 0$, $\delta \in (0, 1)$, and define

$$\tau_R^\varepsilon = \inf\{t \geq 0 : |X_t^\varepsilon| \geq R + 1\} \wedge \inf\{t \geq 0 : |\bar{X}_t| \geq R\}, \quad \text{for } R > 0,$$

so that

$$\begin{aligned}
\mathbb{P}\left\{\sup_{t\leq T} |X_t^\varepsilon - \bar{X}_t| > \delta\right\} &= \mathbb{P}\left\{\sup_{t\leq T} |X_t^\varepsilon - \bar{X}_t| > \delta,\ \sup_{t\leq T} |\bar{X}_t| \geq R\right\} \\
&\quad + \mathbb{P}\left\{\sup_{t\leq T} |X_t^\varepsilon - \bar{X}_t| > \delta,\ \sup_{t\leq T} |\bar{X}_t| < R\right\} \\
&= \mathbb{P}\left\{\sup_{t\leq T} |X_t^\varepsilon - \bar{X}_t| > \delta,\ \sup_{t\leq T} |\bar{X}_t| \geq R\right\} \\
&\quad + \mathbb{P}\left\{\sup_{t\leq T\wedge\tau_R^\varepsilon} |X_t^\varepsilon - \bar{X}_t| > \delta,\ \sup_{t\leq T} |\bar{X}_t| < R\right\} \\
&\leq \mathbb{P}\left\{\sup_{t\leq T} |\bar{X}_t| \geq R\right\} + \mathbb{P}\left\{\sup_{t\leq T\wedge\tau_R^\varepsilon} |X_t^\varepsilon - \bar{X}_t| > \delta\right\}.
\end{aligned} \tag{14}$$

Therefore, since (A4) implies

$$\mathbb{P}\left\{\sup_{t\leq T} |\bar{X}_t| \geq R\right\} \to 0, \text{ as } R \uparrow \infty,$$

to prove (8), it is sufficient to show the convergence of the second summand on the right-hand side of (14), when $\varepsilon \downarrow 0$, for fixed $\delta \in (0, 1)$, $R > 0$. Furthermore, by Markov inequality,

$$\mathbb{P}\left\{\sup_{t\leq T\wedge\tau_R^\varepsilon} |X_t^\varepsilon - \bar{X}_t| > \delta\right\} \leq \delta^{-p}\, \mathbb{E}\left[\sup_{t\leq T\wedge\tau_R^\varepsilon} |X_t^\varepsilon - \bar{X}_t|^p\right], \tag{15}$$

for every $p > 0$, $\delta \in (0, 1)$, and hence showing convergence of the above right-hand side, only, is enough. To keep notation light, we are going to use τ^ε instead of τ_R^ε, as $R > 0$ will be fixed, in what follows.

3.2. Discretization

Fix $\varepsilon > 0$. We show that the expectation on the right-hand side of (15) can be replaced by an expectation of the same quantity, but with the supremum taken over a finite number (diverging to ∞, as $\varepsilon \downarrow 0$) of times t_k, see Corollary 3.7 below.

To start with, we have the following useful a priori estimate.

Lemma 3.1. *For any $p > 1$, the Ornstein–Uhlenbeck process Y^ε satisfies*

$$\mathbb{E}\left[\sup_{t\le T}\left|Y_t^\varepsilon\right|^p\right] \lesssim \varepsilon^{-p}\log^{p/2}\left(1+\varepsilon^{-2}\right).$$

Proof. First, using the decomposition $Y_t^\varepsilon = Y_0^\varepsilon + \left(Y_t^\varepsilon - Y_0^\varepsilon\right)$, Gaussian estimates on Y_0^ε and [15, Theorem 2.2], the result is true in one dimension.

In the infinite-dimensional case, by Hölder's inequality, we can suppose $p > 2$. Therefore, since Q is trace class with eigenvalues satisfying $\sum_{m\in\mathbb{N}} q_m < \infty$, when $\alpha = (p-2)/p$, we obtain that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\le T}\left|Y_t^\varepsilon\right|^p\right] &= \mathbb{E}\left[\sup_{t\le T}\left(\sum_{m\in\mathbb{N},q_m>0} q_m^\alpha q_m^{-\alpha}\left|Y_t^{\varepsilon,m}\right|^2\right)^{p/2}\right]\\ &\lesssim \left(\sum_{m\in\mathbb{N},q_m>0} q_m^{-\alpha p/2}\mathbb{E}\left[\sup_{t\le T}\left|Y_t^{\varepsilon,m}\right|^p\right]\right)\left(\sum_{m\in\mathbb{N}} q_m^{\alpha p/(p-2)}\right)^{(p-2)/2}\\ &\lesssim \varepsilon^{-p}\log^{p/2}\left(1+\varepsilon^{-2}\right),\end{aligned}$$

having used the one-dimensional result for the coordinates $Y_t^{\varepsilon,m} = \langle Y_t^\varepsilon, \mathbf{f}_m\rangle$, $m = 1, 2, \ldots$ □

Remark 3.2. In view of Remark 2.1(iv), the previous result could also be obtained from the analogous result for (9) and parabolic scaling. Indeed, it would be sufficient to prove $\mathbb{E}\left[\sup_{t\le T}|Z_t|^p\right] \lesssim \log^{p/2}(1+T)$ for every $p > 1$.

Now, we introduce the discretization of the time interval $[0, T]$. Let $\Delta > 0$, and let $[T/\Delta]$ be the largest integer less or equal than T/Δ. In what follows, Δ will also depend on ε, in a way to be determined later. Also, to make it easier to bound terms by powers of ε or Δ, without loss of generality, we will always assume that both ε and Δ are less than one.

The next two lemmas control the excursion of X^ε between adjacent nodes in terms of the ratio Δ/ε.

Lemma 3.3. *For any $p > 1$, and any deterministic time $\tau > 0$,*

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\ldots,[T/\Delta]\\ t\le\tau,\, t+k\Delta\le T\wedge\tau^\varepsilon}} |X_{t+k\Delta}^\varepsilon - X_{k\Delta}^\varepsilon|^p\right] \lesssim \left(\frac{\tau}{\varepsilon}\right)^p \log^{p/2}(1+\varepsilon^{-2}).$$

Proof. Since $\beta = 0$, by (10), the increment $X^\varepsilon_{t+k\Delta} - X^\varepsilon_{k\Delta}$ can be written as

$$X^\varepsilon_{t+k\Delta} - X^\varepsilon_{k\Delta} = \int_{k\Delta}^{t+k\Delta} F(s, X^\varepsilon_s)\mathrm{d}s + \int_{k\Delta}^{t+k\Delta} \sigma(s, X^\varepsilon_s)\mathrm{d}W^\varepsilon_s, \quad \text{for} t + k\Delta \le T \wedge \tau^\varepsilon.$$

Therefore, using (A1), (A2), boundedness of X^ε on $[0, \tau^\varepsilon]$, and Lemma 3.1, we obtain that

$$\begin{aligned}\mathbb{E}\left[\sup_{\substack{k=0,1,\ldots,[T/\Delta]\\ t\le\tau,\, t+k\Delta\le T\wedge\tau^\varepsilon}} |X^\varepsilon_{t+k\Delta} - X^\varepsilon_{k\Delta}|^p\right] &\lesssim \tau^p\left(1 + \mathbb{E}\left[\sup_{t\le T\wedge\tau^\varepsilon}\left|Y^\varepsilon_t\right|^p\right]\right)\\ &\lesssim \left(\frac{\tau}{\varepsilon}\right)^p \log^{p/2}(1+\varepsilon^{-2}),\end{aligned}$$

where $W^\varepsilon_t = \int_0^t Y^\varepsilon_s \mathrm{d}s$ was defined in Sect. 2. □

Lemma 3.4. *For any $p > 1$, and any fixed $k \in \{0, 1, \ldots, [T/\Delta]\}$ such that $k\Delta \le T$,*

$$\mathbb{E}\left[|X^\varepsilon_{(k+1)\Delta\wedge\tau^\varepsilon} - X^\varepsilon_{k\Delta\wedge\tau^\varepsilon}|^p\right] \lesssim \Delta^{p/2} + \varepsilon^p \log^{p/2}(1+\varepsilon^{-2}) + \left(\frac{\Delta}{\varepsilon}\right)^{2p} \log^p(1+\varepsilon^{-2}).$$

Proof. It suffices to bound every single term on the right-hand side of the equation

$$\begin{aligned}X^\varepsilon_{(k+1)\Delta\wedge\tau^\varepsilon} - X^\varepsilon_{k\Delta\wedge\tau^\varepsilon} &= \int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} F(s, X^\varepsilon_s)\mathrm{d}s\\ &\quad + \int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} \left(\sigma(s, X^\varepsilon_s) - \sigma(k\Delta\wedge\tau^\varepsilon, X^\varepsilon_{k\Delta\wedge\tau^\varepsilon})\right)\mathrm{d}W^\varepsilon_s\\ &\quad + \int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} \sigma(k\Delta\wedge\tau^\varepsilon, X^\varepsilon_{k\Delta\wedge\tau^\varepsilon})\mathrm{d}W^\varepsilon_s.\end{aligned}$$

First, by (A1) and boundedness of X^ε on $[0, \tau^\varepsilon]$, we have that

$$\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} F(s, X^\varepsilon_s)\mathrm{d}s\right|^p\right] \lesssim \Delta^p.$$

Second, using Hölder's inequality with $q' > 1/p$ and Lemma 3.1,

$$\begin{aligned}&\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} \left(\sigma(s, X^\varepsilon_s) - \sigma(k\Delta\wedge\tau^\varepsilon, X^\varepsilon_{k\Delta\wedge\tau^\varepsilon})\right)\mathrm{d}W^\varepsilon_s\right|^p\right]\\ &\lesssim \mathbb{E}\left[\sup_{t\le T}\left|Y^\varepsilon_t\right|^p \left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} \left|\sigma(s, X^\varepsilon_s) - \sigma(k\Delta\wedge\tau^\varepsilon, X^\varepsilon_{k\Delta\wedge\tau^\varepsilon})\right|\mathrm{d}s\right|^p\right]\\ &\lesssim \varepsilon^{-p}\log^{p/2}(1+\varepsilon^{-2})\left(\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon} \left|\sigma(s, X^\varepsilon_s) - \sigma(k\Delta\wedge\tau^\varepsilon, X^\varepsilon_{k\Delta\wedge\tau^\varepsilon})\right|\mathrm{d}s\right|^{pq'}\right]\right)^{1/q'}.\end{aligned}$$

Since $pq' > 1$ by assumption, we can estimate the integral above using Hölder's inequality with exponents pq' and $pq'/(pq'-1)$, (A2) and Lemma 3.3 to obtain

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\left|\sigma(s,X_s^\varepsilon)-\sigma(k\Delta\wedge\tau^\varepsilon,X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon)\right|\mathrm{d}s\right|^{pq'}\right]\right)^{1/q'}\\
&\lesssim\left(\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\mathrm{d}s\right|^{pq'-1}\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\left|\sigma(s,X_s^\varepsilon)-\sigma(k\Delta\wedge\tau^\varepsilon,X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon)\right|^{pq'}\mathrm{d}s\right]\right)^{1/q'}\\
&\lesssim\Delta^{p-1/q'}\left(\mathbb{E}\left[\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\left(\left|X_s^\varepsilon-X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^{pq'}+(s-k\Delta)^{pq'}\right)\mathrm{d}s\right]\right)^{1/q'}\\
&\lesssim\Delta^{p-1/q'}\left(\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\mathbb{E}\left[\left|X_s^\varepsilon-X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^{pq'}+(s-k\Delta)^{pq'}\right]\mathrm{d}s\right)^{1/q'}\\
&\lesssim\Delta^{p-1/q'}\left(\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}(s-k\Delta)^{pq'}\left(\varepsilon^{-pq'}\log^{pq'/2}(1+\varepsilon^{-2})+1\right)\mathrm{d}s\right)^{1/q'}\\
&\lesssim\varepsilon^{-p}\log^{p/2}(1+\varepsilon^{-2})\Delta^{p-1/q'}\left(\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}(s-k\Delta)^{pq'}\mathrm{d}s\right)^{1/q'}\lesssim\varepsilon^{-p}\log^{p/2}(1+\varepsilon^{-2})\Delta^{2p}.
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\sigma(k\Delta\wedge\tau^\varepsilon,X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon)\mathrm{d}W_s^\varepsilon\right|^p\right]\lesssim\mathbb{E}\left[\left|W_{(k+1)\Delta\wedge\tau^\varepsilon}^\varepsilon-W_{k\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^p\right]\\
&\lesssim\Delta^{p/2}+\varepsilon^p\log^{p/2}(1+\varepsilon^{-2}),
\end{aligned}$$

because, for every $t_2>t_1\geq 0$,

$$\begin{aligned}
W_{t_2}^\varepsilon-W_{t_1}^\varepsilon&=\int_{t_1}^{t_2}\left(\int_{-\infty}^{s}\varepsilon^{-2}e^{-\varepsilon^{-2}(s-r)}\mathrm{d}W_r\right)\mathrm{d}s\\
&=W_{t_2}-W_{t_1}-\int_{-\infty}^{t_2}e^{-\varepsilon^{-2}(t_2-r)}\mathrm{d}W_r+\int_{-\infty}^{t_1}e^{-\varepsilon^{-2}(t_1-r)}\mathrm{d}W_r. \qquad (16)
\end{aligned}$$

□

The next lemma controls the excursion of the limiting process $\bar{X}$ between adjacent nodes.

