QUANTIFYING A CONVERGENCE THEOREM OF GYÖNGY AND KRYLOV

BY KONSTANTINOS DAREIOTIS[1,a], MÁTÉ GERENCSÉR[2,b] AND KHOA LÊ[3,c]

[1]*School of Mathematics, University of Leeds,* [a]*k.dareiotis@leeds.ac.uk*

[2]*Institute of Analysis and Scientific Computing, TU Wien,* [b]*mate.gerencser@tuwien.ac.at*

[3]*Institute für Mathematik, TU Berlin,* [c]*le@math.tu-berlin.de*

We derive sharp strong convergence rates for the Euler–Maruyama scheme approximating multidimensional SDEs with multiplicative noise without imposing *any* regularity condition on the drift coefficient. In case the noise is additive, we show that Sobolev regularity can be leveraged to obtain improved rate: drifts with regularity of order $\alpha \in (0, 1)$ lead to rate $(1 + \alpha)/2$.

1. Introduction. The present article studies stochastic differential equations (SDEs) of the form

$$(1) \qquad dX_t = b(X_t)\, dt + \sigma(X_t)\, dB_t, \quad X_0 = x_0,$$

and their equidistant Euler–Maruyama approximations

$$(2) \qquad dX_t^n = b\big(X_{\kappa_n(t)}^n\big)\, dt + \sigma\big(X_{\kappa_n(t)}^n\big)\, dB_t, \quad X_0^n = x_0,$$

with the notation $\kappa_n(t) = \lfloor nt \rfloor / n$. Here the initial condition is $x_0 \in \mathbb{R}^d$; the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ are measurable functions; B is a d-dimensional standard Brownian motion on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbf{P})$; and the dimension $d \in \mathbb{N}$ is arbitrary. When the coefficients are Lipschitz continuous, the convergence of X^n to X, as well as the rate of convergence, is very well understood. In this article we are interested in the regime where the drift b is far from Lipschitz, maybe not even continuous. In fact, our first result, Theorem 1.2, can be summarised as yielding sharp strong convergence rate without *any* continuity (or even local Sobolev regularity) assumption on b.

Let us recall that when b is irregular, the existence and uniqueness of solutions of (1) relies on the regularising effects of the noise, and therefore some form of nondegeneracy of the diffusion coefficient is necessary. Under the assumption that σ is uniformly elliptic and sufficiently regular, the well-known result of Veretennikov [41] states that (1) is well-posed even with merely bounded and measurable drift coefficient b. However, Veretennikov's proof, using the Yamada–Watanabe principle, was not constructive and did not have any implications on the stability of the solution with respect to the Euler–Maruyama approximations X^n. It took 16 years until the seminal work of Gyöngy and Krylov [14] for the first proof that, under the same weak assumptions, X^n converges in probability to X. This is in stark contrast with the case of regular coefficients, where the well-posedness and the convergence of the Euler–Maruyama approximations follow from essentially the same arguments (which are straightforward applications of the Burkholder–Davis–Gundy and Gronwall's inequalities). The result of Gyöngy and Krylov was qualitative, that is, it provided no rates of convergence. Despite significant interest (see Section 1.1 below), there has not been *any* known upper bound for the error $|X_n - X|$ in the case of bounded measurable b. Our first result, Theorem 1.2 not only does provide an upper bound but also it actually shows that this bound is of

order $n^{-1/2+\varepsilon}$ with arbitrary $\varepsilon > 0$, which is known to be sharp even in the case of smooth coefficients (see [18, 25]).

The optimality of the rate $1/2$ no longer holds when the noise is additive. If σ is simply the identity matrix, then for smooth b the rate of convergence is known to be 1. We are therefore interested in how much, and what kind of, regularity assumption is needed on b to improve the rate of convergence. Our second result, Theorem 1.5 establishes rate $(1 + \alpha)/2$, provided that the drift possesses Sobolev regularity of order $\alpha \in (0, 1)$ with integrability exponent $p \geq \max\{2, d\}$.

The rest of the article is structured as follows. In the remainder of the introduction we briefly overview the relevant literature (Section 1.1), highlight the main aspects of the proof (Section 1.2) and state the main results of the paper (Section 1.3). In Section 2 we introduce the notation and collect/prove a number of auxiliary statements. Section 3 is concerned with some quadrature estimates, which essentially provide the rate of convergence. In Section 4 a general stability estimate is given for approximate solutions of (1), which is to be applied with X and X^n. Section 5 combines all the previous ingredients to provide the proofs of the main results.

1.1. *Literature.*

The strong error analysis of the Euler–Maruyama scheme for SDEs with irregular coefficients has attracted significant attention in recent years. In addition to being a developing branch of the field of "regularisation by noise", it has the practical relevance that SDEs with discontinuous drifts have recently been utilised in applied sciences. As a few examples, they are used in finance for modelling equity markets (see, [17]), in neuroscience for modelling interacting neurons that follow integrate-and-fire type dynamics (see, [12]), and also in the modelling of energy storage problems (see, [39]).

In one direction, see among others [4, 5, 10, 30, 31, 36], low regularity of order $\alpha \in (0, 1)$ of the drift on the Hölder scale is assumed. In these works the regularising properties of the noise at the level of the dual parabolic partial differential equation (PDE) are used in order to close the estimates. This is a technique which originates in the work of Zvonkin and Veretennikov [41, 42]. The version that is most suitable in the setting of numerics is due to Flandoli, Gubinelli and Priola [11], and is known as the *Itô–Tanaka trick*. This method led to various results on the rate of convergence in the case of irregular drift. Concerning (1), in [30] the authors derived a rate proportional to the regularity of the drift, namely, rate $\alpha/2$ for α-Hölder continuous drift b, $\alpha \in (0, 1)$. Notice that this seemed to be consistent with the regular case (see, e.g., [37], Theorem 7.5, where for additive noise the rate $\alpha/2$ is derived for $\alpha \in [1, 2]$). However, it turned out that this rate is suboptimal: in [9] it is shown that in the additive noise case the rate $1/2$ is achieved even for Dini-continuous b, and in dimension $d = 1$, for merely bounded b. The main reason behind this improvement is that one can also leverage the regularising effects of the noise at a purely numerical analytic level: namely, it leads to sharp quadrature-type estimates. For earlier works on such estimates, see [2, 19, 35]. These bounds are closely related to the error analysis of the Euler–Maruyama scheme.

Another direction is explored in, among others, [28, 29, 32, 33], where the irregularity on b is assumed to take the form of discontinuities at finitely many points (or, in higher dimensions, hypersurfaces), outside of which the usual regularity assumptions are imposed. For a detailed account on the development here, including other approximation schemes, we refer to the introduction of Neuenkirch and Szölgyenyi [34]. Another relevant feature of [34] is that, like in [9], the regularising properties of the noise are exploited on two (PDE and quadrature) levels. With this method, in the case of dimension $d = 1$ and additive noise, the authors show that for $b \in W_2^\alpha \cap L_\infty \cap L_1$, for $\alpha \in (0, 1)$, the rate $\min\{(1 + \alpha)/2, 3/4\}$ can be achieved by the Euler–Maruyama scheme. Since piecewise Lipschitz functions belong (at least locally) to $W_2^{1/2-\varepsilon}$ for any $\varepsilon > 0$, this result generalises several previous ones. In the

scalar piecewise Lipschitz case lower bounds have also been obtained by Müller–Gronbach and Yaroslavtseva [33], showing that the rate 3/4 is sharp.

Finally, we mention the recent work [7], which also works on the Hölder scale, but with quite different methods from the above. Instead of relying on PDE theory, it introduces an approach based on stochastic sewing [26]. This approach not only leads to improved rates (in case $b \in C^\alpha$ with additive noise, one gets rate $(1+\alpha)/2$ in [7] vs. $1/2$ in [9] vs. $\alpha/2$ in [30]), but also widely extends the scope of driving noise, covering non-Markovian examples like fractional Brownian motions.

The contributions of the present article in relation to the existing results are as follows.

- In the multiplicative noise case Theorem 1.2 provides the first, and at the same time, sharp, quantification of the qualitative theorem of Gyöngy and Krylov [14]. Previous results imposed, in one way or another, positive regularity (e.g., Hölder [7] or—implicitly—Sobolev [29]). The case of 0 regularity is critical from an analytic point of view, as demonstrated by the fact that up to now not even suboptimal rates were available in this borderline case.
- In the additive noise case Theorem 1.5 shows that if the drift does have additional regularity on the Sobolev scale, then the rate of convergence is also improved. In comparison to [7], regularity is assumed on the Sobolev, rather than the Hölder scale, allowing discontinuous coefficients. Even when $b \in W_p^\alpha$ with $p > d/\alpha$ and therefore one has by Sobolev embedding $b \in C^{\alpha - d/p}$, the present results cannot be recovered from [7]. Indeed, [7] would imply rate $(1+\alpha - d/p)/2$ in this case, while Theorem 1.5 shows rate $(1+\alpha)/2$. In comparison to [34] which considers drifts from Sobolev spaces, our methods give stronger results in two main directions. First, for drifts with Sobolev regularity α the Euler–Maruyama scheme is shown to have rate $(1+\alpha)/2$, which removes the 3/4 threshold from the result of [34]. Second, Theorem 1.5 is valid in all dimensions, in contrast to the restriction $d = 1$ from [34]. Other improvements include higher moment estimates (provided b has Sobolev regularity with higher integrability exponent) and uniform in time error estimates.

1.2. *On the proof.* We now briefly outline the strategy of the proof, also highlighting the differences/similarities to previous works. For simplicity let us consider the additive case, with constant identity diffusion matrix. One has the standard error decomposition

$$(3) \qquad X_t - X_t^n = \int_0^t \left(b(X_s) - b(X_s^n) \right) ds + \int_0^t \left(b(X_s^n) - b(X_{\kappa_n(s)}^n) \right) ds.$$

To have any chance of obtaining a Gronwall-type estimate, one would like to bound the first integral by $\|X - X^n\|$, in some norm $\| \cdot \|$. In all but one of the above-mentioned works this is achieved by means of PDE techniques, using what is known as the Itô–Tanaka trick from [11], which in turn is a variation of the methods of [41, 42]. Recently, in [7] a new approach was introduced to obtain such bounds, based on the sewing methods inspired by [8, 26]. Coincidentally, the sewing method also turns out to be efficient in handling the second integral in (3). For example, if one takes the coefficient b from the Hölder space C^α, $\alpha \in (0, 1)$, then instead of the naive bound of order $\|b\|_{C^\alpha} n^{-\alpha/2}$ one can obtain a bound of order $\|b\|_{C^\alpha} n^{-(1+\alpha)/2}$; see [7], Lemma 4.2.

The lack of regularity of b poses difficulties in both steps. Therefore, in the present article we follow a hybrid path: the first term in (3) is treated by the PDE approach and the second one is estimated by stochastic sewing. Concerning the PDE step, in the Zvonkin/Itô–Tanaka transformation the solution of the corresponding PDE (see (27)) does not have bounded second derivatives, which makes closing the estimates with Gronwall-type argument problematic. This is resolved by relying instead on local L_p-bounds on the second derivatives, replacing Lipschitz bounds with the Hardy–Littlewood maximal inequality, and finally controlling the localisation error. While localisation steps are quite common in the literature when it

comes to uniqueness of SDEs (usually appearing as a simple stopping times argument), they have to be made quantitative for the error analysis. Let us also mention that the use of Hardy–Littlewood maximal inequality in a similar context (albeit without localisation) also appeared in the recent work [5]. Concerning the stochastic sewing step, the main novelty compared to [7] is to exploit Sobolev regularity in estimating approximation errors for "occupation time functionals", that is, bounding quantities of the form

$$\int_0^t \left(f(B_s) - f(B_{\kappa_n(s)}) \right) ds,$$

with f belonging to certain Sobolev space, see Lemma 3.3. This requires a version of the stochastic sewing lemma of [26] allowing singularities, by introducing temporal weights, see Lemma 2.9.

The hybrid approach described above seems useful and efficient. In particular, it has been applied and extended in [27] to obtain sharp rates of convergence of tamed Euler schemes for SDEs with integrable drifts, as considered in Krylov–Röckner [24].

1.3. *Main results.*

ASSUMPTION 1.1. Assume that b is bounded, σ is twice differentiable, its derivatives of order 0, 1, and 2 are bounded functions, and that for some $\lambda > 0$, one has $|y^*(\sigma\sigma^*)(x)y| \geq \lambda^2 |y|^2$ for all $x, y \in \mathbb{R}^d$.

Recall that X and X^n are defined as the solutions of (1) and (2), respectively. The ellipticity of σ is assumed in order to ensure solvability when b is irregular. The fact that under Assumption 1.1 (and therefore also under the stronger Assumption 1.4 below) these solutions exist and are unique, follows immediately from [41]. The nondegeneracy of the noise can also be highly relevant in the context of approximations: as shown [15], even if all coefficients are globally bounded with locally bounded derivatives of any order, with degenerate diffusion coefficient the strong rate of convergence may be worse than any positive power. Our results also imply that such behaviour can be excluded for elliptic diffusion, and see the discussion in Section 1.4 below how the results can be extended to a certain class of degeneracies. The first main result of the article is the following.