Lemma 3.5. *For any $p>1$, any deterministic time $\tau\in(0,1)$, and any fixed $k\in\{0,1,\ldots,[T/\Delta]\}$,*

$$\mathbb{E}\left[\sup_{t\leq\tau,\,t+k\Delta\leq T\wedge\tau^\varepsilon}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^p\right]\lesssim\tau^{\frac{p}{2}}.$$

Proof. Since $\beta=0$, by (13), the increment $\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}$ can be written as

$$\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}=\int_{k\Delta}^{t+k\Delta}\left(F(s,\bar{X}_s)+C(s,\bar{X}_s)\right)\mathrm{d}s$$

$$+\int_{k\Delta}^{t+k\Delta}\sigma(s,\bar{X}_s)\mathrm{d}W_s,\quad \text{for} t+k\Delta\le T\wedge\tau^\varepsilon.$$

Therefore, using (A1), (A2), boundedness of X^ε on $[0,\tau^\varepsilon]$, and Burkholder–Davis–Gundy's inequality, we obtain that

$$\mathbb{E}\left[\sup_{t\le\tau,\, t+k\Delta\le T\wedge\tau^\varepsilon}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^p\right]\lesssim\tau^p$$
$$+\mathbb{E}\left[\sup_{t\le\tau,\, t+k\Delta\le T\wedge\tau^\varepsilon}\left|\int_{k\Delta}^{t+k\Delta}\sigma(s,\bar{X}_s)\mathrm{d}W_s\right|^p\right]\lesssim\tau^p+\tau^{\frac{p}{2}},$$

which proves the lemma since $\tau<1$. □

Corollary 3.6. *For any $p>1$,*

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\dots,[T/\Delta]\\ t\le\Delta,\, t+k\Delta\le T\wedge\tau^\varepsilon}}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^p\right]\lesssim\Delta^{\frac{p}{2}-1}.$$

Proof. The claim easily follows from Lemma 3.5 with $\tau=\Delta$, and the inequality

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\dots,[T/\Delta]\\ t\le\Delta,\, t+k\Delta\le T\wedge\tau^\varepsilon}}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^p\right]\lesssim\sum_{k=0}^{[T/\Delta]}\mathbb{E}\left[\sup_{t\le\Delta,\, t+k\Delta\le T\wedge\tau^\varepsilon}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^p\right]$$
$$\lesssim\sum_{k=0}^{[T/\Delta]}\Delta^{p/2}=\Delta^{\frac{p}{2}-1}.$$

□

Corollary 3.7. *Let $\Delta=\Delta_\varepsilon>0$ depend on ε such that $\Delta/\varepsilon\to 0$, as $\varepsilon\downarrow 0$. Then,*

$$\mathbb{E}\left[\sup_{t\le T\wedge\tau^\varepsilon}|X_t^\varepsilon-\bar{X}_t|^2\right]\lesssim\mathbb{E}\left[\sup_{\substack{k=0,1,\dots,[T/\Delta]\\ k\Delta\le\tau^\varepsilon}}|X_{k\Delta}^\varepsilon-\bar{X}_{k\Delta}|^2\right]+o(1).$$

Proof. First, by Hölder's inequality with $q>1$ and Lemma 3.6, we have that

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\dots,[T/\Delta]\\ t\le\Delta,\, t+k\Delta\le T\wedge\tau^\varepsilon}}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^2\right]\lesssim\left(\mathbb{E}\left[\sup_{\substack{k=0,1,\dots,[T/\Delta]\\ t\le\Delta,\, t+k\Delta\le T\wedge\tau^\varepsilon}}|\bar{X}_{t+k\Delta}-\bar{X}_{k\Delta}|^{2q}\right]\right)^{1/q}$$
$$\lesssim\Delta^{1-1/q}\to 0\text{ as }\varepsilon\downarrow 0,$$

since we have taken $q>1$. Thus, the proof can easily be completed by combining the above and Lemma 3.3, while taking into account

$$|X_t^\varepsilon-\bar{X}_t|^2\lesssim|X_t^\varepsilon-X_{[t/\Delta]\Delta}^\varepsilon|^2+|X_{[t/\Delta]\Delta}^\varepsilon-\bar{X}_{[t/\Delta]\Delta}|^2+|\bar{X}_{[t/\Delta]\Delta}-\bar{X}_t|^2,$$

where $[t/\Delta]$ is again our notation for the floor of t/Δ. □

3.3. Proof of the discretized version

We now discuss our strategy to prove part (ii) of Theorem 2.2. Recall that we want

$$\mathbb{P}\left\{\sup_{t\le T}|X_t^\varepsilon - \bar{X}_t| > \delta\right\} \to 0,$$

for every fixed $\delta > 0$, as $\varepsilon \downarrow 0$. As we have seen, by (14), (15) and Corollary 3.7, it suffices to prove

$$\mathbb{E}\left[\sup_{\substack{k=0,\dots,[T/\Delta]\\ k\Delta\le\tau^\varepsilon}}\left|X_{k\Delta}^\varepsilon - \bar{X}_{k\Delta}\right|^2\right] \to 0, \quad \varepsilon \downarrow 0, \tag{17}$$

for some $\Delta = \Delta_\varepsilon = o(\varepsilon)$. The proof is inspired by [11, Sect. VI.7].

Hereafter, $\partial\sigma$ denotes the derivative of σ with respect its first variable, and $D\sigma$ denotes the derivative of σ with respect its second variable. To start with, by (10) without β-term, (A2), and (16), we have that

$$\begin{aligned}
X_{(k+1)\Delta}^\varepsilon &= X_{k\Delta}^\varepsilon + \int_{k\Delta}^{(k+1)\Delta} F(s, X_s^\varepsilon)\mathrm{d}s + \int_{k\Delta}^{(k+1)\Delta} \sigma(s, X_s^\varepsilon)\mathrm{d}W_s^\varepsilon \\
&= X_{k\Delta}^\varepsilon + \int_{k\Delta}^{(k+1)\Delta} \left(F(s, X_s^\varepsilon) - F(k\Delta, X_{k\Delta}^\varepsilon)\right)\mathrm{d}s \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} F(k\Delta, X_{k\Delta}^\varepsilon)\mathrm{d}s \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \left(\sigma(s, X_s^\varepsilon) - \sigma(k\Delta, X_{k\Delta}^\varepsilon)\right)\mathrm{d}W_s^\varepsilon + \int_{k\Delta}^{(k+1)\Delta} \sigma(k\Delta, X_{k\Delta}^\varepsilon)\mathrm{d}W_s^\varepsilon \\
&= X_{k\Delta}^\varepsilon + \int_{k\Delta}^{(k+1)\Delta} \left(F(s, X_s^\varepsilon) - F(k\Delta, X_{k\Delta}^\varepsilon)\right)\mathrm{d}s \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} F(k\Delta, X_{k\Delta}^\varepsilon)\mathrm{d}s \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \left(\int_{k\Delta}^{s} \left(\partial\sigma(r, X_r^\varepsilon) + D\sigma(r, X_r^\varepsilon)F(r, X_r^\varepsilon)\right)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \left(\int_{k\Delta}^{s} \left(D\sigma(r, X_r^\varepsilon)\sigma(r, X_r^\varepsilon) - D\sigma(k\Delta, X_{k\Delta}^\varepsilon)\sigma(k\Delta, X_{k\Delta}^\varepsilon)\right)\mathrm{d}W_r^\varepsilon\right)\mathrm{d}W_s^\varepsilon \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \left(\int_{k\Delta}^{s} \left(D\sigma(k\Delta, X_{k\Delta}^\varepsilon)\sigma(k\Delta, X_{k\Delta}^\varepsilon) - D\sigma(k\Delta, \bar{X}_{k\Delta})\sigma(k\Delta, \bar{X}_{k\Delta})\right)\mathrm{d}W_r^\varepsilon\right)\mathrm{d}W_s^\varepsilon \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \left(\int_{k\Delta}^{s} D\sigma(k\Delta, \bar{X}_{k\Delta})\sigma(k\Delta, \bar{X}_{k\Delta})\mathrm{d}W_r^\varepsilon\right)\mathrm{d}W_s^\varepsilon \\
&\quad + \int_{k\Delta}^{(k+1)\Delta} \sigma(k\Delta, X_{k\Delta}^\varepsilon)\mathrm{d}W_s \\
&\quad + \sigma(k\Delta, X_{k\Delta}^\varepsilon)\varepsilon^2\left(Y_{k\Delta}^\varepsilon - Y_{(k+1)\Delta}^\varepsilon\right) \\
&= X_{k\Delta}^\varepsilon + I_1^k + I_2^k + I_3^k + I_4^k + I_5^k + I_6^k + I_7^k + I_8^k,
\end{aligned} \tag{18}$$

for any $k = 0, \dots, [T/\Delta]$ such that $(k+1)\Delta \le T$.

Similarly, using (13) instead of (10), the process $\bar{X}$ satisfies

$$\begin{aligned}\bar{X}_{(k+1)\Delta} &= \bar{X}_{k\Delta} + \int_{k\Delta}^{(k+1)\Delta} \big(F(s,\bar{X}_s) - F(k\Delta,\bar{X}_{k\Delta})\big)\,\mathrm{d}s \\ &\quad + \int_{k\Delta}^{(k+1)\Delta} F(k\Delta,\bar{X}_{k\Delta})\mathrm{d}s \\ &\quad + \int_{k\Delta}^{(k+1)\Delta} \big(C(s,\bar{X}_s) - C(k\Delta,\bar{X}_{k\Delta})\big)\,\mathrm{d}s \\ &\quad + \int_{k\Delta}^{(k+1)\Delta} C(k\Delta,\bar{X}_{k\Delta})\mathrm{d}s \\ &\quad + \int_{k\Delta}^{(k+1)\Delta} \big(\sigma(s,\bar{X}_s) - \sigma(k\Delta,\bar{X}_{k\Delta})\big)\,\mathrm{d}W_s + \int_{k\Delta}^{(k+1)\Delta} \sigma(k\Delta,\bar{X}_{k\Delta})\mathrm{d}W_s \\ &= \bar{X}_{k\Delta} + J_1^k + J_2^k + J_3^k + J_4^k + J_5^k + J_6^k. \end{aligned} \tag{19}$$

Having in mind to apply Gronwall's lemma, it turns out to be useful to summarize the contributions of the right-hand sides of (18), (19) as follows:

$$\begin{aligned} X^\varepsilon_{h\Delta} - \bar{X}_{h\Delta} &= \sum_{k=0}^{h-1}\Big(I_2^k - J_2^k\Big) + \sum_{k=0}^{h-1}\Big(I_6^k - J_4^k\Big) + \sum_{k=0}^{h-1}\Big(I_7^k - J_6^k\Big) + \sum_{k=0}^{h-1} I_5^k \\ &\quad + \sum_{k=0}^{h-1}\Big(I_1^k + I_3^k + I_4^k + I_8^k - J_1^k - J_3^k - J_5^k\Big), \end{aligned} \tag{20}$$

for any $h = 1, \ldots, [T/\Delta]$, which splits the difference $X^\varepsilon_{h\Delta} - \bar{X}_{h\Delta}$ into 5 sums.

We at first prove that the 2nd and the 5th sum can be neglected when proving (17). The summands of the 5th sum are discussed in Lemma 3.8 below. The contribution of the 2nd sum though is more delicate and requires a martingale argument similar to that of [11, Theorem VI.7.1].

The remaining sums will be controlled in terms of the difference $X^\varepsilon - \bar{X}$ itself, which allows them to be estimated via Gronwall's lemma.

Of course, under assumption (A1), the function F is uniformly continuous when restricted to $[0, T] \times B_R(0)$, where $B_R(0)$ is the closed ball of radius R in H_d. In what follows, we will denote by $\omega_F : [0, T] \to [0, \infty)$ the (local) modulus of continuity of $F(\cdot, x)$:

$$\big|F(t,x) - F(s,x)\big| \le \omega_F(|t-s|), \quad \text{for every } t, s \in [0,T], \text{ and } x \in B_R(0).$$

Obviously, the function ω_F vanishes at zero, and without loss of generality, it can be chosen to be both non-decreasing and continuous.

Denote by ω_σ the corresponding modulus of continuity of the derivative $D\sigma(\cdot, x)$, and let $\omega_{F,\sigma} = \omega_F + \omega_\sigma$. Recall that, under assumption (A2), one can take $\omega_\sigma(t) = Ct^\gamma$ for some positive constant C and $\gamma \in (0, 1)$.

Lemma 3.8. *For any $p > 1$:*

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_1^k\right|^p + \left|\sum_{k=0}^{h-1} I_3^k\right|^p\right] \lesssim \left(\frac{\Delta}{\varepsilon}\right)^p \log^{p/2}(1+\varepsilon^{-2}) + \omega_F(\Delta)^p;$$

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_4^k\right|^p\right] \lesssim \left(\frac{\Delta^2}{\varepsilon^3}\right)^p \log^{3p/2}(1+\varepsilon^{-2})$$
$$+\left(\frac{\Delta}{\varepsilon^2}\right)^p \log^p(1+\varepsilon^{-2})\,\omega_\sigma(\Delta)^p;$$

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_8^k\right|^p\right] \lesssim \left(\frac{\varepsilon^2}{\Delta}\right)^{p/2} \log^{p/2}(1+\varepsilon^{-2}) + \left(\frac{\varepsilon^2}{\Delta}\right)^p \log^p(1+\varepsilon^{-2})$$
$$+\left(\frac{\Delta}{\varepsilon}\right)^p \log^p(1+\varepsilon^{-2});$$

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} J_1^k\right|^p + \left|\sum_{k=0}^{h-1} J_3^k\right|^p + \left|\sum_{k=0}^{h-1} J_5^k\right|^p\right] \lesssim \Delta^{p/2} + \omega_{F,\sigma}(\Delta)^p.$$

Proof. Throughout this proof, we will frequently make use of (A1), (A2) without explicit mentioning.

For $\sum I_1^k$, by Hölder's inequality and Lemma 3.3,

$$\begin{aligned}\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_1^k\right|^p\right] &\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\Big(\left|X_s^\varepsilon - X_{k\Delta}^\varepsilon\right|\right.\right.\\ &\qquad\left.\left.+\omega_F(s-k\Delta)\Big)\,\mathrm{d}s\right|^p\right]\\ &\lesssim \sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\left[\left|X_{s\wedge\tau^\varepsilon}^\varepsilon - X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^p + \omega_F(\Delta)^p\right]\mathrm{d}s\\ &\lesssim \left(\frac{\Delta}{\varepsilon}\right)^p \log^{p/2}(1+\varepsilon^{-2}) + \omega_F(\Delta)^p.\end{aligned}$$

For $\sum I_3^k$, by Hölder's inequality and Lemma 3.1,

$$\begin{aligned}\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_3^k\right|^p\right] &\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sup_{t\le T}\left|Y_t^\varepsilon\right|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}(s-k\Delta)\,\mathrm{d}s\right|^p\right]\\ &\lesssim \mathbb{E}\left[\sup_{t\le T}\left|Y_t^\varepsilon\right|^p\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}\left|s-k\Delta\right|^p\mathrm{d}s\right]\\ &\lesssim \left(\frac{\Delta}{\varepsilon}\right)^p \log^{p/2}(1+\varepsilon^{-2}).\end{aligned}$$

For $\sum I_4^k$, by Hölder's inequality, Lemmas 3.1 and 3.3,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_4^k\right|^p\right]\\
&\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sup_{t\le T}\left|Y_t^\varepsilon\right|^2\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\left(\left|X_r^\varepsilon - X_{k\Delta}^\varepsilon\right| + \omega_\sigma(r-k\Delta)\right)\mathrm{d}r\right)\mathrm{d}s\right|^p\right]\\
&\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\sup_{t\le T}\left|Y_t^\varepsilon\right|^{2p}\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left|\int_{k\Delta}^{s}\left(\left|X_r^\varepsilon - X_{k\Delta}^\varepsilon\right| + \omega_\sigma(r-k\Delta)\right)\mathrm{d}r\right|^p\mathrm{d}s\right]\\
&\lesssim \varepsilon^{-2p}\log^p(1+\varepsilon^{-2})\left(\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}(s-k\Delta)^{pq'-1}\right.\\
&\qquad\left.\int_{k\Delta}^{s}\left(\mathbb{E}\left[\left|X_{r\wedge\tau^\varepsilon}^\varepsilon - X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^{pq'} + \omega_\sigma(\Delta)^{pq'}\right]\mathrm{d}r\right)\mathrm{d}s\right)^{1/q'}\\
&\lesssim \varepsilon^{-3p}\log^{3p/2}(1+\varepsilon^{-2})\left(\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}(s-k\Delta)^{2pq'}\mathrm{d}s\right)^{1/q'}\\
&\quad+\left(\frac{\Delta}{\varepsilon^2}\right)^p\log^p(1+\varepsilon^{-2})\,\omega_\sigma(\Delta)^p\\
&\lesssim \left(\frac{\Delta^2}{\varepsilon^3}\right)^p\log^{3p/2}(1+\varepsilon^{-2})\\
&\quad+\left(\frac{\Delta}{\varepsilon^2}\right)^p\log^p(1+\varepsilon^{-2})\,\omega_\sigma(\Delta)^p.
\end{aligned}
$$

We now consider $\sum I_8^k$. Here, the idea is to convert Y^ε-increments into X^ε-increments via integration by parts since X^ε-increments are easier to control. This way, applying Lemmas 3.1 and 3.4,

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1} I_8^k\right|^p\right] &\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}\sigma(k\Delta, X_{k\Delta}^\varepsilon)\varepsilon^2\left(Y_{k\Delta}^\varepsilon - Y_{(k+1)\Delta}^\varepsilon\right)\right|^p\right]\\
&\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=1}^{h}\left(\sigma(k\Delta, X_{k\Delta}^\varepsilon) - \sigma((k-1)\Delta, X_{(k-1)\Delta}^\varepsilon)\right)\varepsilon^2 Y_{k\Delta}^\varepsilon\right|^p\right]\\
&\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\sup_{t\le T}\left|\varepsilon^2 Y_t^\varepsilon\right|^p\left|\sum_{k=1}^{h}\left(\left|X_{k\Delta}^\varepsilon - X_{(k-1)\Delta}^\varepsilon\right| + \Delta\right)\right|^p\right]
\end{aligned}
$$

$$\lesssim \mathbb{E}\left[\sup_{t\le T}\left|\varepsilon^2 Y_t^\varepsilon\right|^{pq}\right]^{1/q}\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=1}^{h}\left(\left|X_{k\Delta}^\varepsilon - X_{(k-1)\Delta}^\varepsilon\right| + \Delta\right)\right|^{pq'}\right]^{1/q'}$$

$$\lesssim \varepsilon^p \log^{p/2}(1+\varepsilon^{-2})\Delta^{1/q'-p}\left(\sum_{k=1}^{[T/\Delta]}\mathbb{E}\left[\left|X_{k\Delta\wedge\tau^\varepsilon}^\varepsilon - X_{(k-1)\Delta\wedge\tau^\varepsilon}^\varepsilon\right|^{pq'} + \Delta^{pq'}\right]\right)^{1/q'}$$

$$\lesssim \varepsilon^p \log^{p/2}(1+\varepsilon^{-2})\Delta^{-p}\left(\Delta^{pq'/2} + \varepsilon^{pq'}\log^{pq'/2}(1+\varepsilon^{-2}) + \left(\frac{\Delta}{\varepsilon}\right)^{2pq'}\right.$$
$$\left.\log^{pq'}(1+\varepsilon^{-2})\right)^{1/q'}$$

$$\lesssim \left(\frac{\varepsilon^2}{\Delta}\right)^{p/2}\log^{p/2}(1+\varepsilon^{-2}) + \left(\frac{\varepsilon^2}{\Delta}\right)^{p}\log^{p}(1+\varepsilon^{-2}) + \left(\frac{\Delta}{\varepsilon}\right)^{p}\log^{p}(1+\varepsilon^{-2}).$$

In a similar way, for $\sum J_1^k$ and $\sum J_3^k$, now applying Lemma 3.5,

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}J_1^k\right|^p + \left|\sum_{k=0}^{h-1}J_3^k\right|^p\right]$$
$$\lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\left|\bar X_s - \bar X_{k\Delta}\right| + \omega_{F,\sigma}(s-k\Delta)\right)\mathrm{d}s\right|^p\right]$$
$$\lesssim \sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\left[\left|\bar X_{s\wedge\tau^\varepsilon} - \bar X_{k\Delta\wedge\tau^\varepsilon}\right|^p + \omega_{F,\sigma}(\Delta)^p\right]\mathrm{d}s$$
$$\lesssim \Delta^{p/2} + \omega_{F,\sigma}(\Delta)^p.$$

For the last sum $\sum J_5^k$, by Burkholder–Davis–Gundy's inequality and Lemma 3.5,

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}J_5^k\right|^p\right] \lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\sigma(s,\bar X_s) - \sigma(k\Delta,\bar X_{k\Delta})\right)\mathrm{d}W_s\right|^p\right]$$
$$\lesssim \mathbb{E}\left[\left|\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\left|\sigma(s,\bar X_s) - \sigma(k\Delta,\bar X_{k\Delta})\right|^2\mathrm{d}s\right|^{p/2}\right]$$
$$\lesssim \mathbb{E}\left[\left|\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta\wedge\tau^\varepsilon}^{(k+1)\Delta\wedge\tau^\varepsilon}\left|\sigma(s,\bar X_s) - \sigma(k\Delta,\bar X_{k\Delta})\right|^2\mathrm{d}s\right|^{p}\right]^{1/2}$$
$$\lesssim \left(\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}\mathbb{E}\left[\left|\bar X_{s\wedge\tau^\varepsilon} - \bar X_{k\Delta\wedge\tau^\varepsilon}\right|^{2p} + (s-k\Delta)^{2p}\right]\mathrm{d}s\right)^{1/2} \lesssim \Delta^{p/2}.$$

□

Remark 3.9. The estimates given in Lemma 3.8 motivate the following choice of how $\Delta = \Delta_\varepsilon$ should behave when ε goes to zero:

$$\frac{\Delta^2}{\varepsilon^3}\log^{3/2}(1+\varepsilon^{-2}) \to 0, \quad \frac{\Delta}{\varepsilon^2}\log(1+\varepsilon^{-2})\,\omega_\sigma(\Delta) \to 0, \quad \frac{\varepsilon^2}{\Delta}\log^{1/2}(1+\varepsilon^{-2}) \to 0.$$

Such a choice is always possible. Indeed, under assumption (A2), one can take $\omega_\sigma(t) = Ct^\gamma$ for some positive constant C and $\gamma \in (0, 2/3)$, and therefore the choice $\Delta_\varepsilon = \varepsilon^{\frac{2}{1+\gamma/2}}$ satisfies all the requirements above. We will maintain this choice of Δ in the remainder of the paper.

We now discuss the 2nd sum on the right-hand side of (20), that is

$$\sum_{k=0}^{h-1}\left(\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} D\sigma(k\Delta, \bar{X}_{k\Delta})\sigma(k\Delta, \bar{X}_{k\Delta})\mathrm{d}W_r^\varepsilon\right)\mathrm{d}W_s^\varepsilon - \int_{k\Delta}^{(k+1)\Delta} C(k\Delta, \bar{X}_{k\Delta})\mathrm{d}s\right),$$

the i-th component of which, when plugging in (11), reads

$$\sum_{k=0}^{h-1}\sum_{\ell,m\in\mathbb{N}}\sum_{j=1,\dots,d} D_j\sigma^{i,m}(k\Delta, \bar{X}_{k\Delta})\sigma^{j,\ell}(k\Delta, \bar{X}_{k\Delta})\left(c_{\ell,m}^k(\Delta,\varepsilon) - \delta_{\ell,m}\frac{q_m}{2}\Delta\right),$$

where $c_{\ell,m}^k(\Delta,\varepsilon)$ is given by

$$c_{\ell,m}^k(\Delta,\varepsilon) = \int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\mathrm{d}W_r^{\varepsilon,\ell}\right)\mathrm{d}W_s^{\varepsilon,m}.$$

Taking the conditional expectation of $c_{\ell,m}^k(\Delta,\varepsilon)$ with respect to $\mathcal{F}_{k\Delta}$ yields

$$\begin{aligned}\mathbb{E}\left[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] &= \int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\mathbb{E}\left[Y_r^{\varepsilon,\ell}Y_s^{\varepsilon,m} \mid \mathcal{F}_{k\Delta}\right]\mathrm{d}r\right)\mathrm{d}s \\ &= Y_{k\Delta}^{\varepsilon,\ell}Y_{k\Delta}^{\varepsilon,m}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} e^{-\varepsilon^{-2}(r+s-2k\Delta)}\mathrm{d}r\right)\mathrm{d}s \\ &\quad + \delta_{\ell,m}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} q_\ell\frac{\varepsilon^{-2}}{2}\left(e^{-\varepsilon^{-2}(s-r)} - e^{-\varepsilon^{-2}(r+s-2k\Delta)}\right)\mathrm{d}r\right)\mathrm{d}s,\end{aligned}$$

where the following representation of Y^ε,

$$Y_s^{\varepsilon,m} = Y_{k\Delta}^{\varepsilon,m}e^{-\varepsilon^{-2}(s-k\Delta)} + \int_{k\Delta}^{s} e^{-\varepsilon^{-2}(s-r)}\varepsilon^{-2}\mathrm{d}W_r^m,$$

has been used, and this conditional expectation can easily be calculated as

$$\begin{aligned}\mathbb{E}\left[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] &= \frac{\varepsilon^4}{2}Y_{k\Delta}^{\varepsilon,\ell}Y_{k\Delta}^{\varepsilon,m}\left(e^{-\varepsilon^{-2}\Delta} - 1\right)^2 \\ &\quad + \delta_{\ell,m}\frac{q_m}{2}\left(\Delta + \varepsilon^2\left(-\frac{3}{2} + 2e^{-\varepsilon^{-2}\Delta} - \frac{1}{2}e^{-2\varepsilon^{-2}\Delta}\right)\right).\end{aligned} \tag{21}$$

Now, since $\sum_{j=1,\dots,d} D_j\sigma^{i,m}(k\Delta, \bar{X}_{\tau^\varepsilon\wedge(k\Delta)})\sigma^{j,\ell}(k\Delta, \bar{X}_{\tau^\varepsilon\wedge(k\Delta)})$ is $\mathcal{F}_{k\Delta}$ measurable, for every $\ell, m \in \mathbb{N}$, $i = 1, \dots, d$, each process M_h^i, $h = 1, \dots, [T/\Delta]$, given by

$$M_h^i = \sum_{k=0}^{h-1} \sum_{\ell,m\in\mathbb{N}} \sum_{j=1,\dots,d} D_j\sigma^{i,m}(k\Delta, \bar{X}_{\tau^\varepsilon\wedge(k\Delta)})\sigma^{j,\ell}(k\Delta, \bar{X}_{\tau^\varepsilon\wedge(k\Delta)}) \Big(c_{\ell,m}^k(\Delta,\varepsilon) - \mathbb{E}\Big[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\Big]\Big),$$

is a discrete martingale with respect to the filtration $(\mathcal{F}_{h\Delta})_{h=1}^{[T/\Delta]}$.