THEOREM 1.2. *Given Assumption 1.1, let $\varepsilon \in (0, 1)$, $p \in (0, \infty)$. Then for all $n \in \mathbb{N}$ the following bound holds*:

$$(4) \qquad \left(\mathbf{E} \sup_{t \in [0,1]} |X_t - X_t^n|^p \right)^{1/p} \leq N n^{-1/2+\varepsilon},$$

with some constant N depending only on d, p, ε, λ, and $\sup(|b| + |\sigma| + |\nabla\sigma| + |\nabla^2\sigma|)$.

In the additive case, an even higher rate can be shown given some Sobolev regularity of the drift b. The homogeneous Sobolev spaces $\dot{W}_p^\alpha(\mathbb{R}^d)$, for $\alpha \in (0, 1)$ and $p \in [1, \infty)$, are defined as usual: they contain all measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$(5) \qquad [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)} := \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x) - f(y)|^p}{|x - y|^{d+\alpha p}} \, dx \, dy \right)^{1/p} < \infty.$$

This definition obviously extends to finite-dimensional vector-valued functions.

REMARK 1.3. If $f \in \dot{W}_m^\alpha(\mathbb{R}^d)$ and $\alpha > d/m$ then f has a version that is continuous and $[f]_{C^{\alpha-d/m}(\mathbb{R}^d)} \leq N(d, \alpha, m)[f]_{\dot{W}_m^\alpha(\mathbb{R}^d)}$ (see, e.g., [22], Lemma 2 and Remark 3, p. 203–206). If $\alpha > d/m$, then the elements of $\dot{W}_m^\alpha(\mathbb{R}^d)$ will be treated as continuous functions rather than equivalence classes. In particular, if we write that $b \in \dot{W}_m^\alpha(\mathbb{R}^d)$ (as in Assumption (1.4)) and $\alpha > d/m$, then we automatically mean that, in addition, $b \in C^{\alpha-d/m}$.

ASSUMPTION 1.4. Assume that $\sigma = I$ and b is a bounded measurable function which belongs to $\dot{W}_m^\alpha(\mathbb{R}^d)$ for some $\alpha \in (0, 1)$ and $m \geq \max(d, 2)$.

THEOREM 1.5. *Let Assumption 1.4 hold, let $\varepsilon \in (0, 1)$, $p \in (0, m]$. Then for all $n \in \mathbb{N}$ the following bound holds*:

$$(6) \qquad \left(\mathbf{E} \sup_{t \in [0,1]} |X_t - X_t^n|^p\right)^{1/p} \leq Nn^{-(1+\alpha)/2+\varepsilon},$$

with some constant N depending only on $d, p, \varepsilon, \alpha, m$, $\sup|b|$, and $[b]_{\dot{W}_m^\alpha(\mathbb{R}^d)}$.

REMARK 1.6. In the rest of the paper we will actually assume strict inequality in the conditions $m \geq \max(d, 2)$, and $m \geq p$. To see that this is not a restriction, by Sobolev embeddings we can always slightly increase m at the price of slightly decreasing α. In the situation of Assumption 1.4 even a better embedding is available, namely that b belongs to $\dot{W}_{m/\theta}^{\theta\alpha}$ for any $\theta \in (0, 1)$, see Lemma 2.1. This means a slight loss in the rate, which can just be absorbed in the ε.

REMARK 1.7. Note that there is no moment restriction in Theorem 1.2 but there is the restriction $p \leq m$ in Theorem 1.5. One can increase the moments by sacrificing from the rates, using again the embedding from Lemma 2.1.

As mentioned before, one interesting class of coefficients with Sobolev regularity is that of indicator functions of regular domains. Indeed, it is well known (see, e.g., [40], Section 3.2) that if $D \subset \mathbb{R}^d$ is a bounded Lipschitz domain, then $\mathbf{1}_D \in \dot{W}_p^{1/p-\varepsilon}$ for every $p \in [1, \infty)$, $\varepsilon > 0$. Since multiplying with a bounded Lipschitz function leaves $\dot{W}_p^{1/p-\varepsilon}$ invariant, one gets the following corollary.

COROLLARY 1.8. *Let $\sigma = I$, $\varepsilon \in (0, 1)$, and $p \in (0, \infty)$. Assume that with a finite set of bounded Lipschitz domains $D_1, \ldots, D_k$ and bounded Lipschitz continuous functions $f_1, \ldots, f_k$, b is of the form*

$$b(x) = \sum_{i=1}^{k} f_i(x)\mathbf{1}_{D_i}(x).$$

Then for all $n \in \mathbb{N}$ the following bound holds:

$$\left(\mathbf{E} \sup_{t \in [0,1]} |X_t - X_t^n|^p\right)^{1/p} \leq Nn^{-\frac{1}{2}(1+\frac{1}{\max(2,d,p)})+\varepsilon},$$

with some constant N.

PROOF. By the preceding remarks, b belongs to $\dot{W}_m^{1/m-\varepsilon'}$ for $m = \max(2, d, p)$ and for every $\varepsilon' > 0$, hence satisfying Assumption 1.4. The result follows immediately from Theorem 1.5. □

REMARK 1.9. In the special case of $d = 1$, Corollary 1.8 yields the L_2-rate $3/4 - \varepsilon$. As mentioned above, this is known to be sharp, see [33]. We do not know whether in general dimensions $d > 1$ the L_2-rate $(d + 1)/(2d)$ is sharp, but it is certainly the best known bound at the moment.

1.4. *Localising irregularities.* Let us now briefly outline how to some extent the results can be extended to coefficients with degeneracy and/or growth, provided these properties do not "interfere" with the irregularities.

Let A^1, A^2 be two Lipschitz domains such that $\bar{A}^1 \cup \bar{A}^2 = \mathbb{R}^d$. Take some $\delta > 0$ and denote the δ-fattening of A^i by A^i_δ for $i = 1, 2$. We assume that there exist (globally defined) coefficients b^i, σ^i such that $b = b^i$ and $\sigma = \sigma^i$ on $A^i_{3\delta}$, and that furthermore the corresponding SDEs and their Euler approximations satisfy the strong Markov property and for all $n \in \mathbb{N}$ the bound

$$\mathbf{E} \sup_{t \in [0,1]} |X^i_t - X^{i,n}_t|^p \leq K n^{-\alpha}$$

holds for some $K < \infty$, $p \in [1, \infty)$, and $\alpha > 0$ uniformly over initial conditions. Furthermore, assume the a priori estimates

$$\mathbf{E} \sup_{t \in [0,1]} |X_t|^q + \sup_{n \in \mathbb{N}} \mathbf{E} \|X^n\|^q_{C^\kappa[0,1]} \leq K(q)$$

for some $\kappa \in (0, 1]$ and for all $q \in [1, \infty)$. One may think of an example where b is bounded and σ is elliptic on a ball and both coefficients are Lipschitz but possibly degenerate and have some (linear) growth on the complement of a smaller ball (and therefore "best of both worlds" on the intersection). We then claim that

$$\mathbf{E} \sup_{t \in [0,1]} |X_t - X^n_t|^p \leq N n^{-\alpha+\varepsilon}$$

for any $\varepsilon > 0$, where $N = N(K, (K(q))_{q \in [1,\infty)}, p, \varepsilon, \delta)$. Without loss of generality we assume $x_0 \in \bar{A}^1$, and we define the stopping times $\tau_0 = 0$, and for $k = 0, 1, \ldots,$

$$\tau_{2k+1} = \inf\{t \geq \tau_{2k} : tn \in \mathbb{N}, X^n_t \notin A^1_\delta\}, \qquad \tau_{2k+2} = \inf\{t \geq \tau_{2k+1} : tn \in \mathbb{N}, X^n_t \notin A^2_\delta\}.$$

For any $q \in (1, \infty)$, with dual exponent $q' = q/(q-1)$ we have by the assumptions and standard use of Hölder's and Markov's inequalities

$$\mathbf{E} \sup_{t \in [0,\tau_1]} |X_t - X^n_t|^p$$

$$\lesssim \mathbf{E} \sup_{t \in [0,\tau_1]} |X^1_t - X^{1,n}_t|^p + \left(\mathbf{P}(X^1 \text{ exits } A^1_{3\delta} \text{ before } \tau_1)\right)^{1/q'}$$

$$\lesssim n^{-\alpha} + \left(\mathbf{P}(X^{n,1} \text{ exits } A^1_{2\delta} \text{ before } \tau_1)\right)^{1/q'} + \left(\mathbf{P}\left(\sup_{t \in [0,\tau_1]} |X^1_t - X^{1,n}_t| \geq \delta\right)\right)^{1/q'}$$

$$\lesssim n^{-\alpha} + \left(\mathbf{P}(\|X^n\|_{C^\kappa[0,1]} \geq \delta n^\kappa)\right)^{1/q'} + n^{-\alpha/q'}$$

$$\lesssim n^{-\alpha/q'} + n^{-\kappa q/q'}.$$

The proportionality constants in $\lesssim$ always depend only on the parameters of N mentioned above and q. Choosing q large enough we get a bound of order $n^{-\alpha+\varepsilon}$. Similarly, by the strong Markov property, we have for any $k \geq 0$ that

$$\mathbf{E} \sup_{t \in [\tau_k, \tau_{k+1}]} |X_t - X^n_t|^p \lesssim n^{-\alpha+\varepsilon}.$$

Therefore, for any $m \in \mathbb{N}$ and $q \in (1, \infty)$,

$$\mathbf{E} \sup_{t \in [0,1]} |X_t - X^n_t|^p \lesssim \sum_{k=0}^{m-1} \mathbf{E} \sup_{t \in [\tau_k, \tau_{k+1}]} |X_t - X^n_t|^p + \mathbf{E}\left(\sup_{t \in [0,1]} |X_t - X^n_t|^p \mathbf{1}_{(\tau_m < 1)}\right)$$

$$\lesssim m n^{-\alpha+\varepsilon} + \left(\mathbf{P}(\tau_m < 1)\right)^{1/q'}$$

$$\lesssim m n^{-\alpha+\varepsilon} + \left(\mathbf{P}(\|X^n\|_{C^\kappa[0,1]} \geq \delta m^\kappa)\right)^{1/q'}$$

$$\lesssim m n^{-\alpha+\varepsilon} + m^{-\kappa q/q'}.$$

Choosing $m = n^\varepsilon$ and then q large enough, we get the claim with 2ε in place of ε.

2. Preliminaries.

2.1. *Notation.*

Function spaces. For any function $f : Q \to V$, where $Q \subset \mathbb{R}^k$ is a Borel set and $(V, |\cdot|)$ is a normed space, with the notation $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, let us set the (semi-)norms

$$\|f\|_{C^0(Q,V)} = \sup_{x \in Q} |f(x)|;$$

$$[f]_{C^\gamma(Q,V)} = \sum_{\substack{\ell \in \mathbb{N}_0^k \\ |\ell|_1 = \hat{\gamma}}} \sup_{x \neq y \in Q} \frac{|\partial^\ell f(x) - \partial^\ell f(y)|}{|x-y|^{\bar{\gamma}}}, \quad \gamma > 0, \gamma = \hat{\gamma} + \bar{\gamma}, \hat{\gamma} \in \mathbb{N}_0, \bar{\gamma} \in (0,1];$$

$$\|f\|_{C^\gamma(Q,V)} = \sum_{\substack{\ell \in \mathbb{N}_0^k \\ |\ell|_1 < \gamma}} \|\partial^\ell f\|_{C^0(Q,V)} + [f]_{C^\gamma(Q,V)}, \quad \gamma > 0.$$

In the above $|\ell|_1 = \ell_1 + \cdots + \ell_k$ if $\ell = (\ell_1, \ldots, \ell_k) \in \mathbb{N}_0^k$. By $C^\gamma(Q, V)$ we denote the space of all measurable functions $f : Q \to V$ such that $\|f\|_{C^\gamma(Q,V)} < \infty$.

We will also denote $\mathbb{B} = C^0$, emphasizing that elements of C^0 need not be continuous, but only bounded measurable (and avoiding the confusion with the space of bounded continuous functions). In addition, notice that the elements of $\mathbb{B}$ are function rather than equivalent classes. Similarly, for $k \in \mathbb{N}$, C^k functions are $(k-1)$-times continuously differentiable and the derivatives of $(k-1)$-order are Lipschitz continuous. For $\gamma \in \mathbb{R}_+ \setminus \mathbb{N}_0$, C^γ are of course the usual Hölder spaces. When the domain Q is $\mathbb{R}^d$ and/or the target space V is $\mathbb{R}^d$ or $\mathbb{R}^{d \times d}$, they are suppressed from the notation. Moreover, in $\mathbb{R}^d$ or $\mathbb{R}^{d \times d}$, $|\cdot|$ is understood to be the Euclidean norm. For $\gamma < 0$, we denote by C^γ the space of all tempered distributions f such that

$$\|f\|_{C^\gamma} := \sup_{t \in (0,1)} t^{-\gamma/2} \|\mathscr{P}_t f\|_{L_\infty(\mathbb{R}^d)} < \infty,$$

where $(\mathscr{P}_t)_{t \geq 0}$ is the heat semigroup associated to the standard Gaussian kernel $p_t(x) = (2\pi t)^{-d/2} e^{-|x|^2/(2t)}$. Notice that for $\gamma_1, \gamma_2 \in \mathbb{R}$ with $\gamma_1 < \gamma_2$ we have the continuous embedding $C^{\gamma_2} \hookrightarrow C^{\gamma_1}$. Recall the definition of $\dot{W}_p^\alpha$ from (5). The following simple property is used in some of the remarks in the Introduction.