Lemma 3.10. *For each $i = 1, \dots, d$,*

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|M_h^i\right|^2\right] \lesssim \left(\frac{\Delta}{\varepsilon}\right)^2 \log(1+\varepsilon^{-2}) + \Delta\log^2(1+\varepsilon^{-2}).$$

Proof. Combining Doob's maximal inequality and martingale property gives

$$\begin{aligned}\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|M_h^i\right|^2\right] &\lesssim \mathbb{E}\left[\left|M_{[T/\Delta]}^i\right|^2\right]\\ &\lesssim \sum_{k=0}^{[T/\Delta]-1} \mathbb{E}\left[\Big|\sum_{\ell,m\in\mathbb{N}} c_{\ell,m}^k(\Delta,\varepsilon) - \mathbb{E}\big[c_{\ell,m}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\big]\Big|^2\right],\end{aligned}$$

where

$$\mathbb{E}\left[\Big|\sum_{\ell,m\in\mathbb{N}} c_{\ell,m}^k(\Delta,\varepsilon) - \mathbb{E}\Big[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\Big]\Big|^2\right] \lesssim \mathbb{E}\left[\Big|\sum_{\ell,m\in\mathbb{N}} c_{\ell,m}^k(\Delta,\varepsilon)\Big|^2\right],$$

for each $k = 0, \dots, [T/\Delta] - 1$, because the conditional expectation is an L^2-projection. Thus, by independence of $Y^{\varepsilon,\ell}$ and $Y^{\varepsilon,m}$, for every $\ell \neq m$, we can estimate

$$\begin{aligned}\mathbb{E}\left[\sup_{\substack{h=1,\dots,T/\Delta\\ h\Delta\le\tau^\varepsilon}} \left|M_h^i\right|^2\right] &\lesssim \sum_{k=0}^{T/\Delta-1} \sum_{\ell,m\in\mathbb{N}} \mathbb{E}\left[\Big|\int_{k\Delta}^{(k+1)\Delta} \left(W_s^{\varepsilon,\ell} - W_{k\Delta}^{\varepsilon,\ell}\right) \mathrm{d}W_s^{\varepsilon,m}\Big|^2\right]\\ &\lesssim \sum_{k=0}^{T/\Delta-1} \sum_{\ell,m\in\mathbb{N}} \Delta \int_{k\Delta}^{(k+1)\Delta} \mathbb{E}\left[\left|\left(W_s^{\varepsilon,\ell} - W_{k\Delta}^{\varepsilon,\ell}\right) Y_s^{\varepsilon,m}\right|^2\right] \mathrm{d}s\\ &\lesssim \sum_{k=0}^{T/\Delta-1} \sum_{\ell,m\in\mathbb{N}} \Delta \int_{k\Delta}^{(k+1)\Delta} \mathbb{E}\left[\left|W_s^{\varepsilon,\ell} - W_{k\Delta}^{\varepsilon,\ell}\right|^{2q}\right]^{1/q} \mathbb{E}\left[\left|Y_s^{\varepsilon,m}\right|^{2q'}\right]^{1/q'} \mathrm{d}s\end{aligned}$$

$$\lesssim \sum_{k=0}^{T/\Delta-1} \sum_{\ell,m\in\mathbb{N}} q_\ell q_m \Delta\varepsilon^{-2}\log(1+\varepsilon^{-2}) \int_{k\Delta}^{(k+1)\Delta} \left(\Delta + \varepsilon^2\log(1+\varepsilon^{-2})\right) \mathrm{d}s$$
$$\lesssim \left(\frac{\Delta}{\varepsilon}\right)^2 \log(1+\varepsilon^{-2}) + \Delta\log^2(1+\varepsilon^{-2}).$$

□

To eventually cover the remainder of the 2nd sum on the right-hand side of (20), after subtracting the martingale term M_h, we introduce

$$N_h^i = \sum_{k=0}^{h-1} \sum_{\ell,m\in\mathbb{N}} \sum_{j=1,\dots,d} D_j\sigma^{i,m}(k\Delta,\bar{X}_{k\Delta})\sigma^{j,\ell}(k\Delta,\bar{X}_{k\Delta}) \left(\mathbb{E}\left[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] - \delta_{\ell,m}\frac{q_m}{2}\Delta\right).$$

Lemma 3.11. *For each $i = 1,\dots,d$,*

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|N_h^i\right|^2\right] \lesssim \left(\frac{\varepsilon^2}{\Delta}\right)^2 \log^2(1+\varepsilon^{-2}).$$

Proof. The proof is an easy consequence of (21). Indeed,

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|N_h^i\right|^2\right] \lesssim \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|\sum_{k=0}^{h-1}\sum_{\ell,m\in\mathbb{N}} \left|\mathbb{E}\left[c_{\ell,m}^k(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] - \delta_{\ell,m}\frac{q_m}{2}\Delta\right|\right|^2\right]$$
$$\lesssim \varepsilon^4\log^2(1+\varepsilon^{-2})\Delta^{-1} \sum_{k=0}^{[T/\Delta]-1} \sum_{\ell,m\in\mathbb{N}} q_\ell q_m$$
$$\lesssim \left(\frac{\varepsilon^2}{\Delta}\right)^2 \log^2(1+\varepsilon^{-2}).$$

□

All in all, Lemmas 3.10 and 3.11 together imply

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|\sum_{k=0}^{h-1}\left(I_6^k - J_4^k\right)\right|^2\right] = \mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon}} \left|(M_h + N_h)\right|^2\right]$$
$$\lesssim \left(\frac{\Delta}{\varepsilon}\right)^2 \log(1+\varepsilon^{-2}) + \Delta\log^2(1+\varepsilon^{-2}) + \left(\frac{\varepsilon^2}{\Delta}\right)^2 \log^2(1+\varepsilon^{-2}),$$

showing that the 2nd sum on the right-hand side of (20) can be neglected, like the 5th one, when $\varepsilon \downarrow 0$, and $\Delta = \Delta_\varepsilon$ behaves as described in Remark 3.9.

Recall that we wanted to control the remaining sums in terms of the difference $X^\varepsilon - \bar{X}$ itself, which is obvious for the first and third sum on the right-hand side of (20). However, in case of the fourth sum, applying almost the same martingale

argument used in case of the 2nd sum, each term I_5^k can be formally replaced by $\int_{k\Delta}^{(k+1)\Delta} \left(C(k\Delta, X_{k\Delta}^\varepsilon) - C(k\Delta, \bar{X}_{k\Delta})\right) \mathrm{d}s$, subject to a sufficiently small ε-correction, eventually leading to the wanted contraction argument in this case, too.

On the whole, we have justified that, if $\Delta = \Delta_\varepsilon$ behaves as described in Remark 3.9, then

$$\mathbb{E}\left[\sup_{\substack{k'=0,\dots,h\\ k'\Delta\le\tau^\varepsilon}} \left|X_{k'\Delta}^\varepsilon - \bar{X}_{k'\Delta}\right|^2\right] \lesssim r(\Delta,\varepsilon) + \sum_{k=0}^{h-1} \Delta\mathbb{E}\left[\sup_{\substack{k'=0,\dots,k\\ k'\Delta\le\tau^\varepsilon}} \left|X_{k'\Delta}^\varepsilon - \bar{X}_{k'\Delta}\right|^2\right],$$
$$h = 1,\dots,[T/\Delta],$$

where $r(\Delta,\varepsilon) \to 0,\ \varepsilon \downarrow 0$, finally proving (17), by Gronwall's lemma.

The proof of Theorem 2.2(ii) is thus complete.

4. Weak convergence

In this section we prove part (i) of Theorem 2.2. The idea of proof is similar to the one of part (ii), except that now $\beta \neq 0$ is possible. It is the existence of this bilinear term which prevents us from proving convergence in probability—we only succeed in showing convergence in law (see Remark 4.3(ii)).

First, we prove weak convergence of the bilinear term.

Second, we prove convergence in law of $X^\varepsilon,\ \varepsilon \downarrow 0$, using bounds similar to those obtained in Sect. 3.

4.1. Weak convergence of the bilinear term

For any $\varepsilon > 0$, define the process U^ε by

$$U_t^\varepsilon = \int_0^t \varepsilon\beta(Y_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s, \quad t \in [0,T], \tag{22}$$

where Y^ε is the stationary Ornstein–Uhlenbeck process introduced in Remark 2.1. By (A5), the process U^ε has zero mean, and, using (A3), its second moments,

$$\mathbb{E}\left[\int_0^t \varepsilon \underbrace{\langle\beta(Y_s^\varepsilon, Y_s^\varepsilon), \mathbf{e}_i\rangle}_{\beta^i(Y_s^\varepsilon, Y_s^\varepsilon)} \mathrm{d}s \int_0^t \varepsilon \underbrace{\langle\beta(Y_r^\varepsilon, Y_r^\varepsilon), \mathbf{e}_j\rangle}_{\beta^j(Y_r^\varepsilon, Y_r^\varepsilon)} \mathrm{d}r\right],$$

can be calculated to be

$$\frac{1}{2}\sum_{\ell,m\in\mathbb{N}} \underbrace{\langle\beta(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle}_{\beta^i_{\ell,m}} \underbrace{\langle\beta(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_j\rangle}_{\beta^j_{\ell,m}} q_\ell q_m \left(t + \frac{\varepsilon^2}{2}\left(e^{-2\varepsilon^{-2}t} - 1\right)\right),$$

for $i, j = 1,\dots,d$, and $\ell, m \in \mathbb{N}$.

Recalling (12), using the above short notation, we also have that

$$b^i_{\ell,m} = \beta^i_{\ell,m}\sqrt{\frac{q_\ell q_m}{2}}, \quad i = 1, \dots, d, \ \ell, m \in \mathbb{N}.$$

Next, since $\mathrm{d}Y^{\varepsilon,\ell}_t = -\varepsilon^{-2}Y^{\varepsilon,\ell}_t \mathrm{d}t + \varepsilon^{-2} d\langle W_t, \mathbf{f}_\ell\rangle$, Itô's formula implies

$$\begin{aligned} Y^{\varepsilon,\ell}_t Y^{\varepsilon,m}_t = Y^{\varepsilon,\ell}_0 Y^{\varepsilon,m}_0 - 2\varepsilon^{-2}\int_0^t Y^{\varepsilon,\ell}_s Y^{\varepsilon,m}_s \mathrm{d}s + \varepsilon^{-2}\int_0^t Y^{\varepsilon,\ell}_s d\langle W_s, \mathbf{f}_m\rangle \\ + \varepsilon^{-2} \int_0^t Y^{\varepsilon,m}_s d\langle W_s, \mathbf{f}_\ell\rangle + \frac{t\varepsilon^{-4}}{2} q_\ell \delta_{\ell,m}, \end{aligned}$$

for any $\ell, m \in \mathbb{N}$, and hence

$$\begin{aligned} U^{\varepsilon,i}_t = \int_0^t \varepsilon \sum_{\ell,m\in\mathbb{N}} \beta^i_{\ell,m} Y^{\varepsilon,\ell}_s Y^{\varepsilon,m}_s \mathrm{d}s &= \varepsilon \int_0^t \sum_{\ell,m\in\mathbb{N}} \beta^i_{\ell,m} Y^{\varepsilon,\ell}_s d\langle W_s, \mathbf{f}_m\rangle \\ &\quad - \frac{\varepsilon^3}{2} \sum_{\ell,m\in\mathbb{N}} \beta^i_{\ell,m}\left(Y^{\varepsilon,\ell}_t Y^{\varepsilon,m}_t - Y^{\varepsilon,\ell}_0 Y^{\varepsilon,m}_0\right) + \frac{\varepsilon^{-1}}{4} t \sum_{\ell\in\mathbb{N}} \beta^i_{\ell,\ell} q_\ell \\ &= M^{\varepsilon,i}_t - \frac{1}{2} V^{\varepsilon,i}_t + \frac{\varepsilon^{-1}}{4} t \sum_{\ell\in\mathbb{N}} \beta^i_{\ell,\ell} q_\ell, \end{aligned}$$

where M^ε is a d-dimensional continuous local martingale, while the process V^ε satisfies

$$\mathbb{E}\left[\sup_{t\le T}\left|V^\varepsilon_t\right|^p\right] = \mathbb{E}\left[\sup_{t\le T}\left|\varepsilon^3\left(\beta\left(Y^\varepsilon_t, Y^\varepsilon_t\right) - \beta\left(Y^\varepsilon_0, Y^\varepsilon_0\right)\right)\right|^p\right] \lesssim \varepsilon^p \log^p(1+\varepsilon^{-2}), \quad \forall\, p > 1, \tag{23}$$

by combining (A3) and Lemma3.1.

Remark 4.1. Using $\sum_{\ell,m\in\mathbb{N}} \beta^i_{\ell,m} q_\ell q_m < \infty$ for every $i = 1, \dots, d$, it is possible to prove that M^ε is a square integrable martingale for every $\varepsilon > 0$. However, we will not need this in the following.

The above representation of U^ε, though very simple, has been used in a variety of cases in a fruitful way, see for instance [19] or [10]. Observe that, by (A5), the Itô-correction actually cancels out, being otherwise a contribution of order ε^{-1}. The process U^ε, nevertheless, has got an interesting limit in law:

Proposition 4.2. *The couple of processes (U^ε, W) converges in law, $\varepsilon \downarrow 0$, to a pair of processes (η, ω), where η is a d-dimensional Wiener process with covariance $(\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m})^d_{i,j=1}$, and ω is a Q-Wiener process, like W. Furthermore, η and ω are independent.*

Proof. First, by (23), it is sufficient to prove the proposition for (M^ε, W) instead of (U^ε, W).

Since all components of the processes M^ε, $\varepsilon > 0$, and of W are continuous local martingales, the distributional properties of the limit (η, ω) would follow from [4, Chapter VII, Theorem 1.4], if

$$\mathbb{E}\left[\left(\left[M^{\varepsilon,i}, M^{\varepsilon,j}\right]_t - t\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m}\right)^2\right] \to 0, \quad \varepsilon \downarrow 0, \tag{24}$$

for each $t \in [0, T]$, and $i, j = 1, \ldots, d$, as well as

$$\mathbb{E}\left[\left(\left[M^{\varepsilon,i}, \langle W, \mathbf{f}_m\rangle\right]_t\right)^2\right] \to 0, \quad \varepsilon \downarrow 0,$$

for each $t \in [0, T]$, $i = 1, \ldots, d$, and $m \in \mathbb{N}$.