LEMMA 2.1. *For $\alpha, \theta \in (0, 1)$, $m \in [1, \infty)$, one has the inclusion $\mathbb{B} \cap \dot{W}_m^\alpha \subset \mathbb{B} \cap \dot{W}_{m/\theta}^{\alpha\theta}$.*

PROOF. Let f be a bounded function in $\dot{W}_p^\alpha$. We have

$$[f]_{\dot{W}_{m/\theta}^{\alpha\theta}(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x) - f(y)|^{m/\theta}}{|x-y|^{d+\alpha m}} \, dx \, dy\right)^{\theta/m},$$

$$\leq 2\|f\|_{\mathbb{B}}^{1-\theta} \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x) - f(y)|^m}{|x-y|^{d+\alpha m}} \, dx \, dy\right)^{\theta/m},$$

which means that $[f]_{\dot{W}_{m/\theta}^{\alpha\theta}(\mathbb{R}^d)} \leq 2\|f\|_{\mathbb{B}}^{1-\theta} [f]_{\dot{W}_m^\alpha(\mathbb{R}^d)}^\theta$, completing the proof. $\square$

Matrices. We use the following conventions and basic properties of $d \times d$ matrices. By I we denote the identity matrix and by A^* the transpose of A. The operator norm, determinant and trace of A are denoted respectively by $\|A\|$, $\det(A)$ and $\text{Tr}(A)$. For two symmetric matrices A_1, A_2, by $A_1 \prec A_2$ we mean that $A_2 - A_1$ is nonnegative definite. Recall that the determinant is a differentiable function in a neighbourhood of the identity matrix, and therefore there exists a constant $N = N(d)$ such that $|1 - \det(I + A)| \le N\|A\|$ for $\|A\| \le 1/2$. As a simple consequence, for any fixed K, there exists a constant $N = N(d, K)$ such that on the set $\{A : \|A\| \le K\}$, one has

$$(7) \qquad \left|1 - |\det(I + A)|^{1/2}\right| \le N\|A\|.$$

Indeed, by the previously mentioned bound there exists a $\delta = \delta(d)$ such that $|1 - \det(I + A)| \le 1/2$ for $\|A\| \le \delta$. Since the square root function is Lipschitz on $[1/2, 3/2]$ (even with Lipschitz constant 1), for all A with $\|A\| \le \delta$ we have

$$\left|1 - |\det(I + A)|^{1/2}\right| \le |1 - \det(I + A)| \le N(d)\|A\|.$$

This is the claimed bound for $\|A\| \le \delta$. On the other hand, if $K \ge \delta$ and $\delta \le \|A\| \le K$, then trivially we have

$$\left|1 - |\det(I + A)|^{1/2}\right| \le 1 + N'(d)\|I + A\|^{d/2} \le N''(d, \delta, K)\|A\|.$$

Convention on constants. In proofs (and only in proofs) of theorems/lemmas/propositions we use the shorthand $f \lesssim g$ to mean that there exists a constant N such that $f \le Ng$, and that N does *not* depend on any other parameters than the ones specified in the theorem/lemma/proposition. Whenever a constant depends on any other parameters, they are indicated by parentheses in $N(\cdot)$.

2.2. *Heat kernel bounds.* For a symmetric positive definite $d \times d$ matrix Σ, let p_Σ be the density of a d-dimensional Gaussian vector with mean zero and covariance matrix Σ defined by

$$(8) \qquad p_\Sigma(x) = \frac{(\det \Sigma^{-1})^{1/2}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}x^* \Sigma^{-1} x\right), \quad x \in \mathbb{R}^d.$$

Let $\Sigma^{1/2}$ be a square matrix such that $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^*$ and introduce the notation $x_\Sigma = \Sigma^{-1/2}x$ where $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$. The exponential in (8) can be rewritten as $\exp(-\frac{1}{2}|x_\Sigma|^2)$. It is then straightforward to see that for any $k \ge 0$,

$$(9) \qquad |x_\Sigma|^k p_\Sigma(x) \le N p_{\Sigma/2}(x),$$

where N depends only on k and d. For $t > 0$, we use the shorthand $p_t = p_{tI}$. For a measurable function $f : \mathbb{R}^d \to \mathbb{R}$ we write $\mathcal{P}_\Sigma f := p_\Sigma * f$ and $\mathcal{P}_t f := p_t * f$. For all $t > 0$, $\theta \in [1, \infty]$, and $\alpha \ge 0$ one has the bounds

$$(10) \qquad \|p_t\|_{L_\theta(\mathbb{R}^d)} \le N t^{-\frac{d}{2}(1 - \frac{1}{\theta})},$$

$$(11) \qquad \|\nabla p_t\|_{L_\theta(\mathbb{R}^d)} \le N t^{-\frac{1}{2} - \frac{d}{2}(1 - \frac{1}{\theta})},$$

$$(12) \qquad \||\cdot|^\alpha p_t(\cdot)\|_{L_\theta(\mathbb{R}^d)} \le N' t^{\frac{\alpha}{2} - \frac{d}{2}(1 - \frac{1}{\theta})},$$

$$(13) \qquad \||\cdot|^\alpha \nabla^2 p_t(\cdot)\|_{L_\theta(\mathbb{R}^d)} \le N' t^{\frac{\alpha}{2} - 1 - \frac{d}{2}(1 - \frac{1}{\theta})},$$

with some constants $N = N(d, \theta)$ and $N' = N'(\alpha, d, \theta)$.

LEMMA 2.2. *Let $p \in [1, \infty)$ and $\alpha \in (0, 1)$. Then, for all $f \in \dot{W}_p^\alpha(\mathbb{R}^d)$, $t \in (0, 1]$,*

$$\|\nabla^2 \mathscr{P}_t f\|_{L_p(\mathbb{R}^d)} \le N t^{-1+\frac{\alpha}{2}} [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)},$$

where N depends only on α, p, and d.

PROOF. For $p, q \in [1, \infty)$, $\frac{1}{q} + \frac{1}{p} = 1$, by Hölder's inequality, we have for any $\beta > 0$

$$\|\nabla^2 \mathscr{P}_t f\|_{L_p(\mathbb{R}^d)}^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (\nabla^2 p_t)(y) f(x - y) \, dy \right|^p dx$$

$$= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (\nabla^2 p_t)(y)(f(x - y) - f(x)) \, dy \right|^p dx$$

$$\le \||\cdot|^\beta \nabla^2 p_t(\cdot)\|_{L_q(\mathbb{R}^d)}^p \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|f(x - y) - f(x)|^p}{|y|^{p\beta}} \, dx \, dy.$$

Choosing $\beta = d/p + \alpha$ gives

$$\|\nabla^2 \mathscr{P}_t f\|_{L_p(\mathbb{R}^d)}^p \le \||\cdot|^\beta \nabla^2 p_t(\cdot)\|_{L_q(\mathbb{R}^d)}^p [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)}^p,$$

and the claim follows by (13). The case $p = \infty$ follows similarly. □

LEMMA 2.3. *Let $\beta \in \mathbb{R}$, $\alpha \ge (-\beta) \vee 0$. There exists a constant N depending only on α, β and d such that for all $t \in (0, 1]$ and for all $f \in C^\beta(\mathbb{R}^d)$,*

$$\|\mathscr{P}_t f\|_{C^{\beta+\alpha}(\mathbb{R}^d)} \le N t^{-\alpha/2} \|f\|_{C^\beta(\mathbb{R}^d)}.$$

PROOF. For $\beta \ge 0$ these are well-known estimates that follow from direct computations with Gaussian densities. If $\beta < 0$, then the case $\alpha + \beta \in [0, 1]$ is shown in [7], Proposition 3.7. The case $\alpha + \beta > 1$ now follows by using the semigroup property. If $\alpha + \beta \in (1, 2]$, for example, then

$$\|\mathscr{P}_t f\|_{C^{\beta+\alpha}(\mathbb{R}^d)} \le \|\nabla \mathscr{P}_{\frac{t}{2}+\frac{t}{2}} f\|_{C^{\beta+\alpha-1}(\mathbb{R}^d)} + \|\mathscr{P}_t f\|_{B(\mathbb{R}^d)}$$

$$\lesssim t^{-(\beta+\alpha-1)/2} \|\mathscr{P}_{\frac{t}{2}} f\|_{C^1(\mathbb{R}^d)} + t^{\beta/2} \|f\|_{C^\beta(\mathbb{R}^d)}$$

$$\lesssim t^{-(\beta+\alpha-1)/2} t^{(\beta-1)/2} \|f\|_{C^\beta(\mathbb{R}^d)} + t^{\beta/2} \|f\|_{C^\beta(\mathbb{R}^d)}$$

$$\lesssim t^{-\alpha/2} \|f\|_{C^\beta(\mathbb{R}^d)}.$$

The case $\alpha + \beta \in (k, k + 1]$, for $k > 1$, follows by induction. □

Let Δ denote the Laplacian on $\mathbb{R}^d$. The next lemma is folklore, but since we did not find an exact reference in this form, we provide a short proof.

LEMMA 2.4. *Let $\beta \in \mathbb{R} \setminus \mathbb{Z}$. There exists a constant N depending on β such that for all $f \in C^\beta(\mathbb{R}^d)$*

$$\|(1 - \Delta)^{-1} f\|_{C^{\beta+2}(\mathbb{R}^d)} \le N \|f\|_{C^\beta(\mathbb{R}^d)}.$$

PROOF. If $\beta \in \mathbb{R} \setminus \mathbb{Z}$ and $\beta > 0$, these are the usual Schauder estimates. Hence, we only deal with the case $\beta < 0$. First, let us look at the case $-1 < \beta < 0$. Recall the following simple properties of $\mathscr{P}_t$: it commutes with differentiation $\nabla \mathscr{P}_t f = \mathscr{P}_t \nabla f$, it satisfies the semigroup property $\mathscr{P}_t \mathscr{P}_s = \mathscr{P}_{t+s}$, and one has

$$(1 - \Delta)^{-1} f = \int_0^\infty e^{-t} \mathscr{P}_t f \, dt.$$

Combining with Lemma 2.3 gives

$$\|(1-\Delta)^{-1}f\|_{C^1(\mathbb{R}^d)} \lesssim \int_0^1 t^{(\beta-1)/2}\|f\|_{C^\beta(\mathbb{R}^d)}\,dt + \sup_{t\geq 1}\|\mathscr{P}_t f\|_{C_1(\mathbb{R}^d)}.$$

For the second term on the right-hand side we have the following. Suppose that $\gamma \geq 0$. Then, for all $t \geq 1$, we have

$$\|\mathscr{P}_t f\|_{C^\gamma(\mathbb{R}^d)} \leq \|\mathscr{P}_1 f\|_{C^\gamma(\mathbb{R}^d)} \lesssim \|\mathscr{P}_{1/2} f\|_{L_\infty(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)},$$

where for the second inequality we used the semigroup property and Lemma 2.3 and for the last one the definition of C^β. Consequently, $\|(1-\Delta)^{-1}f\|_{C^1(\mathbb{R}^d)} \leq N\|f\|_{C^\beta(\mathbb{R}^d)}$ and we only have to show (recall that $\beta + 1 \in (0, 1)$) that

$$\left[\nabla(1-\Delta)^{-1}f\right]_{C^{\beta+1}(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}.$$

It is known that (see, e.g., [21], Exercise 3.4.4, p. 39)

$$\left[\nabla(1-\Delta)^{-1}f\right]_{C^{\beta+1}(\mathbb{R}^d)} \lesssim \sup_{\varepsilon>0} \varepsilon^{-\beta/2}\|\nabla\mathscr{P}_\varepsilon\nabla(1-\Delta)^{-1}f\|_{L_\infty(\mathbb{R}^d)}.$$

Let $\varepsilon \in (0, 1]$. By the above-mentioned properties of $\mathscr{P}$ and by Lemma 2.3 we have

$$
\begin{aligned}
&\|\nabla\mathscr{P}_\varepsilon\nabla(1-\Delta)^{-1}f\|_{L_\infty(\mathbb{R}^d)}\\
&\quad\leq \int_0^1 \|\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}f\|_{L_\infty(\mathbb{R}^d)}\,dt + \int_1^\infty e^{-t}\|\nabla^2\mathscr{P}_{t+\varepsilon}f\|_{L_\infty(\mathbb{R}^d)}\,dt\\
&\quad\lesssim \int_0^1 (t+\varepsilon)^{-1/2}\|\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}f\|_{L_\infty(\mathbb{R}^d)}\,dt + \|f\|_{C^\beta(\mathbb{R}^d)}\\
&\quad\leq \int_0^\infty (t+\varepsilon)^{-1/2}\|\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}f\|_{L_\infty(\mathbb{R}^d)}\,dt + \|f\|_{C^\beta(\mathbb{R}^d)}\\
&\quad\lesssim \int_0^\infty (t+\varepsilon)^{-1/2}(t+\varepsilon)^{(\beta-1)/2}\|f\|_{C^\beta(\mathbb{R}^d)}\,dt + \|f\|_{C^\beta(\mathbb{R}^d)}\\
&\quad\lesssim \varepsilon^{\beta/2}\|f\|_{C^\beta(\mathbb{R}^d)}.
\end{aligned}
$$

We now show that the same estimate holds for $\varepsilon > 1$. From (11) we have $\|\nabla\mathscr{P}_t g\|_{L_\infty} \leq \|\nabla p_t\|_{L_1}\|g\|_{L_\infty} \lesssim t^{-1/2}\|g\|_{L_\infty}$ for all $t > 0$. Consequently, for $\varepsilon > 1$, we have

$$
\begin{aligned}
\|\nabla\mathscr{P}_\varepsilon\nabla(1-\Delta)^{-1}f\|_{L_\infty(\mathbb{R}^d)} &\leq \int_0^\infty e^{-t}\|\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}f\|_{L_\infty(\mathbb{R}^d)}\,dt\\
&\lesssim \int_0^\infty e^{-t}(t+\varepsilon)^{-1/2}\|\nabla\mathscr{P}_{\frac{t+\varepsilon}{2}}f\|_{L_\infty(\mathbb{R}^d)}\,dt\\
&\lesssim \varepsilon^{-1/2}\|\mathscr{P}_{\frac{1}{2}}f\|_{C^1(\mathbb{R}^d)}\\
&\lesssim \varepsilon^{\beta/2}\|f\|_{C^\beta(\mathbb{R}^d)}.
\end{aligned}
$$

(14)

Consequently,

$$\left[\partial_i(1-\Delta)^{-1}f\right]_{C^{\beta+1}(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}.$$

This combined with $\|(1-\Delta)^{-1}f\|_{C^1(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}$ shows that

$$\|(1-\Delta)^{-1}f\|_{C^{2+\beta}(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}.$$

The case $-2 < \beta < -1$ is treated similarly as above. One can easily see that $\|(1 - \Delta)^{-1} f\|_{L_\infty(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}$ and what is left to be checked is that $[(1 - \Delta)^{-1} f]_{C^{\beta+2}(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}$. This is very similar to the previous argument, with the difference that one uses

$$\|(1 - \Delta)^{-1} f\|_{C^{2+\beta}(\mathbb{R}^d)} \lesssim \sup_{\varepsilon > 0} \varepsilon^{-\beta/2} \|\nabla^2 \mathscr{P}_\varepsilon (1 - \Delta)^{-1} f\|_{L_\infty(\mathbb{R}^d)},$$

(see, e.g., [21], Exercise 3.3.6, p. 40) and that in order to get a bound similar to (14), one needs the bound $\|\nabla^2 \mathscr{P}_t g\|_{L_\infty} \leq N t^{-1} \|g\|_{L_\infty}$ for all $t > 0$.

Finally, for the case $\beta < -2$, we have for $\varepsilon \in (0, 1]$

$$\|\mathscr{P}_\varepsilon (1 - \Delta)^{-1} f\|_{L_\infty(\mathbb{R}^d)}$$

$$\leq \int_0^{1-\varepsilon} e^{-t} \|\mathscr{P}_{t+\varepsilon} f\|_{L_\infty(\mathbb{R}^d)} + \int_{1-\varepsilon}^\infty e^{-t} \|\mathscr{P}_{t+\varepsilon} f\|_{L_\infty(\mathbb{R}^d)}$$

$$\lesssim \int_0^{1-\varepsilon} (t + \varepsilon)^{\beta/2} \sup_{t \in (0, 1-\varepsilon)} \left((t + \varepsilon)^{-\beta/2} \|\mathscr{P}_{t+\varepsilon} f\|_{L_\infty(\mathbb{R}^d)}\right) dt + \|f\|_{C^\beta(\mathbb{R}^d)}$$

$$\lesssim \varepsilon^{1+\beta/2} \|f\|_{C^\beta(\mathbb{R}^d)},$$

which of course implies that

$$\|(1 - \Delta)^{-1} f\|_{C^{\beta+2}(\mathbb{R}^d)} \lesssim \|f\|_{C^\beta(\mathbb{R}^d)}.$$

This finishes the proof. $\square$

LEMMA 2.5. *Let $t > s > 0$, $\alpha \in (0, 1)$, $m \in [1, \infty)$, and $p \in [1, m]$. Then there exist constants N and $N' = N'(\alpha)$ such that the following bounds hold for all $f \in L_m(\mathbb{R}^d)$ and $g \in \dot{W}_m^\alpha(\mathbb{R}^d)$:*

$$(15) \qquad \|f(B_t)\|_{L_p(\Omega)} \leq N t^{-d/(2m)} \|f\|_{L_m(\mathbb{R}^d)},$$

$$(16) \qquad \|g(B_t) - g(B_s)\|_{L_p(\Omega)} \leq N' s^{-d/(2m)} |t - s|^{\alpha/2} [g]_{\dot{W}_m^\alpha(\mathbb{R}^d)}.$$

PROOF. Since $\|\cdot\|_{L_p(\Omega)} \leq \|\cdot\|_{L_m(\Omega)}$, it suffices to show the result when $p = m$. We start with (15). It is evident that

$$\mathbf{E}|f(B_t)|^m = \int_{\mathbb{R}^d} |f(x)|^m p_t(x) \, dx = \||f|^m p_t\|_{L_1(\mathbb{R}^d)} \leq \|f\|_{L_m(\mathbb{R}^d)}^m \|p_t\|_{L_\infty(\mathbb{R}^d)}.$$

By (10), $\|p_t\|_{L_\infty(\mathbb{R}^d)} \lesssim t^{-d/2}$, yielding the claim. Moving on to (16), first one has

$$(17) \qquad \|g(B_t) - g(B_s)\|_{L_p(\Omega))}^p = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |g(x) - g(y)|^p p_{t-s}(x - y) p_s(y) \, dx \, dy.$$

If $p = m$, the right-hand side can be bounded by

$$[g]_{\dot{W}_m^\alpha(\mathbb{R}^d)}^m \|(\cdot)^{\alpha m + d} p_{t-s}(\cdot)\|_{L_\infty(\mathbb{R}^d)} \|p_s\|_{L_\infty(\mathbb{R}^d)} \lesssim [g]_{\dot{W}_m^\alpha(\mathbb{R}^d)}^m |t - s|^{\alpha m/2} s^{-d/2},$$

using (10) and (12). Thus, (16) follows. $\square$

PROPOSITION 2.6. *Let $p \in [1, \infty)$, $\delta \in (0, 1)$, and $\alpha \in (0, 2\delta]$. Then, one has for all $f \in \dot{W}_p^\alpha(\mathbb{R}^d)$*

$$(18) \qquad \|\mathscr{P}_t f - \mathscr{P}_s f\|_{L_p(\mathbb{R}^d)} \leq N |t - s|^\delta s^{\alpha/2 - \delta} [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)},$$

for all $0 < s < t$, where N is a constant depending only on d, p, α, δ. Moreover, for $\alpha = 2\delta$ the estimate also holds with $s = 0$ with the convention $\mathscr{P}_0 f = f$.

PROOF. One has

$$\|\mathscr{P}_t f - \mathscr{P}_s f\|_{L_p(\mathbb{R}^d)} \le \int_s^t \left\|\frac{\partial}{\partial r}\mathscr{P}_r f\right\|_{L_p(\mathbb{R}^d)} dr = \int_s^t \|\Delta\mathscr{P}_r f\|_{L_p(\mathbb{R}^d)} dr$$

$$\lesssim [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)} \int_s^t r^{\frac{\alpha}{2}-\delta} r^{-1+\delta} dr$$

$$\lesssim [f]_{\dot{W}_p^\alpha(\mathbb{R}^d)} s^{\frac{\alpha}{2}-\delta}(t-s)^\delta,$$

where we have used Lemma 2.2. If $\alpha = 2\delta$, then letting $s \to 0$ and applying Fatou's lemma yields the inequality for $s = 0$. □

PROPOSITION 2.7. *Let $d \in \mathbb{N}$, $K > 0$ and let Σ, $\bar{\Sigma}$ be symmetric invertible matrices such that $K^{-1}I \prec \Sigma\bar{\Sigma}^{-1} \prec KI$. Then for all $x, y \in \mathbb{R}^d$, and $\alpha \in [0,1]$ one has the bounds*

$$(19) \qquad |p_\Sigma(x) - p_\Sigma(y)| \le N|x-y|^\alpha \|\Sigma^{-1}\|^{\alpha/2}(p_{\Sigma/2}(x) + p_{\Sigma/2}(y)),$$

$$(20) \qquad |p_\Sigma(x) - p_{\bar{\Sigma}}(x)| \le N\|I - \Sigma\bar{\Sigma}^{-1}\|(p_{\Sigma/2}(x) + p_{\bar{\Sigma}/2}(x)),$$

where N is a constant depending only on d, K.

PROOF. We start with (19). The case $\alpha = 0$ is trivial, and therefore it suffices to show the bound in the case $\alpha = 1$, the remaining cases follow by interpolation. For all $k = 1, \ldots, d$, one has

$$\partial_{x_k} p_\Sigma(x) = \partial_{x_k}\left(-\frac{1}{2}x^*\Sigma^{-1}x\right)p_\Sigma(x).$$

It is easy to see that $|\partial_{x_k}|x_\Sigma|^2| \le 2|x_\Sigma|\|\Sigma^{-1/2}\|$, and therefore by (9) one gets

$$|\nabla p_\Sigma(x)| \lesssim \|\Sigma^{-1}\|^{1/2}|p_{\Sigma/2}(x)|.$$

Now take $x, y \in \mathbb{R}^d$, and assume without loss of generality that $|x_\Sigma| \le |y_\Sigma|$. Define $\bar{x}$ to be the minimizer of the distance from y to the set $\{z : |z_\Sigma| = |x_\Sigma|\}$. By definition, $|\bar{x} - y| \le |x - y|$ and every point z on the line segment between $\bar{x}$ and y satisfies $|z_\Sigma| \ge |x_\Sigma|$. Moreover, $|\bar{x}_\Sigma| = |x_\Sigma|$ and therefore $p_\Sigma(x) = p_\Sigma(\bar{x})$. Denoting by $e = \frac{y-\bar{x}}{|y-\bar{x}|}$ the unit vector in the direction of $y - \bar{x}$, one has

$$|p_\Sigma(x) - p_\Sigma(y)| = |p_\Sigma(\bar{x}) - p_\Sigma(y)| \le |y - \bar{x}| \int_0^1 |\partial_e p_\Sigma(\bar{x} + \theta(y-\bar{x}))| d\theta$$

$$\lesssim |y - \bar{x}| \int_0^1 \|\Sigma^{-1}\|^{1/2} p_{\Sigma/2}(\bar{x} + \theta(y-\bar{x})) d\theta$$

$$\lesssim |y - x|\|\Sigma^{-1}\|^{1/2} p_{\Sigma/2}(x),$$

as claimed. Moving on to (20),

$$p_\Sigma(x) - p_{\bar{\Sigma}}(x) = (1 - (\det\Sigma\bar{\Sigma}^{-1})^{1/2})p_\Sigma(x)$$

$$+ \frac{(\det\bar{\Sigma}^{-1})^{1/2}}{(2\pi)^{d/2}}\left(\exp\left(-\frac{1}{2}|x_\Sigma|^2\right) - \exp\left(-\frac{1}{2}|x_{\bar{\Sigma}}|^2\right)\right).$$

Thanks to (7) applied with $A = \Sigma\bar{\Sigma}^{-1} - I$, the first term can be immediately seen to be bounded by the right-hand side of (20). Concerning the second, one can write

$$(\det\bar{\Sigma}^{-1})^{1/2}\left|\exp\left(-\frac{1}{2}|x_\Sigma|^2\right) - \exp\left(-\frac{1}{2}|x_{\bar{\Sigma}}|^2\right)\right|$$

$$\lesssim (\det \bar{\Sigma}^{-1})^{1/2}(|x_{\Sigma}|^2 - |x_{\bar{\Sigma}}|^2)\left|\exp\left(-\frac{1}{2}|x|^2_{\Sigma}\right) + \exp\left(-\frac{1}{2}|x_{\bar{\Sigma}}|^2\right)\right|$$

$$\lesssim (|x_{\Sigma}|^2 - |x_{\bar{\Sigma}}|^2)(p_{\Sigma}(x) + p_{\bar{\Sigma}}(x)).$$

Note that

$$|x_{\Sigma}|^2 - |x_{\bar{\Sigma}}|^2 = x_{\Sigma}^*(I - (\Sigma^{1/2})^*\bar{\Sigma}^{-1}\Sigma^{1/2})x_{\Sigma}.$$

The matrix $A := I - (\Sigma^{1/2})^*\bar{\Sigma}^{-1}\Sigma^{1/2}$ is real symmetric and hence has the form $A = Q^*DQ$ where Q is an orthogonal matrix and D is a diagonal matrix. Hence, we have

$$x_{\Sigma}^*Ax_{\Sigma} = (Qx_{\Sigma})^*D(Qx_{\Sigma}) \leq |Qx_{\Sigma}|^2 \operatorname{Tr}(D) = |x_{\Sigma}|^2 \operatorname{Tr}(A).$$

Furthermore, writing $A = I - \Sigma^{-1/2}\Sigma\bar{\Sigma}^{-1}\Sigma^{1/2}$, we obtain that

$$\operatorname{Tr}(A) = \operatorname{Tr}(I - \Sigma^{-1/2}\Sigma\bar{\Sigma}^{-1}\Sigma^{1/2}) = \operatorname{Tr}(I - \Sigma\bar{\Sigma}^{-1}) \lesssim \|I - \Sigma\bar{\Sigma}^{-1}\|.$$