First, fix $t \in [0, T]$, as well as $i, j = 1, \ldots, d$. Then, the quadratic covariation $\left[M^{\varepsilon,i}, M^{\varepsilon,j}\right]_t$ is given by

$$\left[M^{\varepsilon,i}, M^{\varepsilon,j}\right]_t = \varepsilon^2 \int_0^t \sum_{m\in\mathbb{N}} \sum_{\ell,\ell'\in\mathbb{N}} \beta^i_{\ell,m}\beta^j_{\ell',m} q_m Y^{\varepsilon,\ell}_s Y^{\varepsilon,\ell'}_s \mathrm{d}s,$$

so that

$$\begin{aligned}
&\mathbb{E}\left[\left(\left[M^{\varepsilon,i}, M^{\varepsilon,j}\right]_t - t\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m}\right)^2\right] \\
&\quad = \varepsilon^4 \iint_0^t \sum_{m,\underline{m}\in\mathbb{N}} \sum_{\substack{\ell,\ell'\in\mathbb{N}\\ \underline{\ell},\underline{\ell}'\in\mathbb{N}}} \beta^i_{\ell,m}\beta^j_{\ell',m}\beta^i_{\underline{\ell},\underline{m}}\beta^j_{\underline{\ell}',\underline{m}} q_m q_{\underline{m}} \mathbb{E}\left[Y^{\varepsilon,\ell}_s Y^{\varepsilon,\ell'}_s Y^{\varepsilon,\underline{\ell}}_r Y^{\varepsilon,\underline{\ell}'}_r\right] \mathrm{d}s\mathrm{d}r \\
&\qquad - 2\varepsilon^2 \int_0^t \sum_{m\in\mathbb{N}} \sum_{\ell,\ell'\in\mathbb{N}} \beta^i_{\ell,m}\beta^j_{\ell',m} q_m \mathbb{E}\left[Y^{\varepsilon,\ell}_s Y^{\varepsilon,\ell'}_s\right] \mathrm{d}s \left(t\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m}\right) \\
&\qquad + \left(t\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m}\right)^2.
\end{aligned}$$

Now, using that one can easily calculate $\mathbb{E}\left[Y^{\varepsilon,\ell}_s Y^{\varepsilon,\ell'}_s\right] = \frac{\varepsilon^{-2}}{2} q_\ell \delta_{\ell,\ell'}$, it follows from Isserlis–Wick's theorem, see [14, Theorem 1.28], that

$$\mathbb{E}\left[Y^{\varepsilon,\ell}_s Y^{\varepsilon,\ell'}_s Y^{\varepsilon,\underline{\ell}}_r Y^{\varepsilon,\underline{\ell}'}_r\right] = \frac{\varepsilon^{-4}}{4}\left(q_\ell q_{\underline{\ell}} \delta_{\ell,\ell'}\delta_{\underline{\ell},\underline{\ell}'} + q_\ell q_{\ell'} e^{-2\varepsilon^{-2}|s-r|}\left(\delta_{\ell,\underline{\ell}}\delta_{\ell',\underline{\ell}'} + \delta_{\ell,\underline{\ell}'}\delta_{\ell',\underline{\ell}}\right)\right),$$

which yields

$$\mathbb{E}\left[\left(\left[M^{\varepsilon,i}, M^{\varepsilon,j}\right]_t - t\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m}\right)^2\right]$$

$$= \left(t \sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m} - t \sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m} \right)^2 + O(\varepsilon^2) \lesssim \varepsilon^2,$$

proving (24).

Second, fix $t \in [0, T]$, as well as $i = 1, \dots, d$, $m \in \mathbb{N}$. Then,

$$\left[M^{\varepsilon,i}, \langle W, \mathbf{f}_m \rangle \right]_t = \int_0^t \beta^i (\varepsilon Y^\varepsilon_s, Q\mathbf{f}_m)\, \mathrm{d}s,$$

where, using Lemma 3.1,

$$\mathbb{E}\left[|\int_0^t \beta^i (\varepsilon Y^\varepsilon_s, Q\mathbf{f}_m)\, \mathrm{d}s\, |^2 \right] = \mathbb{E}\left[|\, \beta^i (\varepsilon \int_0^t Y^\varepsilon_s \mathrm{d}s\, , Q\mathbf{f}_m)\, |^2 \right]$$
$$\lesssim \mathbb{E}[\, |\, \overbrace{\varepsilon \int_0^t Y^\varepsilon_s \mathrm{d}s}^{\varepsilon W_t - \varepsilon^3 (Y^\varepsilon_t - Y^\varepsilon_0)}\, |^2\, q_m^2\,] \xrightarrow{\varepsilon \downarrow 0} 0,$$

finishing the proof of the proposition. □

Remark 4.3. (i) Of course, a d-dimensional Wiener process with covariance $(\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m})^d_{i,j=1}$ can always be represented by $\sum_{\ell,m\in\mathbb{N}} b_{\ell,m} \bar{W}^{\ell,m}$, where $\{\bar{W}^{\ell,m}\}_{\ell,m\in\mathbb{N}}$ is a family of independent one-dimensional standard Wiener processes.

(ii) We would like to stress that we do not expect a much stronger convergence of U^ε, when $\varepsilon \downarrow 0$, as the one stated in the above proposition. Indeed, it turns out to be that the sequence $\{M^\varepsilon\}_{\varepsilon>0}$ is not even a Cauchy sequence in $L^2(\Omega; \mathbb{R}^d)$. To see this, for fixed $0 < \varepsilon < \underline{\varepsilon}$, and some $1 \le i \le d$, consider

$$\mathbb{E}\left[\sup_{t\le T} \left| M^{\varepsilon,i}_t - M^{\underline{\varepsilon},i}_t \right|^2 \right] = \mathbb{E}\left[\sup_{t\le T} \left| \int_0^t \sum_{\ell,m\in\mathbb{N}} \beta^i_{\ell,m} \left(\varepsilon Y^{\varepsilon,\ell}_s - \underline{\varepsilon} Y^{\underline{\varepsilon},\ell}_s \right) d\langle W_s, \mathbf{f}_m \rangle \right|^2 \right].$$

But, by Burkholder–Davis–Gundy's inequality, the above expectation can be bound from below by

$$\mathbb{E}\left[\int_0^T \sum_{m\in\mathbb{N}} \left(\sum_{\ell\in\mathbb{N}} \beta^i_{\ell,m} \left(\varepsilon Y^{\varepsilon,\ell}_s - \underline{\varepsilon} Y^{\underline{\varepsilon},\ell}_s \right) \right)^2 q_m \mathrm{d}s \right]$$
$$= T \sum_{\ell,m\in\mathbb{N}} (\beta^i_{\ell,m})^2 q_\ell q_m \left(1 - \frac{2\varepsilon^{-1}\underline{\varepsilon}^{-1}}{\varepsilon^{-2} + \underline{\varepsilon}^{-2}} \right),$$

where

$$\lim_{\varepsilon\to 0} \left(1 - \frac{2\varepsilon^{-1}\underline{\varepsilon}^{-1}}{\varepsilon^{-2} + \underline{\varepsilon}^{-2}} \right) = 1, \quad \text{for every fixed} \underline{\varepsilon} > 0,$$

so that $\{M^{\varepsilon,i}\}_{\varepsilon>0}$ cannot be Cauchy in $L^2(\Omega)$.

4.2. Weak convergence of solutions

We now prove $X^\varepsilon \to \bar{X}$, in law, when $\varepsilon \downarrow 0$.

First, for each $\varepsilon > 0$, let $\hat{X}^\varepsilon$ be the solution of

$$\hat{X}^\varepsilon_t = x_0 + \int_0^t \Big(F(s, \hat{X}^\varepsilon_s) + C(s, \hat{X}^\varepsilon_s)\Big)\,\mathrm{d}s + \int_0^t \sigma(s, \hat{X}^\varepsilon_s)\mathrm{d}W_s + U^\varepsilon_t, \quad t \in [0, T], \tag{25}$$

where U^ε is given by (22), and let $\tau^\varepsilon_R = \inf\{t \geq 0 : |X^\varepsilon_t| \geq R\} \wedge \inf\{t \geq 0 : |\hat{X}^\varepsilon_t| \geq R\}$.

Note that, if (A4), then the coefficients F, C, σ, β must have properties such that each of the above equations admits global solutions on $[0, T]$, too.

Next, taking into account $\mathbb{E}\left[\left|\sup_{s\in[0,T]} \varepsilon\beta(Y^\varepsilon_s, Y^\varepsilon_s)\right|^p\right] \lesssim \varepsilon^{-p}\log^p(1+\varepsilon^{-2})$ we can estimate increments of U^ε with

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\ldots,[T/\Delta]\\ t\leq\tau,\, t+k\Delta\leq T\wedge\tau^\varepsilon_R}} |U^\varepsilon_{t+k\Delta} - U^\varepsilon_{k\Delta}|^p\right] \lesssim \left(\frac{\tau}{\varepsilon}\right)^p \log^p(1+\varepsilon^{-2}). \tag{26}$$

As a consequence, it can easily be verified that the analogous of Lemmas 3.3 and 3.4 would still be valid for the process X^ε, despite $\beta \neq 0$, on the one hand, and that the following versions

$$\mathbb{E}\left[\sup_{t\leq\tau,\, t+k\Delta\leq T\wedge\tau^\varepsilon_R} |\hat{X}^\varepsilon_{t+k\Delta} - \hat{X}^\varepsilon_{k\Delta}|^p\right] \lesssim \tau^{\frac{p}{2}} + \left(\frac{\tau}{\varepsilon}\right)^p \log^p(1+\varepsilon^{-2}),$$
$$p > 1,\ \tau \in (0,1),\ k \in \{0, 1, \ldots, [T/\Delta]\},$$

and

$$\mathbb{E}\left[\sup_{\substack{k=0,1,\ldots,[T/\Delta]\\ t\leq\Delta,\, t+k\Delta\leq T\wedge\tau^\varepsilon_R}} |\hat{X}^\varepsilon_{t+k\Delta} - \hat{X}^\varepsilon_{k\Delta}|^p\right] \lesssim \Delta^{\frac{p}{2}-1} + \frac{\Delta^{p-1}}{\varepsilon^p}\log^p(1+\varepsilon^{-2}), \quad p > 1,$$

of Lemmas 3.5 and 3.6, respectively, would hold true when replacing $\bar{X}$ by $\hat{X}^\varepsilon$, on the other. We point out that the proof of this claim differs from those in Sect. 3 only for the term U^ε, which however can be controlled by (26).

Therefore, when expanding X^ε and $\hat{X}^\varepsilon$ as in (18) and (19), but including the β-term, and then arguing as in the proof of Theorem 2.2(ii) in Sect. 3, it would immediately follow that $X^\varepsilon_{\cdot\wedge\tau^\varepsilon_R} - \hat{X}^\varepsilon_{\cdot\wedge\tau^\varepsilon_R} \to 0$, in probability, $\varepsilon \downarrow 0$, for any $R > 0$, once the following lemma is also available.

Lemma 4.4. *Assume that $\Delta = \Delta_\varepsilon$ behaves as described in Remark 3.9. Then,*

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau_R^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} D\sigma(r,X_r^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon\right|^2\right]\to 0,\quad \varepsilon\downarrow 0.$$

Proof. To start with, write

$$\begin{aligned}
&\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} D\sigma(r,X_r^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon\\
&\quad=\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\left(D\sigma(r,X_r^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)-D\sigma(k\Delta,X_{k\Delta}^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\right)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon\\
&\qquad+\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} D\sigma(k\Delta,X_{k\Delta}^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon,
\end{aligned}$$

which creates two summands, for any fixed $0\le k\le [T/\Delta]-1$.

We estimate the impact of each summand separately.

First, using $|D\sigma(r,X_r^\varepsilon)-D\sigma(k\Delta,X_{k\Delta}^\varepsilon)|\lesssim |X_r^\varepsilon-X_{k\Delta}^\varepsilon|+\omega_\sigma(\Delta)$, we obtain that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau_R^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\left(D\sigma(r,X_r^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)-D\sigma(k\Delta,X_{k\Delta}^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\right)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon\right|^2\right]\\
&\lesssim \varepsilon^{-4}\log^2(1+\varepsilon^{-2})\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau_R^\varepsilon}}\left|\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s}\left(|X_r^\varepsilon-X_{k\Delta}^\varepsilon|+\omega_\sigma(\Delta)\right)\mathrm{d}r\right)\mathrm{d}s\right|^2\right]\\
&\lesssim \varepsilon^{-4}\log^2(1+\varepsilon^{-2})\mathbb{E}\left[\sum_{k=0}^{\lceil T\wedge\tau_R^\varepsilon\rceil/\Delta-1}\int_{k\Delta}^{(k+1)\Delta}\left|\int_{k\Delta}^{s}\left(|X_r^\varepsilon-X_{k\Delta}^\varepsilon|+\omega_\sigma(\Delta)\right)\mathrm{d}r\right|^2\mathrm{d}s\right]\\
&\lesssim \varepsilon^{-4}\log^2(1+\varepsilon^{-2})\sum_{k=0}^{[T/\Delta]-1}\int_{k\Delta}^{(k+1)\Delta}(s-k\Delta)\int_{k\Delta}^{s}\left(\mathbb{E}\left[|X_{r\wedge\tau_R^\varepsilon}^\varepsilon-X_{k\Delta\wedge\tau_R^\varepsilon}^\varepsilon|^2\right]+\omega_\sigma(\Delta)^2\right)\mathrm{d}r\mathrm{d}s\\
&\lesssim\left(\frac{\Delta^2}{\varepsilon^3}\right)^2\log^3(1+\varepsilon^{-2})+\left(\frac{\Delta}{\varepsilon^2}\right)^2\log^2(1+\varepsilon^{-2})\,\omega_\sigma(\Delta)^2.
\end{aligned}$$

Second, we approach

$$\sum_{k=0}^{h-1}\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} D\sigma(k\Delta,X_{k\Delta}^\varepsilon)\varepsilon\beta(Y_r^\varepsilon,Y_r^\varepsilon)\mathrm{d}r\right)\mathrm{d}W_s^\varepsilon \tag{27}$$

following the method used when discussing the 2nd sum on the right-hand side of (20) in the proof of Theorem 2.2(ii), but now for triple moments of Y^ε.