Hence, we have

$$(\det \bar{\Sigma}^{-1})^{1/2}\left|\exp\left(-\frac{1}{2}|x_{\Sigma}|^2\right) - \exp\left(-\frac{1}{2}|x_{\bar{\Sigma}}|^2\right)\right| \lesssim \|I - \Sigma\bar{\Sigma}^{-1}\||x_{\Sigma}|^2(p_{\Sigma}(x) + p_{\bar{\Sigma}}(x)).$$

By our assumptions on $\Sigma\bar{\Sigma}^{-1}$, we also have $|x_{\Sigma}| \lesssim |x_{\bar{\Sigma}}|$. Using (9) once again, we see that

$$|x_{\Sigma}|^2(p_{\Sigma}(x) + p_{\bar{\Sigma}}(x)) \lesssim |x_{\Sigma}|^2 p_{\Sigma}(x) + |x_{\bar{\Sigma}}|^2 p_{\bar{\Sigma}}(x) \lesssim p_{\Sigma/2}(x) + p_{\bar{\Sigma}/2}(x).$$

This completes the proof. $\square$

COROLLARY 2.8. *Let Σ_0 be a symmetric and positive definite $d \times d$-matrix and $x \in \mathbb{R}^d$. Then the function $\Sigma \mapsto p_{\Sigma}(x)$ is differentiable at Σ_0 and there exists a constant N depending only on d such that*

(21) $$\|\partial_{\Sigma} p_{\Sigma_0}(x)\| \leq N\|\Sigma_0^{-1}\|p_{\Sigma_0/2}(x).$$

PROOF. The differentiability is obvious, therefore we only need to show the bound (21). Take Σ' to be an arbitrary but fixed matrix with unit norm and $h > 0$. Set $\Sigma = \Sigma_0 + h\Sigma'$ and $\bar{\Sigma} = \Sigma_0$. Notice that for small enough h, these two matrices satisfy the condition of Proposition 2.7 with, say, $K = 2$. Applying (20), we get

$$|p_{\Sigma_0}(x) - p_{\Sigma_0 + h\Sigma'}(x)| \lesssim \|h\Sigma'\Sigma_0^{-1}\||p_{\Sigma_0/2}(x) + p_{(\Sigma_0 + h\Sigma')/2}(x)|.$$

Upon dividing by h and letting $h \to 0$, we get that $\langle \partial_{\Sigma} p_{\Sigma_0}(x), \Sigma' \rangle$, the directional derivative of $p_{\cdot}(x)$ at Σ_0 in the direction of Σ', satisfies

$$|\langle \partial_{\Sigma} p_{\Sigma_0}(x), \Sigma' \rangle| \lesssim \|\Sigma_0^{-1}\|p_{\Sigma_0/2}(x).$$

Taking suprema over Σ' gives (21). $\square$

2.3. *Sewing.* We need a weighted version of the stochastic sewing lemma. Such a version appeared recently in [3], but our formulation below allows for a larger range of exponents δ_1, δ_2. Set $[S, T]_{\leq} = \{(s, t) : S < s < t \leq T\}$ and

$$[S, T]_{\leq}^* = \{(s, t) : S < s < t \leq T, |t - s| \leq S\}.$$

For functions $\mathcal{A}$ of one variable we use the notation $\mathcal{A}_{s,t} = \mathcal{A}_t - \mathcal{A}_s$ and for functions A of two variables we use the notation $\delta A_{s,u,t} = A_{s,t} - A_{s,u} - A_{u,t}$. On $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$, the conditional expectation with respect to $\mathcal{F}_s$ is denoted by $\mathbf{E}^s$.

LEMMA 2.9. *Fix $p \geq 2$ and let A be a function from $[0, 1]_{\leq}$ to $L_p(\Omega)$ such that $A_{s,t}$ is $\mathscr{F}_t$-measurable for all $(s, t) \in [0, 1]_{\leq}$. Suppose that there exist $\varepsilon_1, \varepsilon_2 > 0, \delta_1, \delta_2 \geq 0$ and $C_1, C_2 < \infty$ satisfying $1/2 + \varepsilon_1 - \delta_1 > 0, 1 + \varepsilon_2 - \delta_2 > 0$ and the following: for all $S \in (0, 1]$, $(s, t) \in [S, 1]_{\leq}^*$ and $u \in [s, t]$, the bounds*

$$
(22) \qquad \|A_{s,t}\|_{L_p(\Omega)} \leq C_1 S^{-\delta_1} |t - s|^{1/2 + \varepsilon_1},
$$

$$
(23) \qquad \|\mathbf{E}^s \delta A_{s,u,t}\|_{L_p(\Omega)} \leq C_2 S^{-\delta_2} |t - s|^{1 + \varepsilon_2}
$$

hold. Then there exists a unique $(\mathscr{F}_t)_{t \in [0,1]}$-adapted function $\mathscr{A} : [0, 1] \to L_p(\Omega)$ such that $\mathscr{A}_0 = 0$ and for all $S \in (0, 1]$, there exist constants K_1, K_2 such that for all $(s, t) \in [S, 1]_{\leq}^*$ one has

$$
(24) \qquad \|\mathscr{A}_{s,t} - A_{s,t}\|_{L_p(\Omega)} \leq K_1 |t - s|^{1/2 + \varepsilon_1} + K_2 |t - s|^{1 + \varepsilon_2},
$$

$$
(25) \qquad \|\mathbf{E}^s (\mathscr{A}_{s,t} - A_{s,t})\|_{L_p(\Omega)} \leq K_2 |t - s|^{1 + \varepsilon_2}.
$$

Furthermore, the above bounds hold with $K_1 = NC_1 S^{-\delta_1}, K_2 = NC_2 S^{-\delta_2}$, where N depends only on $p, \varepsilon_1, \varepsilon_2, \delta_1, \delta_2$. Moreover, $\mathscr{A}$ satisfies the bounds, for all $S \in (0, 1], (s, t) \in [S, 1]_{\leq}^*$,

$$
(26) \qquad \|\mathscr{A}_{s,t}\|_{L_p(\Omega)} \leq N(C_1 S^{-\delta_1} |t - s|^{1/2 + \varepsilon_1} + C_2 S^{-\delta_2} |t - s|^{1 + \varepsilon_2}).
$$

PROOF. Note that our assumptions imply that the conditions of the usual stochastic sewing lemma from [26] are satisfied on each interval $I_n := [2^{-n}, 2^{-n+1}]$. Therefore we get adapted processes $\mathscr{A}^{(n)}$ on I_n such that $\mathscr{A}^{(n)}_{2^{-n}} = 0$ and that for all $s, t \in I_n$ one has the bounds (24)–(25)–(26) hold with $\mathscr{A}^{(n)}$ in place of $\mathscr{A}$ and 2^{-n} in place of S, with K_1, K_2 as specified in the statement. In particular, one gets

$$
\|\mathscr{A}^{(n)}_{2^{-n}, 2^{-n+1}}\|_{L_p(\Omega)} \lesssim (C_1 + C_2) 2^{-n\kappa}
$$

with $\kappa := \min(1/2 + \varepsilon_1 - \delta_1, 1 + \varepsilon_2 - \delta_2)$. By assumption $\kappa > 0$, and therefore if we define

$$
\mathscr{A}_t = \sum_{n=1}^{\infty} \mathbf{1}_{t \geq 2^{-n}} \mathscr{A}^{(n)}_{2^{-n}, 2^{-n+1} \wedge t},
$$

then the sum converges in $L_p(\Omega)$ so the definition is indeed meaningful. Clearly $\mathscr{A}$ is adapted, $\mathscr{A}_0 = 0$, and since $\mathscr{A}_{s,t} = \mathscr{A}^{(n)}_{s,t}$ for $(s, t) \in I_n$, the bounds (24)–(26) hold for $(s, t) \in I_n$, $S = 2^{-n}$. Extending it to general pairs (s, t) is then standard; see, for example, [6]. $\square$

2.4. *PDE estimates.* In this section, we obtain estimates concerning the PDE associated with (1). Although such results seem folklore, due to the lack of exact reference to our knowledge, we provide a short proof. In the following, for a function u on $\mathbb{R}^d$, $\nabla^2 u$ is understood to be a matrix. For matrices A, B, just like for any other vectors, the $\cdot$ notation denotes the scalar product $A \cdot B = \sum_{i,j} A_{i,j} B_{i,j}$. Consider the PDE

$$
(27) \qquad \begin{aligned} \partial_t u &= \frac{1}{2}(\sigma\sigma^*) \cdot \nabla^2 u + b \cdot \nabla u - \theta u + f \quad \text{in } (0, 1) \times \mathbb{R}^d, \\ u &= 0 \quad \text{on } \{0\} \times \mathbb{R}^d, \end{aligned}
$$

for $\theta > 0$. We look for solutions in the mixed order Sobolev spaces $W_p^{1,2}((0, 1) \times \mathbb{R}^d)$ defined as the completion of the set of compactly supported smooth functions with respect to the norm

$$
\|u\|_{W_p^{1,2}((0,1) \times \mathbb{R}^d)} = \sum_{k \in \mathbb{N}_0, \ell \in \mathbb{N}_0^d, 2k + |\ell|_1 \leq 2} \|\partial_t^k \partial_x^\ell u\|_{L_p((0,1) \times \mathbb{R}^d)}.
$$

LEMMA 2.10.   *Let $f \in L_p((0,1) \times \mathbb{R}^d) \cap L_\infty((0,1) \times \mathbb{R}^d)$ for $p > 1$, and $b \in L_\infty(\mathbb{R}^d)$. Then, (27) has a unique solution u in the class $W_p^{1,2}((0,1) \times \mathbb{R}^d)$. Moreover, there exists a constant N depending only on $p, \lambda, \|\sigma\|_{C^2}, T$, and $\|b\|_{\mathbb{B}(\mathbb{R}^d)}$, such that*

$$\|u\|_{W_p^{1,2}((0,1) \times \mathbb{R}^d)} \leq N \|f\|_{L_p((0,1) \times \mathbb{R}^d)}. \tag{28}$$

In addition, there exists $\theta_0 > 0$ such that for all $\theta > \theta_0$

$$\|\nabla u\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \leq N\theta^{-1/2} \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)}. \tag{29}$$

PROOF.   The existence, the uniqueness, and estimate (28) follow from [22], Theorem 10, page 123. For (29), let us assume first that $b = 0$. We have that $\tilde{u}(t, x) := e^{\theta t} u(t, x)$ satisfies

$$\partial_t \tilde{u} = \frac{1}{2}(\sigma \sigma^*) \cdot \nabla^2 \tilde{u} + e^{\theta t} f. \tag{30}$$

Hence, we have

$$\nabla \tilde{u}(t, x) = \int_0^t \int_{\mathbb{R}^d} \nabla_x p_{t-s}^\sigma(x, y) e^{\theta s} f(s, y) \, dy \, ds,$$

where $p^\sigma(x, y)$ is the Green's function of the operator $\partial_t - \frac{1}{2}(\sigma \sigma^*) \cdot \nabla^2$. It is well known (see, e.g., [13], (6.13), p. 24), that there is a constant N depending on d and $\|\sigma\|_{C^2}$ such that $\sup_{x \in \mathbb{R}^d} \|\nabla_x p_t^\sigma(x, \cdot)\|_{L_1(\mathbb{R}^d)} \leq Nt^{-1/2}$ for all $t \in (0, 1]$, which implies

$$|\nabla u(t, x)| \lesssim \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \int_0^t |t - s|^{-1/2} e^{-\theta(t-s)} \, ds \lesssim \theta^{-1/2} \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)},$$

where the last inequality can be easily seen from a change of variables. This shows (29) in the case $b = 0$. For the general case, notice that by assumption we have that $f \in L_q((0,1) \times \mathbb{R}^d)$ for all $q \in [p, \infty)$, which shows that in fact (28) holds for all such q in place of p. From this and the embedding $W_q^{1,2}((0,1) \times \mathbb{R}^d) \subset C^{1-(d+2)/(2q), 2-(d+2)/q}((0,1) \times \mathbb{R}^d)$ for q large enough, it follows that $b \cdot \nabla u + f \in L_p((0,1) \times \mathbb{R}^d) \cap L_\infty((0,1) \times \mathbb{R}^d)$. Hence, we can replace f in our previous analysis by $b \cdot \nabla u + f$, which gives

$$\|\nabla u\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \lesssim \theta^{-1/2} \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} + \theta^{-1/2} \|\nabla u\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} < \infty.$$

If θ is large enough, (29) clearly follows. $\square$

2.5. *A Gronwall-type lemma.*   We use a somewhat nonstandard form of Gronwall's lemma, stated in the following lemma.

LEMMA 2.11.   *Let $(Y)_{t \in [0,1]}$ be an adapted, nonnegative process, such that $\mathbf{E} \sup_{t \in [0,1]} Y_t^2 < \infty$. Let $(A_t)_{t \in [0,1]}$ be an adapted, continuous increasing process, and let $C, R \geq 0$ be constants. Assume that for all stopping times τ, τ', with $0 \leq \tau \leq \tau' \leq 1$ we have*

$$\mathbf{E} \sup_{t \in [\tau, \tau']} Y_t \leq C\mathbf{E}Y_\tau + CR + C\mathbf{E} \int_\tau^{\tau'} Y_s \, dA_s. \tag{31}$$

Then, there exist a constant $\bar{C} \geq 0$ depending only on C such that for all $m \in \mathbb{N}$ we have

$$\mathbf{E} \sup_{t \in [0,1]} Y_t \leq \bar{C}^m \mathbf{E}Y_0 + \bar{C}^m R + \bar{C}\left(\mathbf{E} \sup_{t \in [0,1]} Y_t^2\right)^{1/2}\left(\mathbf{P}\left(A_1 \geq \frac{m}{2C}\right)\right)^{1/2}. \tag{32}$$

PROOF.   Let us set $\tau_0 := 0$ and for $m \in \mathbb{N}$ let us define inductively

$$\tau_m := \inf\{t > \tau_{m-1} : A_t - A_{\tau_{m-1}} > (2C)^{-1}\} \wedge 1.$$

By (31) it follows that for all $m \in \mathbb{N}$

$$(33) \qquad \mathbf{E} \sup_{t \in [\tau_{m-1}, \tau_m]} Y_t \leq 2C\mathbf{E}Y_{\tau_{m-1}} + 2CR.$$

Consequently, we have

$$(34) \qquad \mathbf{E} \sup_{t \in [0, \tau_m]} Y_t \leq \sum_{j=1}^m \mathbf{E} \sup_{t \in [\tau_{j-1}, \tau_j]} Y_t \leq 2C \sum_{j=1}^m \mathbf{E}Y_{\tau_{j-1}} + m2CR.$$

Notice that (33), upon iteration, implies

$$\mathbf{E}Y_{\tau_m} \leq (2C)^m \mathbf{E}Y_0 + R \sum_{j=1}^m (2C)^j \leq \tilde{C}^m \mathbf{E}Y_0 + \tilde{C}^m R,$$

for some $\tilde{C}$ depending only on C. This combined with (34), gives

$$\mathbf{E} \sup_{t \in [0, \tau_m]} Y_t \leq 2Cm\tilde{C}^m \mathbf{E}Y_0 + 2Cm\tilde{C}^m R + m2CR \leq \bar{C}^m \mathbf{E}Y_0 + \bar{C}^m R,$$

for an appropriate $\bar{C}$ depending only on C. Finally, by Hölder's inequality we have

$$\mathbf{E} \sup_{t \in [0,1]} Y_t \leq \mathbf{E} \sup_{t \in [0, \tau_m]} Y_t + \left(\mathbf{E} \sup_{t \in [0,1]} Y_t^2\right)^{1/2} (\mathbf{P}(\tau_m < 1))^{1/2}$$

$$\leq \bar{C}^m \mathbf{E}Y_0 + \bar{C}^m R + \left(\mathbf{E} \sup_{t \in [0,1]} Y_t^2\right)^{1/2} (\mathbf{P}(\tau_m < 1))^{1/2},$$

which implies (32), since $\{\tau_m < 1\} \subset \{A_1 \geq m(2C)^{-1}\}$ by the definition of τ_m. This finishes the proof. $\square$

3. Integral estimates.   In this section, we obtain estimates with sharp rates for the quantity

$$\int_0^1 \left(f_r(X_r^n) - f_r(X_{\kappa_n(r)}^n)\right) dr$$

where f is a bounded measurable function on $[0, 1] \times \mathbb{R}^d$. The main results of the current section are Corollaries 3.2 and 3.5. Our method is based on stochastic sewing techniques, Lemma 2.9. The analytic properties which we utilize to verify the hypotheses of Lemma 2.9 are the heat kernel estimates from Section 2.2 and estimates on the density of the Euler–Maruyama approximations, see (36) below. We consider the cases of multiplicative noise and additive noise, corresponding to Assumptions 1.1 and 1.4, separately.

3.1. *Multiplicative noise.*   As in [7], we first consider the driftless Euler–Maruyama scheme

$$(35) \qquad d\bar{X}_t^n = \sigma\left(\bar{X}_{\kappa_n(t)}^n\right) dB_t, \quad \bar{X}_0^n = x_0.$$

We will sometimes denote the dependence on the initial condition by writing $\bar{X}_t^n(x_0)$, but most of the time this dependence will not play any role and therefore will be omitted from the notation.

Let us recall some key estimates for the transition probabilities of the $\bar{X}_t^n$. Notice that a trivial induction argument shows that $\bar{X}_t^n$ has a density for all $t > 0$. A useful bound for

the density in L_p spaces is due to Gyöngy and Krylov, see Lemma 5.1 below. Notice that in Lemma 5.1 one cannot put derivatives on the test function G. For this reason, another bound on the transition probabilities of $\bar{X}$ was derived in [7]: under Assumption 1.1, by [7], Theorem 5.1, for any $G \in C^1$, $t = 1/n, 2/n, \ldots, 1$, one has

$$(36) \qquad \left| \mathbf{E}(\nabla G)(\bar{X}_t^n) \right| \leq N \|G\|_\mathbb{B} t^{-1/2} + N \|G\|_{C^1} e^{-cn},$$

with some constant $N = N(d, \lambda, \|\sigma\|_{C^2})$ and $c = c(d, \|\sigma\|_{C^2}) > 0$.

Estimate (36) is applied to obtain the following result, extending [7], Lem. 6.1, from Hölder continuous functions to bounded measurable functions.

LEMMA 3.1. *Let $y \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $p > 0$. Suppose that Assumption 1.4 holds and that $\bar{X}^n := \bar{X}^n(y)$ is the solution of (35). Then for all $f \in \mathbb{B}([0,1] \times \mathbb{R}^d)$, $0 \leq s \leq t \leq 1$, $n \in \mathbb{N}$, one has the bound*

$$(37) \qquad \left\| \int_s^t \left(f_r(\bar{X}_r^n) - f_r(\bar{X}_{\kappa_n(r)}^n) \right) dr \right\|_{L_p(\Omega)} \leq N \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} n^{-1/2+2\varepsilon} |t-s|^{1/2+\varepsilon},$$

with some $N = N(p, d, \varepsilon, \lambda, \|\sigma\|_{C^2})$.

PROOF. Parts of the proof are unchanged from [7], Lem 6.1, therefore occasionally we shall refer back to arguments therein. By the linearity of the left-hand side in f, we may and will assume $\|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} = 1$. Set $0 \leq s \leq t \leq 1$,

$$A_{s,t} := \mathbf{E}^s \int_s^t \left(f_r(\bar{X}_r^n) - f_r(\bar{X}_{\kappa_n(r)}^n) \right) dr.$$

We will use Lemma 2.9 with $\delta_1 = \delta_2 = 0$ (which is the same as the original version of stochastic sewing lemma from [26]). It is straightforward to verify that $\mathbf{E}^s \delta A_{s,u,t} = 0$ for any $0 \leq s \leq u \leq t \leq 1$. Therefore (23) is verified with $C_2 = 0$.

The bulk of the proof is the verification of (22) with $C_1 = N n^{-1/2+2\varepsilon}$, $\varepsilon_1 = \varepsilon$. Define k by $k/n = \kappa_n(s)$. If $t \in [s, (k+4)/n]$, then $|t-s| \leq 4n^{-1}$ and it is evident that

$$(38) \qquad \|A_{s,t}\|_{L_p(\Omega)} \leq 2|t-s| \lesssim n^{-1/2+\varepsilon} |t-s|^{1/2+\varepsilon},$$

as required. For the case $t > (k+4)/n$, one first writes

$$|A_{s,t}| = |I_1 + I_2| := \left| \left(\int_s^{(k+4)/n} + \int_{(k+4)/n}^t \right) \mathbf{E}^s \left(f_r(\bar{X}_r^n) - f_r(\bar{X}_{\kappa_n(r)}^n) \right) dr \right|.$$

The term $|I_1|$ is easily seen to be bounded by the right-hand-side of (38) even for all ω, and so clearly $\|I_1\|_{L_p(\Omega)} \lesssim n^{-1/2+\varepsilon} |t-s|^{1/2+\varepsilon}$. Therefore, (23) is verified once we show

$$(39) \qquad \|I_2\|_{L_p(\Omega)} \lesssim |t-s|^{1/2+\varepsilon} n^{-1/2+2\varepsilon}.$$

For $r \in [(k+4)/n, t]$, we have $s \leq (k+1)/n \leq \kappa_n(r)$ so that we can write

$$I_2 = \mathbf{E}^s \int_{(k+4)/n}^t \mathbf{E}^{(k+1)/n} \left(\mathbf{E}^{\kappa_n(r)} f_r(\bar{X}_r^n) - f_r(\bar{X}_{\kappa_n(r)}^n) \right) dr.$$

Denote $C(x) = (\sigma \sigma^*)(x)$. We have

$$\mathbf{E}^{\kappa_n(r)} f_r(\bar{X}_r^n) = \mathbf{E}^{\kappa_n(r)} f_r \left(\bar{X}_{\kappa_n(r)}^n + (B_r - B_{\kappa_n(r)}) \sigma(\bar{X}_{\kappa_n(r)}^n) \right) = \mathscr{P}_{C(\bar{X}_{\kappa_n(r)}^n)(r-\kappa_n(r))} f_r(\bar{X}_{\kappa_n(r)}^n),$$

so with $g_r^n(x) := \mathscr{P}_{C(x)(r-\kappa_n(r))} f_r(x) - f_r(x)$ we have

$$(40) \qquad I_2 = \mathbf{E}^s \int_{(k+4)/n}^t \mathbf{E}^{(k+1)/n} g_r^n(\bar{X}_{\kappa_n(r)}^n) dr.$$

By [7], eq. (6.7), one has the bound

$$\|g_r^n\|_{C^\beta} \lesssim \|f\|_{\mathbb{B}} \, n^{\beta/2} = n^{\beta/2} \tag{41}$$

for all $\beta \in [-1, 0]$, $r \in [0, 1]$, and $n \in \mathbb{N}$. Define the operator $\tilde{\mathscr{P}}$ by $(\tilde{\mathscr{P}}h)(x) = \mathbf{E}h(\bar{X}_{1/n}^n(x))$ and denote $\tilde{g} = \tilde{\mathscr{P}}g$. By the tower rule and applying the Markov property twice one gets

$$
\begin{aligned}
\mathbf{E}^{(k+1)/n} g_r^n(\bar{X}_{\kappa_n(r)}^n) &= \mathbf{E}^{(k+1)/n} \mathbf{E}^{\kappa_n(r)-1/n} g_r^n(\bar{X}_{\kappa_n(r)}^n) \\
&= \mathbf{E}^{(k+1)/n} \tilde{g}_r^n(\bar{X}_{\kappa_n(r)-1/n}^n) \\
&= \left(\mathbf{E}\tilde{g}_r^n(\bar{X}_{\kappa_n(r)-(k+2)/n}^n(y)) \right)\big|_{y=\bar{X}_{(k+1)/n}^n}.
\end{aligned}
\tag{42}
$$

Introduce the functions u and $\tilde{u}$ as the solutions of the equations

$$(1 - \Delta)u = g, \qquad (1 - \Delta)\tilde{u} = \tilde{g}.$$

It follows from Lemma 2.4 that these solutions exist, are unique, and satisfy the bounds

$$\|u\|_{C^{\beta+2}} \le N(\beta)\|g\|_{C^\beta}, \qquad \|\tilde{u}\|_{C^{\beta+2}} \le N(\beta)\|\tilde{g}\|_{C^\beta} \tag{43}$$

for all $\beta \in \mathbb{R} \setminus \mathbb{Z}$. Denote $\delta = \kappa_n(r) - (k+2)/n$. One then has, by (42), (36), and (43)

$$
\begin{aligned}
\left| \mathbf{E}^{(k+1)/n} g(\bar{X}_{\kappa_n(r)}^n) \right| &\le \sup_{y \in \mathbb{R}^d} \left| \mathbf{E}(\tilde{u} - \Delta\tilde{u})(\bar{X}_\delta^n(y)) \right| \\
&\lesssim \|\tilde{u}\|_{C^1} \delta^{-1/2} + \|\tilde{u}\|_{C^2} e^{-cn} \\
&\lesssim \|\tilde{g}\|_{C^{-1+\varepsilon}} \delta^{-1/2} + \|\tilde{g}\|_{C^\varepsilon} e^{-cn}.
\end{aligned}
\tag{44}
$$

First we find a bound on $\|\tilde{g}\|_{C^\varepsilon}$. Clearly one has $\|\tilde{g}\|_{C^0} \le \|g\|_{C^0}$. Furthermore,

$$
\begin{aligned}
\left| \tilde{g}(x) - \tilde{g}(y) \right| &= \left| \int_{\mathbb{R}^d} \left(p_{\frac{C(x)}{n}}(x - z) - p_{\frac{C(y)}{n}}(y - z) \right) g(z)\, dz \right| \\
&\le \|g\|_{\mathbb{B}} \int_{\mathbb{R}^d} \left| p_{\frac{C(x)}{n}}(x - z) - p_{\frac{C(x)}{n}}(y - z) \right| + \left| p_{\frac{C(x)}{n}}(y - z) - p_{\frac{C(y)}{n}}(y - z) \right| dz.
\end{aligned}
$$