Indeed, define

$$c_{\ell,m,n}^k(\Delta,\varepsilon)=\int_{k\Delta}^{(k+1)\Delta}\left(\int_{k\Delta}^{s} Y_r^{\varepsilon,\ell}Y_r^{\varepsilon,m}\mathrm{d}r\right)Y_s^{\varepsilon,n}\mathrm{d}s,$$

and take the conditional expectation with respect to $\mathcal{F}_{k\Delta}$, that is

$$\mathbb{E}\left[c^k_{\ell,m,n}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] = \int_{k\Delta}^{(k+1)\Delta} \left(\int_{k\Delta}^{s} \mathbb{E}\left[Y^{\varepsilon,\ell}_r Y^{\varepsilon,m}_r Y^{\varepsilon,n}_s \mid \mathcal{F}_{k\Delta}\right] \mathrm{d}r\right) \mathrm{d}s.$$

Since

$$\begin{aligned}\mathbb{E}\left[Y^{\varepsilon,\ell}_r Y^{\varepsilon,m}_r Y^{\varepsilon,n}_s \mid \mathcal{F}_{k\Delta}\right] = {} & Y^{\varepsilon,\ell}_{k\Delta} Y^{\varepsilon,m}_{k\Delta} Y^{\varepsilon,n}_{k\Delta} e^{-\varepsilon^{-2}(s+2r-3k\Delta)} \\ & + \left(Y^{\varepsilon,\ell}_{k\Delta}\delta_{m,n}q_n + Y^{\varepsilon,m}_{k\Delta}\delta_{\ell,n}q_n + Y^{\varepsilon,n}_{k\Delta}\delta_{\ell,m}q_\ell\right) \\ & \frac{\varepsilon^{-2}}{2}\left(e^{-\varepsilon^{-2}(s-k\Delta)} - e^{-\varepsilon^{-2}(s+2r-3k\Delta)}\right),\end{aligned}$$

we have that

$$\begin{aligned}\mathbb{E}\left[c^k_{\ell,m,n}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right] = {} & Y^{\varepsilon,\ell}_{k\Delta} Y^{\varepsilon,m}_{k\Delta} Y^{\varepsilon,n}_{k\Delta} \frac{\varepsilon^4}{2}\left(1 - e^{-\varepsilon^{-2}\Delta} - \frac{1}{3} + \frac{1}{3}e^{-3\varepsilon^{-2}\Delta}\right) \\ & + \left(Y^{\varepsilon,\ell}_{k\Delta}\delta_{m,n}q_n + Y^{\varepsilon,m}_{k\Delta}\delta_{\ell,n}q_n + Y^{\varepsilon,n}_{k\Delta}\delta_{\ell,m}q_\ell\right) \\ & \times \frac{\varepsilon^2}{2}\left(\frac{\Delta}{\varepsilon^2}e^{-\varepsilon^{-2}\Delta} + \frac{1}{2} - \frac{1}{2}e^{-\varepsilon^{-2}\Delta} + \frac{1}{6} - \frac{1}{6}e^{-3\varepsilon^{-2}\Delta}\right).\end{aligned}$$

Next, for each $i = 1, \dots, d$, the process $M^i_h,\ h = 1, \dots, [T/\Delta]$, given by

$$M^i_h = \sum_{k=0}^{h-1} \sum_{\ell,m,n\in\mathbb{N}} \sum_{j=1,\dots,d} D_j\sigma^{i,n}(k\Delta, X^\varepsilon_{\tau^\varepsilon_R\wedge k\Delta})\varepsilon\beta^j_{\ell,m}\left(c^k_{\ell,m,n}(\Delta,\varepsilon) - \mathbb{E}\left[c^k_{\ell,m,n}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right]\right),$$

is a martingale with respect to the filtration $(\mathcal{F}_{h\Delta})_{h=1}^{[T/\Delta]}$, and arguing as in the proof of Lemma 3.10 yields

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,[T/\Delta]\\ h\Delta\le\tau^\varepsilon_R}} \left|M^i_h\right|^2\right] \lesssim \frac{\Delta^3}{\varepsilon^4}\log^3(1+\varepsilon^{-2}), \quad i = 1, \dots, d.$$

So, it remains to prove that the remainder, after subtracting the martingale term M_h from (27), also vanishes, when $\varepsilon \downarrow 0$. For $i = 1, \dots, d$, the ith coordinate of this remainder reads

$$N^i_h = \sum_{k=0}^{h-1} \sum_{\ell,m,n\in\mathbb{N}} \sum_{j=1,\dots,d} D_j\sigma^{i,n}(k\Delta, X^\varepsilon_{k\Delta})\varepsilon B^j_{\ell,m}\mathbb{E}\left[c^k_{\ell,m,n}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right],$$

and we can easily calculate the below bound,

$$\mathbb{E}\left[\sup_{\substack{h=1,\dots,T/\Delta\\ h\Delta\le\tau^\varepsilon_R}} \left|N^i_h\right|^2\right] \lesssim \Delta^{-1}\sum_{k=0}^{T/\Delta-1} \mathbb{E}\left[\left|\varepsilon\mathbb{E}\left[c^k_{\ell,m,n}(\Delta,\varepsilon) \mid \mathcal{F}_{k\Delta}\right]\right|^2\right]$$

$$\lesssim \left(\frac{\varepsilon^2}{\Delta}\right)^2 \log^3(1+\varepsilon^{-2}),$$

finishing the proof of the lemma. □

Corollary 4.5. *For any $R > 0$, if $\Delta = \Delta_\varepsilon$ behaves as described in Remark 3.9,*

$$\mathbb{E}\left[\sup_{t\le T\wedge\tau_R^\varepsilon} |X_t^\varepsilon - \hat{X}_t^\varepsilon|^2\right] \to 0, \quad \varepsilon \downarrow 0,$$

and hence $X^\varepsilon_{\cdot\wedge\tau_R^\varepsilon} - \hat{X}^\varepsilon_{\cdot\wedge\tau_R^\varepsilon} \to 0$, in probability, $\varepsilon \downarrow 0$, in particular.

The above corollary suggests that it would be sufficient to show that $\hat{X}^\varepsilon_{\cdot\wedge\tau_R^\varepsilon} \to \bar{X}_{\cdot\wedge\tau_R^\varepsilon}$, in law, when $\varepsilon \downarrow 0$, subject to some procedure allowing to let R go to infinity, afterwards. So, we at first prove the weak convergence for fixed R and then discuss the limit-procedure for $R \to \infty$.

Modify the coefficients F, σ outside the set $\{(t, x) : |x| < R\}$ in such a way that the new coefficients F_R, σ_R, but also $D\sigma_R$, are globally bounded, and that both functions $F_R(t, \cdot)$ and $D\sigma_R(t, \cdot)$ are globally Lipschitz, uniformly in $t \in [0, T]$.

Of course, $\hat{X}^\varepsilon_{\cdot\wedge\tau_R^\varepsilon}$ coincides with $\hat{X}^{\varepsilon,R}_{\cdot\wedge\tau_R^\varepsilon}$, where $\hat{X}^{\varepsilon,R}$ denotes the solution to the equation obtained when replacing the coefficients of (25) by F_R, σ_R, and the Stratonovich correction C_R associated with σ_R. Also, let $\bar{X}^R$ denote the solution to the equation obtained when replacing the coefficients of (13) by F_R, σ_R, C_R.

Proposition 4.6. *Fix $R > 0$. Then, $\hat{X}^{\varepsilon,R}$ converges to $\bar{X}^R$, in law, when $\varepsilon \downarrow 0$.*

Proof. Since

$$\hat{X}_t^{\varepsilon,R} - U_t^\varepsilon = x_0 + \int_0^t \left(F_R(s, \hat{X}_s^{\varepsilon,R}) + C_R(s, \hat{X}_s^{\varepsilon,R})\right) \mathrm{d}s + \int_0^t \sigma_R(s, \hat{X}_s^{\varepsilon,R}) \mathrm{d}W_s,$$

by boundedness of the coefficients on the above right-hand side, we obtain that

$$\mathbb{E}\left[\sup_{t\le T} |\hat{X}_t^{\varepsilon,R} - U_t^\varepsilon|\right] \lesssim |x_0| + T + \mathbb{E}\left[\sup_{t\le T} \left|\int_0^t \sigma_R(s, \hat{X}_s^{\varepsilon,R}) \mathrm{d}W_s\right|\right],$$

where Burkholder–Davis–Gundy's inequality gives $\mathbb{E}\left[\sup_{t\le T} |\int_0^t \sigma_R(s, \hat{X}_s^{\varepsilon,R}) \mathrm{d}W_s|\right] \lesssim T^{1/2}$.

Similarly, $\mathbb{E}\left[|(\hat{X}_{t_2}^{\varepsilon,R} - U_{t_2}^\varepsilon) - (\hat{X}_{t_1}^{\varepsilon,R} - U_{t_1}^\varepsilon)|^p\right] \lesssim |t_2 - t_1|^{p/2}$, for any $|t_2 - t_1| < 1$, and any $p > 1$. Thus, by Kolmogorov–Chentsov's theorem, for every $\alpha \in (0, 1)$, one can find $\Delta \in (0, 1)$ such that

$$\mathbb{P}\left\{\sup_{t_1,t_2\in[0,T],\, |t_2-t_1|\le\Delta} \frac{|(\hat{X}_{t_2}^{\varepsilon,R} - U_{t_2}^\varepsilon) - (\hat{X}_{t_1}^{\varepsilon,R} - U_{t_1}^\varepsilon)|}{|t_2 - t_1|^\gamma} \le const\right\} \ge 1-\alpha, \quad \forall \varepsilon > 0,$$

where *const* depends on γ, but not on ε, and $\gamma \in (0, 1/2)$ can be freely chosen.

We therefore have equi-boundedness and equi-continuity of $\{\hat{X}^{\varepsilon,R} - U^{\varepsilon}\}_{\varepsilon>0}$ with arbitrarily high probability, and hence the family $\{\hat{X}^{\varepsilon,R} - U^{\varepsilon}\}_{\varepsilon>0}$ is tight with respect to the uniform topology in $C([0,T],\mathbb{R}^d)$, first applying Arzelà–Ascoli, followed by Prokhorov's theorem. Moreover, $\{U^{\varepsilon}\}_{\varepsilon>0}$ is trivially tight by Proposition 4.2, so that adding $\hat{X}^{\varepsilon,R} - U^{\varepsilon}$ and U^{ε} would make $\{\hat{X}^{\varepsilon,R}\}_{\varepsilon>0}$ tight, too.

All in all, the family of triples $\{(\hat{X}^{\varepsilon,R}, U^{\varepsilon}, W)\}_{\varepsilon>0}$ is tight.

Next, for $\varepsilon > 0$, let $\mathbb{P}^{R,\varepsilon}$ be the pushforward measure $\mathbb{P} \circ (\hat{X}^{\varepsilon,R}, U^{\epsilon}, W)^{-1}$ on the space

$$\tilde{\Omega} = C([0,T], H_d) \times C([0,T], H_d) \times C([0,T], H_\infty)$$

equipped with the Borel-σ-algebra $\mathcal{B}$, and let (ξ, η, ω) denote the coordinate process on $\tilde{\Omega}$.

By tightness of $\{(\hat{X}^{\varepsilon,R}, U^{\epsilon}, W)\}_{\varepsilon>0}$, there exists a subsequence $(\varepsilon_n)_{n\in\mathbb{N}}$ such that $\mathbb{P}^{R,\varepsilon_n}$ weakly converges to a probability measure $\mathbb{P}^R$ on $(\tilde{\Omega}, \mathcal{B})$, when $n \uparrow \infty$.

Let $\tilde{\mathcal{F}}$ be the $\mathbb{P}^R$-completion of $\mathcal{B}$, and let $(\tilde{\mathcal{F}}_t)_{t\in[0,T]}$ be the smallest filtration the process (ξ, η, ω) is adapted to, on the one hand, and which satisfies the usual conditions with respect to $\mathbb{P}^R$, on the other. Also, introduce $\tilde{\mathcal{F}}^n$, $(\tilde{\mathcal{F}}^n_t)_{t\in[0,T]}$ in a similar way with respect to $\mathbb{P}^{R,\varepsilon_n}$, $n \in \mathbb{N}$.

Now, it easily follows from Proposition 4.2 that, on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^R)$, the following distributional properties must hold for the pair of processes (η, ω): η is a d-dimensional Wiener process with covariance $(\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m})^d_{i,j=1}$, ω is a Q-Wiener process, η and ω are independent.

Introduce

$$M^R_t = \xi_t - x_0 - \int_0^t (F_R(s,\xi_s) + C_R(s,\xi_s))\,\mathrm{d}s - \eta_t, \quad t \in [0,T], \tag{28}$$

and observe that each component of both processes M^R and ω, but also

$$M^{R,i}_t M^{R,j}_t - \int_0^t \sum_{m\in\mathbb{N}} \sigma^{i,m}_R(s,\xi_s)\sigma^{j,m}_R(s,\xi_s) q_m \mathrm{d}s, \quad t \in [0,T], \quad i,j = 1,\dots,d,$$

$$M^{R,i}_t \omega^m_t - \int_0^t \sigma^{i,m}_R(s,\xi_s) q_m \mathrm{d}s, \quad t \in [0,T], \quad i = 1,\dots,d, \quad m \in \mathbb{N},$$

$$\omega^\ell_t \omega^m_t - t\delta_{\ell,m} q_m, \quad t \in [0,T], \quad \ell, m \in \mathbb{N},$$

are continuous local martingales with respect to $(\tilde{\mathcal{F}}^n_t)_{t\in[0,T]}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}^n, \mathbb{P}^{R,\varepsilon_n})$, for any $n \in \mathbb{N}$, and hence they are continuous local martingales with respect to $(\tilde{\mathcal{F}}_t)_{t\in[0,T]}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^R)$, too, by [12, IX. Cor.1.19].