The first term in the integral is bounded via (19):

$$\left| p_{\frac{C(x)}{n}}(x - z) - p_{\frac{C(x)}{n}}(y - z) \right| \lesssim |x - y|^\varepsilon n^{\varepsilon/2} \left\| (C(x))^{-1} \right\|^{\varepsilon/2} \left| p_{\frac{C(x)}{2n}}(x - z) + p_{\frac{C(x)}{2n}}(y - z) \right|.$$

By Assumption 1.1, we have that $\|(C(x))^{-1}\| \le N$. Concerning the other term, we wish to apply (20). To this end, using again Assumption 1.1, we have

$$\left\| I - C(x)(C(y))^{-1} \right\| \le \left\| (C(y))^{-1} \right\| \left\| C(y) - C(x) \right\| \lesssim |x - y|.$$

By (20) this implies

$$\left| p_{\frac{C(x)}{n}}(y - z) - p_{\frac{C(y)}{n}}(y - z) \right| \lesssim |x - y| \left| p_{\frac{C(x)}{2n}}(y - z) + p_{\frac{C(y)}{2n}}(y - z) \right|.$$

At this point we simply use the fact that any heat kernel p_Σ has unit integral to conclude that for any x, y with $|x - y| \le 1$

$$\left| \tilde{g}(x) - \tilde{g}(y) \right| \lesssim \|g\|_{\mathbb{B}} |x - y|^\varepsilon n^{\varepsilon/2},$$

or,

$$\|\tilde{g}\|_{C^\varepsilon} \lesssim n^{\varepsilon/2} \|g\|_{\mathbb{B}}. \tag{45}$$

Next, we bound $\|\tilde{g}\|_{C^{-1+\varepsilon}}$. Recall that $\tilde{g} = \tilde{\mathscr{P}}((1-\Delta)u)$, and one can write the trivial bounds

$$(46) \qquad \|\tilde{\mathscr{P}}u\|_{C^{-1+\varepsilon}} \leq \|\tilde{\mathscr{P}}u\|_{C^0} \leq \|u\|_{C^0} \lesssim \|g\|_{C^{-1+\varepsilon}}.$$

Also

$$(47) \qquad \|\partial_k \tilde{\mathscr{P}} \partial_k u\|_{C^{-1+\varepsilon}} \lesssim \|\tilde{\mathscr{P}} \partial_k u\|_{C^\varepsilon} \lesssim n^{\varepsilon/2} \|\partial_k u\|_{C^\varepsilon} \lesssim n^{\varepsilon/2} \|g\|_{C^{-1+\varepsilon}},$$

where in the second inequality we used the previous argument. Putting $U = \partial_k u$, it remains to estimate the commutator $\partial_k \tilde{\mathscr{P}} U - \tilde{\mathscr{P}} \partial_k U$ in the norm $C^{-1+\varepsilon}$. It turns out that it can even be bounded in C^0. Indeed,

$$
\begin{aligned}
(\partial_k \tilde{\mathscr{P}} U - \tilde{\mathscr{P}} \partial_k U)(x) &= \int_{\mathbb{R}^d} \left(\partial_{y_k} p_{\frac{C(y)}{n}}(x-z)\right)\big|_{y=x} U(z)\, dz \\
(48) \qquad &= \int_{\mathbb{R}^d} \left(\partial_\Sigma p_\Sigma(x-z)\right)\big|_{\Sigma=C(x)/n} \frac{\partial_{x_k} C(x)}{n} U(z)\, dz \\
&\lesssim \int_{\mathbb{R}^d} p_{\frac{C(x)}{2n}}(x-z) U(z)\, dz \lesssim \|U\|_{\mathbb{B}} \lesssim \|u\|_{C^{1+\varepsilon}} \lesssim \|g\|_{C^{-1+\varepsilon}},
\end{aligned}
$$

where for $\partial_\Sigma p_\Sigma$ we used (21). From (46), (47), (48), we can conclude

$$(49) \qquad \|\tilde{g}\|_{C^{-1+\varepsilon}} \lesssim n^{\varepsilon/2} \|g\|_{C^{-1+\varepsilon}}.$$

Putting (45) and (49) into (44), and then using (41),

$$
\begin{aligned}
\left|\mathbf{E}^{(k+1)/n} g\left(\bar{X}^n_{\kappa_n(r)}\right)\right| &\lesssim n^{\varepsilon/2} \delta^{-1/2} \|g\|_{C^{-1+\varepsilon}} + n^{\varepsilon/2} e^{-cn} \|g\|_{\mathbb{B}} \\
&\lesssim n^{-1/2+\varepsilon} \delta^{-1/2} + N n^{\varepsilon/2} e^{-cn}.
\end{aligned}
$$

Recall that we defined $\delta = \kappa_n(r) - (k+2)/n$ and we are in the situation of (40). In particular, $\delta \geq 2/n$ and so we can further write

$$\left|\mathbf{E}^{(k+1)/n} g\left(\bar{X}^n_{\kappa_n(r)}\right)\right| \lesssim n^{-1/2+2\varepsilon} \delta^{-1/2+\varepsilon}.$$

Substituting this bound back into (40) and integrating, we get (39).

At this point, the conditions of Lemma 2.9 are satisfied. It only remains to identify the process $\mathscr{A}$. We claim that it is given by

$$\hat{\mathscr{A}}_t = \int_0^t \left(f_r(\bar{X}^n_r) - f_r(\bar{X}^n_{\kappa_n(r)})\right) dr.$$

Clearly $\hat{\mathscr{A}}$ is adapted and starts from 0. Moreover, $A_{s,t} = \mathbf{E}^s \hat{\mathscr{A}}_{s,t}$ and therefore (25) is satisfied with $K_2 = 0$. On the other hand, one has the trivial bound

$$\|\hat{\mathscr{A}}_{s,t} - A_{s,t}\|_{L_p(\Omega)} \leq \|\hat{\mathscr{A}}_{s,t}\|_{L_p(\Omega)} + \|A_{s,t}\|_{L_p(\Omega)} \leq 4|s-t|,$$

which verifies (24) with $K_1 = 4$. It therefore follows that $\hat{\mathscr{A}} = \mathscr{A}$, and the bound (26) is precisely our claimed bound (37). $\quad\square$

COROLLARY 3.2. *Let* $\varepsilon \in (0, 1/2)$, $p > 0$. *Suppose that Assumption* 1.1 *holds and that* X^n *is the solution of* (2). *Then for all* $f \in \mathbb{B}([0,1] \times \mathbb{R}^d)$, $n \in \mathbb{N}$, *one has the bound*

$$(50) \qquad \left\| \int_0^\cdot \left(f_r(X^n_r) - f_r(X^n_{\kappa_n(r)})\right) dr \right\|_{L_p(\Omega; \mathbb{B}[0,1])} \leq N \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} n^{-1/2+2\varepsilon},$$

with some $N = N(p, d, \varepsilon, \lambda, \|b\|_{\mathbb{B}}, \|\sigma\|_{C^2})$.

PROOF. Owning to Jensen's inequality, it suffices to prove the statement for $p \geq 2$. For any continuous process Z, let us denote

$$h(Z) = \left\| \int_0^{\cdot} \left(f_r(Z_r) - f_r(Z_{\kappa_n(r)}) \right) dr \right\|_{\mathbb{B}([0,1])}.$$

From Kolmogorov's continuity criterion and (37), one immediately gets

$$(51) \qquad \left\| h(\bar{X}^n) \right\|_{L_{2p}(\Omega)} \leq N \| f \|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} n^{-1/2 + 2\varepsilon}.$$

Let us set

$$\rho = \exp\left(-\int_0^1 (\sigma^{-1} b)(X_{\kappa_n(r)}^n) \, dB_r - \frac{1}{2} \int_0^1 |(\sigma^{-1} b)(X_{\kappa_n(r)}^n)|^2 \, dr \right)$$

and define the measure $\tilde{\mathbf{P}}$ by $d\tilde{\mathbf{P}} = \rho \, d\mathbf{P}$. Since $\sigma^{-1} b$ is a bounded measurable function, $\mathbf{E}\rho^{\theta}$ is finite for every $\theta \in \mathbb{R}$. By Girsanov's theorem, $\tilde{\mathbf{P}}$ is a probability measure and X^n solves (35) with a $\tilde{\mathbf{P}}$-Wiener process $\tilde{B}$ in place of B. In other words, the law of X^n under $\tilde{\mathbf{P}}$ and the law of $\bar{X}^n$ under $\mathbf{P}$ coincide. Therefore,

$$\mathbf{E}h(X^n)^p = \tilde{\mathbf{E}}(h(X^n)^p \rho^{-1}) \leq (\tilde{\mathbf{E}}h(X^n)^{2p})^{1/2} (\tilde{\mathbf{E}}\rho^{-2})^{1/2} = (\mathbf{E}h(\bar{X}^n)^{2p})^{1/2} (\mathbf{E}\rho^{-1})^{1/2}.$$

Note that $\mathbf{E}\rho^{-1}$ is bounded by a constant depending only on the supremum of $\sigma^{-1} b$, which in turn is bounded by $\lambda^{-1} \| b \|_{\mathbb{B}}$. Combining this with (51), we get the desired bound (50). $\qquad \square$

3.2. *Additive noise.*

LEMMA 3.3. *Let $\alpha \in (0, 1)$, $p \geq 2$, $\varepsilon \in (0, 1/2)$, $\alpha' \in (1 - 2\varepsilon, 1)$, and $m \geq d$ such that $m > p$. Let $f \in \mathbb{B}([0, 1], \dot{W}_m^{\alpha}(\mathbb{R}^d)) \cap \mathbb{B}([0, 1] \times \mathbb{R}^d)$, and $g \in \mathbb{B}([0, 1], C^{\alpha'}(\mathbb{R}^d))$. Then for all $S \in (0, 1]$, $(s, t) \in [S, 1]_{\leq}^*$, and $n \in \mathbb{N}$ one has the bound*

$$
\left\| \int_s^t g_r(B_r)(f_r(B_r) - f_r(B_{\kappa_n(r)})) \, dr \right\|_{L_p(\Omega)}
$$

$$(52) \qquad \leq N \| g \|_{\mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))} \left(\sup_{r \in [0,1]} [f_r]_{\dot{W}_m^{\alpha}(\mathbb{R}^d)} + \| f \|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \right)$$

$$\times n^{-(1+\alpha)/2 + \varepsilon} |t - s|^{1/2 + \varepsilon} S^{-d/(2m)},$$

where N is a constant depending only on d, p, α, m, and ε.

PROOF. By linearity of the left-hand side in both g and f, we may and will assume

$$\| g \|_{\mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))} = \sup_{r \in [0,1]} [f_r]_{\dot{W}_m^{\alpha}(\mathbb{R}^d)} + \| f \|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} = 1.$$

We define for $(s, t) \in [0, 1]_{\leq}$

$$A_{s,t} := \mathbf{E}^s \int_s^t g_r(B_s)(f_r(B_r) - f_r(B_{\kappa_n(r)})) \, dr.$$

Let us check the conditions of Lemma 2.9, with $\delta_1 = \delta_2 = d/(2m)$. We begin by showing that (22) holds with $C_1 = N n^{-(1+\alpha)/2 + \varepsilon}$, $\varepsilon_1 = \varepsilon$. Take $S \in (0, 1]$ and $(s, t) \in [S, 1]_{\leq}^*$. Define k by $k/n = \kappa_n(s)$. Suppose first that $t \in [(k + 4)/n, 1]$. By using the fact that $\| g_r \|_{C^{\alpha'}(\mathbb{R}^d)} \leq 1$ for all $r \in [0, 1]$ and that $p < m$, we have that

$$(53) \qquad \| A_{s,t} \|_{L_p(\Omega)} \leq \tilde{A}_{s,t} := \int_s^t \| \mathbf{E}^s (f_r(B_r) - f_r(B_{\kappa_n(r)})) \|_{L_m(\Omega)} \, dr.$$

Notice that

$$\tilde{A}_{s,t} = I_1 + I_2 := \left(\int_s^{(k+4)/n} + \int_{(k+4)/n}^t \right) \left\| \mathbf{E}^s \left(f_r(B_r) - f_r(B_{\kappa_n(r)}) \right) \right\|_{L_m(\Omega)} dr.$$

One has, by (15), (18), using $n^{-1} < t - s$ and $[f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)} \le 1$ for all $r \in [0, 1]$,

$$I_2 \le \int_{(k+4)/n}^t \left\| \mathscr{P}_{r-s} f_r(B_s) - \mathscr{P}_{\kappa_n(r)-s} f_r(B_s) \right\|_{L_m(\Omega)} dr$$

$$\lesssim \int_{(k+4)/n}^t s^{-d/(2m)} \left\| \mathscr{P}_{r-s} f_r - \mathscr{P}_{\kappa_n(r)-s} f_r \right\|_{L_m(\mathbb{R}^d)} dr$$

$$\lesssim \int_{(k+4)/n}^t n^{-(1+\alpha)/2} (\kappa_n(r) - s)^{-1/2} s^{-d/(2m)} dr$$

$$\lesssim n^{-(1+\alpha)/2} (t - s)^{1/2} s^{-d/(2m)}$$

$$\lesssim n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2+\varepsilon} S^{-d/(2m)}.$$