Therefore, applying [3, Theorem 8.2] to the pair of process (M^R, ω) yields

$$M^R_t = \int_0^t \sigma_R(s,\xi_s) dW^R_s, \quad \omega_t = \int_0^t 1\, dW^R_s = W^R_t, \quad t \in [0,T],$$

on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^R)$, or an enlargement of this space we still denote by $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbb{P}^R)$, where W^R is another Q-Wiener process, which, by the above representation, even $\mathbb{P}^R$- almost surely coincides with ω, so that

$$M_t^R = \int_0^t \sigma_R(s, \xi_s) d\omega_s, \quad t \in [0, T], \quad \mathbb{P}^R - \text{a.s.}$$

Thus, Eq. (28) can be written as

$$\xi_t = x_0 + \int_0^t (F_R(s, \xi_s) + C_R(s, \xi_s))\, \mathrm{d}s + \int_0^t \sigma_R(s, \xi_s) d\omega_s + \eta_t, \quad t \in [0, T], \mathbb{P}^R - \text{a.s.},$$

where ω is a Q-Wiener process, while η is a d-dimensional Wiener process, independent of ω, and with covariance $(\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m})^d_{i,j=1}$. Observe that the process $\bar{X}^R$ satisfies the same type of equation, as $\sum_{\ell,m\in\mathbb{N}} b_{\ell,m} \bar{W}^{\ell,m}$ from (13) is a d-dimensional Wiener process with covariance $(\sum_{\ell,m\in\mathbb{N}} b^i_{\ell,m} b^j_{\ell,m})^d_{i,j=1}$, too. But, since this type of equation admits a unique strong solution, the laws of ξ and $\bar{X}^R$ must be the same, proving $\hat{X}^{\varepsilon_n,R} \to \bar{X}^R$, in law, when $n \uparrow \infty$. However, the same argument applies to any converging subsequence, and the limit will always be the same, finally proving $\hat{X}^{\varepsilon,R} \to \bar{X}^R$, in law, when $\varepsilon \downarrow 0$. □

It remains to discuss how R can be taken to infinity.

Recall that $\bar{X}$ is the solution of (13), and it is not difficult to see that $\bar{X}^R$ converges to $\bar{X}$, in law, as $R \to \infty$.

Now take a function $\varphi_R \in C(C([0, T], \mathbb{R}^d), [0, 1])$, such that $\varphi_R(u) = 0$, if $\sup_{t\in[0,T]} |u_t| \le R - 1$, and $\varphi_R(u) = 1$, if $\sup_{t\in[0,T]} |u_t| > R$.

Then,

$$\mathbb{P}\{\tau_R^\varepsilon < T\} \le \mathbb{P}\left\{\sup_{t\in[0,T]} |\hat{X}_t^{\varepsilon,R}| \ge R\right\} \le \mathbb{E}\left[\varphi_R(\hat{X}^{\varepsilon,R})\right],$$

and because $\hat{X}^{\varepsilon,R} \to \bar{X}^R$, in law, when $\varepsilon \downarrow 0$, we deduce that

$$\limsup_{\varepsilon\to 0} \mathbb{P}\{\tau_R^\varepsilon < T\} \le \mathbb{E}\left[\varphi_R(\bar{X}^R)\right] \le \mathbb{P}\left\{\sup_{t\in[0,T]} |\bar{X}_t^R| \ge R - 1\right\} = \mathbb{P}\left\{\sup_{t\in[0,T]} |\bar{X}_t| \ge R - 1\right\},$$

where the last probability converges to zero, when $R \to \infty$, because $\bar{X}$ is a global solution.

As a consequence, for any $\psi \in C_b(C([0, T], \mathbb{R}^d), \mathbb{R})$,

$$\begin{aligned}\left|\mathbb{E}\left[\psi(X^\varepsilon)\right] - \mathbb{E}\left[\psi(\bar{X})\right]\right| \le{}& \left|\mathbb{E}\left[\psi(X^\varepsilon)\right] - \mathbb{E}\left[\psi(X^\varepsilon_{\cdot\wedge\tau_R^\varepsilon})\right]\right| + \left|\mathbb{E}\left[\psi(X^\varepsilon_{\cdot\wedge\tau_R^\varepsilon})\right] - \mathbb{E}\left[\psi(\hat{X}^{\varepsilon,R}_{\cdot\wedge\tau_R^\varepsilon})\right]\right| \\ &+ \left|\mathbb{E}\left[\psi(\hat{X}^{\varepsilon,R}_{\cdot\wedge\tau_R^\varepsilon})\right] - \mathbb{E}\left[\psi(\hat{X}^{\varepsilon,R})\right]\right| + \left|\mathbb{E}\left[\psi(\hat{X}^{\varepsilon,R})\right] - \mathbb{E}\left[\psi(\bar{X}^R)\right]\right| \\ &+ \left|\mathbb{E}\left[\psi(\bar{X}^R)\right] - \mathbb{E}\left[\psi(\bar{X})\right]\right|.\end{aligned}$$

Here, when taking R large enough, we can make all the summands on the right-hand side, except for the second and fourth, arbitrarily small, uniformly in ε, and, for fixed R, the remaining terms go to zero, when $\varepsilon \downarrow 0$.

Thus, by a diagonal argument, the convergence in law of $X^\varepsilon \to \bar{X}$, $\varepsilon \downarrow 0$, follows, completing the proof of the theorem.

5. Application to climate models

We now apply Theorem 2.2 to perform stochastic model reduction for a subclass of the stochastic climate models given by (4), (5) in the introduction: we restrict ourselves to a simpler version of (5), omitting fast forcing $\varepsilon^{-2} f^2_{\varepsilon^{-1}t}$ and $\varepsilon^{-1} A^2_2 Y^\varepsilon_t$, on the one hand, but also neglecting the interaction $B^2_{12}(X^\varepsilon_t, Y^\varepsilon_t)$, on the other. While the first two terms we omit are technically demanding but look doable from a wider prospective, which is beyond this paper, the term $\varepsilon^{-1} B^2_{12}(X^\varepsilon_t, Y^\varepsilon_t)$ involving the neglected interaction is notoriously hard and beyond our understanding, right now.

For each $\varepsilon > 0$, let $(X^\varepsilon, Y^\varepsilon)$ be a pair of processes satisfying

$$\frac{\mathrm{d}X^\varepsilon_t}{\mathrm{d}t} = F^1_t + A^1_1 X^\varepsilon_t + A^1_2 Y^\varepsilon_t + B^1_{11}(X^\varepsilon_t, X^\varepsilon_t) + B^1_{12}(X^\varepsilon_t, Y^\varepsilon_t) + \varepsilon B^1_{22}(Y^\varepsilon_t, Y^\varepsilon_t), \tag{29}$$

$$\frac{\mathrm{d}Y^\varepsilon_t}{\mathrm{d}t} = \varepsilon^{-2} A^2_1 X^\varepsilon_t + \varepsilon^{-2} B^2_{11}(X^\varepsilon_t, X^\varepsilon_t) - \varepsilon^{-2} Y^\varepsilon_t + \varepsilon^{-2} \dot{W}_t, \tag{30}$$

where $A^1_1 : H_d \to H_d$, $A^1_2 : H_\infty \to H_d$, $A^2_1 : H_d \to H_\infty$ are bounded linear operators, $B^1_{11} : H_d \times H_d \to H_d$, $B^1_{12} : H_d \times H_\infty \to H_d$, $B^1_{22} : H_\infty \times H_\infty \to H_d$, $B^2_{11} : H_d \times H_d \to H_\infty$ are continuous bilinear maps, and $F^1 : [0, T] \to H_d$ is a deterministic continuous external force. Stochastic basis and Wiener process W are taken to be the same as in Remark 2.1.

In what follows, the above equations will always have initial conditions (x_0, y_0), where $x_0 \in H_d$ can be chosen arbitrarily, while $y_0 = \int_{-\infty}^0 \varepsilon^{-2} e^{\varepsilon^{-2} s} \mathrm{d}W_s$ will be fixed to ensure pseudo-stationarity of the scaled unresolved variables. Note that fixing $y_0 \in H_\infty$ this way would not restrict the initial data of the reduced equations.

In fluid dynamics settings like (1), it is customary to assume that A is self-adjoint, and that the full nonlinearity is skew-symmetric: $\langle B(z', z), z\rangle_H = 0$, $z, z' \in H$, see [18]. We therefore make the following assumptions on the projected coefficients:

(C1) $A^2_1 = (A^1_2)^*$;
(C2) $\langle B^1_{11}(x', x), x\rangle_{H_d} = 0$, for all $x, x' \in H_d$;
(C3) $\langle B^1_{12}(x', y), x\rangle_{H_d} = -\langle B^2_{11}(x', x), y\rangle_{H_\infty}$, for all $x, x' \in H_d$, $y \in H_\infty$.

Also, without loss of generality, we can assume that B^1_{22} is symmetric in the sense of $\langle B^1_{22}(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d} = \langle B^1_{22}(\mathbf{f}_m, \mathbf{f}_\ell), \mathbf{e}_i\rangle_{H_d}$, for all i, ℓ, m; and finally, we will need the analogue of (A5), that is

(C4) $\sum_{\ell\in\mathbb{N}} \langle B^1_{22}(\mathbf{f}_\ell, \mathbf{f}_\ell), \mathbf{e}_i\rangle_{H_d}\, q_\ell = 0$, for all $i = 1, \ldots, d$.

Note that the latter condition is indeed satisfied for many fluid dynamics models—it usually holds independently of the structure of the noise because $\langle B^1_{22}(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d}$ would be zero on the diagonal, when $\ell = m$, for all i.

Next, we bring Eqs. (29), (30) into a form which makes them comparable to (6), (7).

Using the definition of y_0, we have the following mild formulation of (30),

$$Y^\varepsilon_t = \tilde{Y}^\varepsilon_t + \int_0^t \varepsilon^{-2} e^{-\varepsilon^{-2}(t-s)} \left(A^2_1 X^\varepsilon_s + B^2_{11}(X^\varepsilon_s, X^\varepsilon_s)\right) \mathrm{d}s, \quad t \in [0, T], \tag{31}$$

where

$$\tilde{Y}^\varepsilon_t = \int_{-\infty}^t \varepsilon^{-2} e^{-\varepsilon^{-2}(t-s)} \mathrm{d}W_s, \quad t \in \mathbb{R},$$

is a stationary Ornstein–Uhlenbeck process. Plugging (31) into (29), X^ε alternatively satisfies

$$\begin{aligned} X^\varepsilon_t = x_0 &+ \int_0^t \left(F^1_s + A^1_1 X^\varepsilon_s + B^1_{11}(X^\varepsilon_s, X^\varepsilon_s)\right) \mathrm{d}s + \int_0^t A^1_2 Z^\varepsilon_s \mathrm{d}s + \int_0^t B^1_{12}\left(X^\varepsilon_s, Z^\varepsilon_s\right) \mathrm{d}s \\ &+ \int_0^t A^1_2 \tilde{Y}^\varepsilon_s \mathrm{d}s + \int_0^t B^1_{12}(X^\varepsilon_s, \tilde{Y}^\varepsilon_s) \mathrm{d}s \\ &+ \int_0^t \varepsilon B^1_{22}(\tilde{Y}^\varepsilon_s, \tilde{Y}^\varepsilon_s) \mathrm{d}s + 2\int_0^t \varepsilon B^1_{22}(\tilde{Y}^\varepsilon_s, Z^\varepsilon_s) \mathrm{d}s + \int_0^t \varepsilon B^1_{22}\left(Z^\varepsilon_s, Z^\varepsilon_s\right) \mathrm{d}s, t \in [0, T], \end{aligned} \tag{32}$$

when using the abbreviation

$$Z^\varepsilon_s = \int_0^s \varepsilon^{-2} e^{-\varepsilon^{-2}(s-r)} \left(A^2_1 X^\varepsilon_r + B^2_{11}(X^\varepsilon_r, X^\varepsilon_r)\right) \mathrm{d}r.$$

Since Z^ε_s is close to $A^2_1 X^\varepsilon_s + B^2_{11}(X^\varepsilon_s, X^\varepsilon_s)$, for small ε, and since both terms $B^1_{22}(\tilde{Y}^\varepsilon_s, Z^\varepsilon_s)$, $B^1_{22}\left(Z^\varepsilon_s, Z^\varepsilon_s\right)$ will be shown to vanish with ε, too, the process X^ε should be close to $\tilde{X}^\varepsilon$ satisfying

$$\begin{aligned} \tilde{X}^\varepsilon_t = x_0 &+ \int_0^t \left(F^1_s + A^1_1 \tilde{X}^\varepsilon_s + B^1_{11}(\tilde{X}^\varepsilon_s, \tilde{X}^\varepsilon_s)\right) \mathrm{d}s + \int_0^t A^1_2 \left(A^2_1 \tilde{X}^\varepsilon_s + B^2_{11}(\tilde{X}^\varepsilon_s, \tilde{X}^\varepsilon_s)\right) \mathrm{d}s \\ &+ \int_0^t B^1_{12}\left(\tilde{X}^\varepsilon_s, \left(A^2_1 \tilde{X}^\varepsilon_s + B^2_{11}(\tilde{X}^\varepsilon_s, \tilde{X}^\varepsilon_s)\right)\right) \mathrm{d}s \\ &+ \int_0^t A^1_2 \tilde{Y}^\varepsilon_s \mathrm{d}s + \int_0^t B^1_{12}(\tilde{X}^\varepsilon_s, \tilde{Y}^\varepsilon_s) \mathrm{d}s + \int_0^t \varepsilon B^1_{22}(\tilde{Y}^\varepsilon_s, \tilde{Y}^\varepsilon_s) \mathrm{d}s, \quad t \in [0, T], \end{aligned} \tag{33}$$

which is an equation of type (6) with

$$\begin{aligned} F(t, x) &= F^1_t + A^1_1 x + B^1_{11}(x, x) + A^1_2 \left(A^2_1 x + B^2_{11}(x, x)\right) + B^1_{12}\left(x, \left(A^2_1 x + B^2_{11}(x, x)\right)\right), \\ \sigma(t, x) &= A^1_2 + B^1_{12}(x, \cdot), \\ \beta &= B^1_{22}. \end{aligned}$$

Thus, in this setting, the analogue of (13) would read

$$\begin{aligned}\bar{X}_t =& x_0 + \int_0^t \left(F_s^1 + A_1^1\bar{X}_s + B_{11}^1(\bar{X}_s, \bar{X}_s)\right) \mathrm{d}s + \int_0^t A_2^1 \left(A_1^2\bar{X}_s + B_{11}^2(\bar{X}_s, \bar{X}_s)\right) \mathrm{d}s \\ &+ \int_0^t B_{12}^1 \left(\bar{X}_s, \left(A_1^2\bar{X}_s + B_{11}^2(\bar{X}_s, \bar{X}_s)\right)\right) \mathrm{d}s + \int_0^t C(\bar{X}_s)\,\mathrm{d}s \\ &+ A_2^1 W_t + \int_0^t B_{12}^1(\bar{X}_s, \mathrm{d}W_s) + \sum_{\ell,m\in\mathbb{N}} b_{\ell,m}\bar{W}_t^{\ell,m}, \quad t \in [0, T], \end{aligned} \tag{34}$$

where the Stratonovich correction term $C : H_d \to H_d$ simplifies to

$$\langle C(x), \mathbf{e}_i\rangle_{H_d} = \frac{1}{2}\sum_{m\in\mathbb{N}} q_m \sum_{j=1}^{d} \langle B_{12}^1(\mathbf{e}_j, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d} \langle B_{12}^1(x, \mathbf{f}_m), \mathbf{e}_j\rangle_{H_d}, \quad i = 1, \dots, d,$$

and

$$b_{\ell,m}^i = \langle B_{22}^1(\mathbf{f}_\ell, \mathbf{f}_m), \mathbf{e}_i\rangle_{H_d} \sqrt{\frac{q_\ell q_m}{2}}, \quad i = 1, \dots, d, \ \ell, m \in \mathbb{N}.$$

Proposition 5.1. *When assuming (C1)–(C3), Eq.* (34) *admits a unique global strong solution on* $[0, T]$.

Proof. First, regularity of coefficients guarantees the existence of a unique local strong solution. Second, by Itô's formula,

$$\begin{aligned}\frac{1}{2}|\bar{X}_{t\wedge\tau}|^2 =& \frac{1}{2}|x_0|^2 + \int_0^{t\wedge\tau} \langle F_s^1 + A_1^1\bar{X}_s + B_{11}^1(\bar{X}_s, \bar{X}_s), \bar{X}_s\rangle\,\mathrm{d}s \\ &+ \int_0^{t\wedge\tau} \langle A_2^1\left(A_1^2\bar{X}_s + B_{11}^2(\bar{X}_s, \bar{X}_s)\right), \bar{X}_s\rangle\,\mathrm{d}s \\ &+ \int_0^{t\wedge\tau} \langle B_{12}^1\left(\bar{X}_s, \left(A_1^2\bar{X}_s + B_{11}^2(\bar{X}_s, \bar{X}_s)\right)\right), \bar{X}_s\rangle\,\mathrm{d}s + \int_0^{t\wedge\tau} \langle C(\bar{X}_s), \bar{X}_s\rangle\,\mathrm{d}s \\ &+ \int_0^{t\wedge\tau} \langle A_2^1\mathrm{d}W_s, \bar{X}_s\rangle + \int_0^{t\wedge\tau} \langle B_{12}^1(\bar{X}_s, \mathrm{d}W_s), \bar{X}_s\rangle + \sum_{\ell,m\in\mathbb{N}} \int_0^{t\wedge\tau} \langle b_{\ell,m}, \bar{X}_s\rangle\, d\bar{W}_s^{\ell,m} \\ &+ \frac{1}{2}\sum_{m\in\mathbb{N}} |A_2^1\mathbf{f}_m|^2 q_m(t\wedge\tau) + \frac{1}{2}\sum_{m\in\mathbb{N}} \int_0^{t\wedge\tau} |B_{12}^1(\bar{X}_s, \mathbf{f}_m)|^2 q_m\,\mathrm{d}s + \frac{1}{2}\sum_{\ell,m\in\mathbb{N}} |b_{\ell,m}|^2(t\wedge\tau),\end{aligned}$$

for any fixed $t \in [0, T]$, and any stopping time τ smaller than a possible explosion time.

Applying (C1)–(C3), we have the identities

$$\begin{aligned}&\langle B_{11}^1(\bar{X}_s, \bar{X}_s), \bar{X}_s\rangle_{H_d} = 0, \\ &\langle A_2^1 B_{11}^2(\bar{X}_s, \bar{X}_s), \bar{X}_s\rangle_{H_d} = \langle B_{11}^2(\bar{X}_s, \bar{X}_s), A_1^2\bar{X}_s\rangle_{H_\infty}, \\ &\langle B_{12}^1(\bar{X}_s, A_1^2\bar{X}_s), \bar{X}_s\rangle_{H_d} = -\langle B_{11}^2(\bar{X}_s, \bar{X}_s), A_1^2\bar{X}_s\rangle_{H_\infty}, \\ &\langle B_{12}^1(\bar{X}_s, B_{11}^2(\bar{X}_s, \bar{X}_s)), \bar{X}_s\rangle_{H_d} = -\|B_{11}^2(\bar{X}_s, \bar{X}_s)\|_{H_\infty}^2,\end{aligned}$$

leading to

$$\mathbb{E}\left[\sup_{t'\le t}|\bar{X}_{t'\wedge\tau}|^2\right] \lesssim \left(1+\int_0^t \mathbb{E}\left[\sup_{s'\le s}|\bar{X}_{s'\wedge\tau}|^2\right]\mathrm{d}s\right),$$

again using the regularity of the coefficients combined with Burkholder–Davis–Gundy's inequality. Thus, by Gronwall, the local solution $\bar{X}$ has to be global on $[0, T]$. □

Remark 5.2. In a very similar way, it can be shown that both Eqs. (32) and (33) admit unique global strong solutions on $[0, T]$, too, and hence those proofs are omitted. As a consequence, simply substituting the solution of (32) into (31), for each $\varepsilon > 0$, there is a unique pair of processes $(X^\varepsilon, Y^\varepsilon)$ satisfying (29), (30) on $[0, T]$.

Theorem 5.3. *Assume (C1)–(C3), fix $\varepsilon > 0$, and let $(X^\varepsilon, Y^\varepsilon)$ be the unique pair of processes satisfying* (29), (30) *on a given climate time interval* $[0, T]$.

(i) If (C4), then X^ε converges in law, $\varepsilon \downarrow 0$, to the unique process $\bar{X}$ satisfying (34).

(ii) However, if (C4) comes via $B^1_{22} = 0$, then the stronger convergence (8) *holds true.*

Proof. Recall the process $\tilde{X}^\varepsilon$ satisfying (33), which is an equation of type (6) with coefficients F, σ, β satisfying (A1)–(A3). Furthermore, by Proposition 5.1 and Remark 5.2, condition (A4) is satisfied, too, while (A5) and (C4) actually are the same condition.

All in all, Theorem 2.2 implies that both parts (i) and (ii) of Theorem 5.3 hold true when replacing X^ε by $\tilde{X}^\varepsilon$.

Thus, it is sufficient to prove convergence in probability of $X^\varepsilon - \tilde{X}^\varepsilon$ to zero, $\varepsilon \downarrow 0$, uniformly on compact subsets of a localizing stochastic interval, which can easily be shown following the lines of proof of Theorem 2.2.

Indeed, by localization and discretization arguments, one would first derive

$$\mathbb{E}\left[\sup_{\substack{k'=0,\dots,h\\ k'\Delta\le\tau^\varepsilon_R}}\left|X^\varepsilon_{k'\Delta}-\tilde{X}^\varepsilon_{k'\Delta}\right|^2\right] \lesssim r(\Delta,\varepsilon)+\sum_{k=0}^{h-1}\Delta\mathbb{E}\left[\sup_{\substack{k'=0,\dots,k\\ k'\Delta\le\tau^\varepsilon_R}}\left|X^\varepsilon_{k'\Delta}-\tilde{X}^\varepsilon_{k'\Delta}\right|^2\right],$$
$$h = 1,\dots,[T/\Delta],$$

where $\tau^\varepsilon_R = \inf\{t \ge 0 : |X^\varepsilon_t| \ge R\} \wedge \inf\{t \ge 0 : |\tilde{X}^\varepsilon_t| \ge R\}$, and $r(\Delta, \varepsilon) \to 0$, $\varepsilon \downarrow 0$, for a suitable choice of $\Delta = \Delta_\varepsilon$. Then, combining Gronwall's lemma and Markov's inequality, one would obtain

$$\lim_{\varepsilon\to 0}\mathbb{P}\left\{\sup_{t\le T\wedge\tau^\varepsilon_R}\|X^\varepsilon_t-\tilde{X}^\varepsilon_t\|_{H_d}>\delta\right\}=0,\quad \forall\,\delta>0,$$

which yields the convergences stated in parts (i) and (ii) of Theorem 5.3 up to time τ^ε_R. Since $\bar{X}$ is globally defined, both types of convergence can be extended to the whole

interval $[0, T]$, using similar arguments given in the proof of the corresponding parts of Theorem 2.2. □

Funding Information Open access funding provided by Scuola Normale Superiore within the CRUI-CARE Agreement.

REFERENCES

[1] R.N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 60, 185–201, 1982.

[2] Z. Brzeźniak, M. Capiński, and F. Flandoli. A convergence result for stochastic partial differential equations. Stochastics, 24(4):423–445, 1988.

[3] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications* 2nd edn. Cambridge University Press, Cambridge, 2014.

[4] S. N. Ethier and T. G. Kurtz. *Markov processes—characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[5] F. Flandoli and U. Pappalettera. 2D Euler equations with Stratonovich transport noise as a large scale stochastic model reduction. *J. Nonlinear Sci.*, 31:24, 2021.

[6] C. Franzke and A. J. Majda. Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.*, 63(2):457–479, 2006.

[7] C. Franzke, A. J. Majda, and E. Vanden-Eijnden. Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.*, 62(6):1722–1745, 2005.

[8] M.I. Freidlin and A.D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. Springer New York, 2012.

[9] K. Hasselmann. Stochastic climate models Part I. Theory. *Tellus*, 28:473–485, 1976.

[10] B. Iftimie, È. Pardoux, and A. Piatnitski. Homogenization of a singular random one-dimensional pde. *Ann. Inst. H. Poincarè Probab. Statist.*, 44(3):519–543, 06 2008.

[11] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. North-Holland Publishing Co., second edition, 1989.

[12] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes* Grundlehren der mathematischen Wissenschaften 288. Springer, Berlin, second edition, 2002.

[13] A. Jain, I. Timofeyev, and E. Vanden-Eijnden. Stochastic mode-reduction in models with conservative fast sub-systems. *Commun. Math. Sci.*, 13(2):297–314, 2015.

[14] S. Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[15] C. Jia and G. Zhao. Moderate maximal inequalities for the Ornstein-Uhlenbeck process. *Proc. Amer. Math. Soc.*,https://doi.org/10.1090/proc/14804, 2020.

[16] T. G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Functional Analysis*, 12:55–67, 1973.

[17] A. J. Majda, I. Timofeyev, and E. Vanden Eijnden. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, 54(8):891–974, 2001.
[18] A. J. Majda and X. Wang. *Non-linear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press, Cambridge, 2006.
[19] S. Olla. Homogenization of diffusion processes in random fields, 1994.
[20] E. Pardoux and Yu. Veretennikov. On the Poisson Equation and Diffusion Approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.
[21] E. Pardoux and Yu. Veretennikov. On the Poisson Equation and Diffusion Approximation. II. *Ann. Probab.*, 31(3):1166–1192, 2003.
[22] C. Penland and L. Matrosova. A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation. *Journal of Climate*, 7(9):1352–1372, 09 1994.
[23] G. Tessitore and J. Zabczyk. Wong-Zakai approximations of stochastic evolution equations. *Journal of Evolution Equations*, 6(4):621–655, 2006.
[24] K. Twardowska. Approximation theorems of Wong-Zakai type for stochastic differential equations in infinite dimensions. *Dissertationes Math. (Rozprawy Mat.)*, 325, 1993.
[25] G. K. Vallis. *Atmospheric and oceanic fluid dynamics : fundamentals and large-scale circulation*. Cambridge University Press, first edition, November 2006.
[26] E. Wong and M. Zakai. On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Statist.*, 36(5):1560–1564, 10 1965.

Sigurd Assing
Department of Statistics
The University of Warwick
Coventry
CV4 7AL
UK
E-mail: s.assing@warwick.ac.uk

Franco Flandoli and Umberto Pappalettera
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa
Italy
E-mail: franco.flandoli@sns.it

Umberto Pappalettera
E-mail: umberto.pappalettera@sns.it

Accepted: 15 April 2021