Next, we deal with the term I_1. If $\alpha \le d/m$, we use $\|f_r\|_{\mathbb{B}} \le 1$, $r \in [0, 1]$, in a trivial way to get

$$I_1 \le 8n^{-1} \lesssim n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2+\varepsilon} S^{-\alpha/2}$$

$$\le n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2+\varepsilon} S^{-d/(2m)},$$

where we also used $n^{-1} \le |t - s| \le S$. If $\alpha > d/m$, then we use that $[f_r]_{C^{\alpha-d/m}(\mathbb{R}^d)} \lesssim [f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)}$ (see Remark 1.3) and we see that

$$I_1 \lesssim n^{-1-\alpha/2+d/(2m)} \| f \|_{\mathbb{B}([0,1], C^{\alpha-d/m})}$$

$$\le n^{-1-\alpha/2+d/(2m)}$$

$$\le n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2-d/(2m)+\varepsilon}$$

$$\le n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2+\varepsilon} S^{-d/(2m)},$$

where we have used that $n^{-1} \le |t - s| \le S$ and that $d/m < \alpha < 1$. Consequently, for $t \in [(k+4)/n, 1]$, we have shown that

$$\tilde{A}_{s,t} \lesssim n^{-(1+\alpha)/2+\varepsilon} |t - s|^{1/2+\varepsilon} S^{-d/(2m)}.$$

We now move to the case $t \in (s, (k+4)/n)$. We have two subcases, either $k \ge 1$ or $k = 0$. Suppose first that $k \ge 1$. We have

(54)
$$\tilde{A}_{s,t} = \int_s^{t \wedge \frac{k+1}{n}} \left\| \mathbf{E}^s \left(f_r(B_r) - f_r(B_{\kappa_n(r)}) \right) \right\|_{L_m(\Omega)} dr$$

$$+ \int_{t \wedge \frac{k+1}{n}}^t \left\| \mathbf{E}^s \left(f_r(B_r) - f_r(B_{\kappa_n(r)}) \right) \right\|_{L_m(\Omega)} dr.$$

Next, we see that

$$\int_s^{t \wedge \frac{k+1}{n}} \left\| \mathbf{E}^s \left(f_r(B_r) - f_r(B_{\kappa_n(r)}) \right) \right\|_{L_m(\Omega)} dr$$

$$= \int_s^{t \wedge \frac{k+1}{n}} \left\| \mathscr{P}_{r-s} f_r(B_s) - f_r(B_{k/n}) \right\|_{L_m(\Omega)} dr$$

$$\leq \int_s^{t \wedge \frac{k+1}{n}} \|\mathscr{P}_{r-s} f_r(B_s) - f_r(B_s)\|_{L_m(\Omega)} \, dr$$
$$+ \left(t \wedge \frac{k+1}{n} - s\right) \|f_r(B_s) - f_r(B_{k/n})\|_{L_m(\Omega)}.$$

For the first term at the right-hand side above we have by (15) and Proposition 2.6

$$\int_s^{t \wedge \frac{k+1}{n}} \|\mathscr{P}_{r-s} f_r(B_s) - f_r(B_s)\|_{L_m(\Omega)} \, dr \lesssim \int_s^{t \wedge \frac{k+1}{n}} s^{-d/(2m)} \|\mathscr{P}_{r-s} f_r - f_r\|_{L_m(\mathbb{R}^d)} \, dr$$
$$\lesssim \int_s^{t \wedge \frac{k+1}{n}} s^{-d/(2m)} |r-s|^{\alpha/2} \, dr$$
$$\lesssim |t-s|^{1+(\alpha/2)} s^{-d/(2m)}$$
$$\lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)},$$

where we have used that $|t-s| \leq 4n^{-1}$ in the last inequality. For the second term, by (16)

$$\left(t \wedge \frac{k+1}{n} - s\right) \|f_r(B_s) - f_r(B_{k/n})\|_{L_m(\Omega)} \lesssim |t-s| n^{-\alpha/2} (k/n)^{-d/(2m)}$$
$$\lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} s^{-d/(2m)}$$
$$\lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)},$$

where we have used that $|t-s| \leq 4n^{-1}$ and $k/n \leq s \leq 4k/n$. Hence,

$$(55) \quad \int_s^{t \wedge \frac{k+1}{n}} \|\mathbf{E}^s (f_r(B_r) - f_r(B_{\kappa_n(r)}))\|_{L_m(\Omega)} \, dr \lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)}.$$

For the second term at the right-hand side of (54) we can assume that $t > (k+1)/n$ and then we have by (16)

$$\int_{\frac{k+1}{n}}^t \|\mathbf{E}^s (f_r(B_r) - f_r(B_{\kappa_n(r)}))\|_{L_m(\Omega)} \, dr \leq \int_{\frac{k+1}{n}}^t \|f_r(B_r) - f_r(B_{\kappa_n(r)})\|_{L_m(\Omega)} \, dr$$
$$\lesssim \int_{\frac{k+1}{n}}^t |r - \kappa_n(r)|^{\alpha/2} (\kappa_n(r))^{-d/(2m)} \, dr$$
$$\lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)},$$

where we have used that $|t-s| \leq 4n^{-1}$ and $S \leq s \leq (k+1)/n \leq \kappa_n(r)$ for $r \geq (k+1)/n$. Consequently, for the case $t \in (s, (k+4)/n)$ and $k \geq 1$ we get

$$\tilde{A}_{s,t} \lesssim n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)}.$$

Finally, for the case $t \in (s, (k+4)/n)$ and $k = 0$ we have the following. Assume first that $\alpha < d/m$. From $\|f_r\|_{\mathbb{B}(\mathbb{R}^d)} \leq 1$, $r \in [0,1]$, we have the trivial estimate

$$\tilde{A}_{s,t} \leq 2|t-s|.$$

By definition of $[S,1]^*_{\leq}$ we have $|t-s| \leq S \leq s \leq n^{-1}$, and so one also has

$$|t-s| \leq n^{-1/2+\varepsilon} |t-s|^{1/2+\varepsilon} \leq n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-\alpha/2}$$
$$\leq n^{-(1+\alpha)/2+\varepsilon} |t-s|^{1/2+\varepsilon} S^{-d/(2m)}.$$

Consequently,

$$\tilde{A}_{s,t} \leq 2n^{-(1+\alpha)/2+\varepsilon}|t-s|^{1/2+\varepsilon}S^{-d/(2m)}.$$

If $\alpha > d/m$, we can use that $[f_r]_{C^{\alpha-d/m}(\mathbb{R}^d)} \lesssim [f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)}$ as before. We then get

$$\tilde{A}_{s,t} \lesssim n^{-\alpha/2+d/(2m)}|t-s| \sup_{r\in[0,1]} [f_r]_{C^{\alpha-d/m}(\mathbb{R}^d)} \leq n^{-\alpha/2+d/(2m)}|t-s|.$$

Since $S \leq n^{-1}$ we have that $n^{d/(2m)} \leq S^{-d/(2m)}$ which combined with $|t-s| \leq n^{-1}$ gives

$$\tilde{A}_{s,t} \lesssim n^{-(1+\alpha)/2+\varepsilon}|t-s|^{1/2+\varepsilon}S^{-d/(2m)}.$$

By combining all of the cases above, we can conclude that the bound

(56) $$\tilde{A}_{s,t} \lesssim n^{-(1+\alpha)/2+\varepsilon}|t-s|^{1/2+\varepsilon}S^{-d/(2m)}$$

holds for all $(s,t) \in [S,1]_{\leq}^*$, which in particular implies (22) for $A_{s,t}$ with $C_1 = Nn^{-(1+\alpha)/2+\varepsilon}$ (see (53)).

We now move to (23). We will show that it holds with $\varepsilon_2 = (\alpha'-1+2\varepsilon)/2 > 0$. Let $(s,t) \in [S,1]_{\leq}^*$ and let $u \in (s,t)$. We have

$$\mathbf{E}^s \delta A_{s,u,t} = \int_u^t \mathbf{E}^s\big((g_r(B_s) - g_r(B_u))\mathbf{E}^u\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\big)\,dr,$$

which implies that

$$\big\|\mathbf{E}^s \delta A_{s,u,t}\big\|_{L_p(\Omega)} \leq \int_u^t \big\|g_r(B_s) - g_r(B_u)\big\|_{L_{pm/(m-p)}(\Omega)} \big\|\mathbf{E}^u\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\big\|_{L_m(\Omega)}\,dr$$

$$\lesssim |s-u|^{\alpha'/2} \int_u^t \big\|\mathbf{E}^u\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\big\|_{L_m(\Omega)}\,dr$$

$$\lesssim |s-u|^{\alpha'/2}\tilde{A}_{u,t}$$

$$\lesssim |t-s|^{\alpha'/2}n^{-(1+\alpha)/2+\varepsilon}|t-s|^{1/2+\varepsilon}S^{-d/(2m)}$$

$$\lesssim n^{-(1+\alpha)/2+\varepsilon}|t-s|^{1+\varepsilon_2}S^{-d/(2m)},$$

where the first inequality follows from Hölder inequality, for the last inequality we have used (56) (notice that $(u,t) \in [S,1]_{\leq}^*$).

Summarising, we have shown that $A_{s,t}$ satisfies the conditions of Lemma 2.9 with $\delta_1 = \delta_2 = d/(2m)$, $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = (\alpha'-1+2\varepsilon)/2$, and $C_1 = C_2 = Nn^{-(1+\alpha)/2+\varepsilon}$. Consequently, by Lemma 2.9, there exists a unique process $\mathscr{A}_t$ satisfying (24) and (25). Let us now set

$$\bar{\mathscr{A}}_t := \int_0^t g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\,dr.$$

Since $\|g_r\|_{C^{\alpha'}(\mathbb{R}^d)}, \|f_r\|_{\mathbb{B}(\mathbb{R}^d)} \leq 1$, for $r \in [0,1]$, we have the trivial estimates

$$\|\bar{\mathscr{A}}_{s,t} - A_{s,t}\|_{L_p(\Omega)} \leq 4|t-s| \leq 4|t-s|^{1/2+\varepsilon},$$

$$\big\|\mathbf{E}^s(\bar{\mathscr{A}}_{s,t} - A_{s,t})\big\|_{L_p(\Omega)} \leq 2|t-s|^{1+\alpha'/2} \leq 2|t-s|^{1+\varepsilon_2},$$

which show that $\bar{\mathscr{A}}$ satisfies (24)–(25), and therefore $\mathscr{A} = \bar{\mathscr{A}}$. The desired inequality (52) now follows from (26). $\square$

LEMMA 3.4. *Let $\alpha \in (0, 1)$, $p \geq 2$, $\varepsilon \in (0, 1/2)$, $\alpha' \in (1 - 2\varepsilon, 1)$, and $m \geq d$ such that $m > p$. Then, for all $f \in \mathbb{B}([0, 1], \dot{W}_m^\alpha(\mathbb{R}^d)) \cap \mathbb{B}([0, 1] \times \mathbb{R}^d)$, and $g \in \mathbb{B}([0, 1], C^{\alpha'}(\mathbb{R}^d))$ one has the bound*

$$
(57) \quad \left\| \int_0^{\cdot} g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right\|_{L_p(\Omega; \mathbb{B}[0,1])}
$$
$$
\leq N \|g\|_{\mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))} \Big(\sup_{r \in [0,1]} [f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)} + \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \Big) n^{-(1+\alpha)/2 + \varepsilon},
$$

where N is a constant depending only on d, p, α, m, and ε.

PROOF. As before, we may and will assume

$$
\|g\|_{\mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))} = \sup_{r \in [0,1]} [f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)} + \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} = 1.
$$

We will first show that for $q \in (p, m)$ and all $t \in [0, 1]$ we have

$$
(58) \quad \left\| \int_0^t g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right\|_{L_q(\Omega)} \lesssim n^{-(1+\alpha)/2 + \varepsilon}.
$$

For any $\ell \in \mathbb{N}$, by Lemma 3.3 we get

$$
\left\| \int_{t2^{-\ell}}^t g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right\|_{L_q(\Omega)}
$$
$$
\leq \sum_{k=0}^{\ell-1} \left\| \int_{t2^{k-\ell}}^{t2^{k-\ell+1}} g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right\|_{L_q(\Omega)}
$$
$$
\lesssim n^{-(1+\alpha)/2 + \varepsilon} \sum_{k=0}^{\ell-1} (t2^{k-\ell})^{1/2+\varepsilon} (t2^{k-\ell})^{-d/(2m)}.
$$

Then, notice that since $d \leq m$, we have

$$
\sum_{k=0}^{\ell-1} (t2^{k-\ell})^{1/2+\varepsilon} (t2^{k-\ell})^{-d/(2m)} \leq 2^{-\ell\varepsilon} \sum_{k=0}^{\ell-1} 2^{k\varepsilon} = 2^{-\ell\varepsilon} \frac{2^{\ell\varepsilon} - 1}{2^\varepsilon - 1} \leq \frac{1}{2^\varepsilon - 1}.
$$

Hence, (58) follows from Fatou's lemma by letting $\ell \to \infty$. Since (58) holds for all $t \in [0, 1]$, we also get that for all $t \in [0, 1]$

$$
(59) \quad \left\| \int_t^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right\|_{L_q(\Omega)} \lesssim n^{-(1+\alpha)/2 + \varepsilon}.
$$

Let τ be a stopping time bounded by 1, taking only finitely many values $t_1, t_2, \ldots, t_k$. We have

$$
\mathbf{E}\left| \int_\tau^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right|^q
$$
$$
(60) \quad = \sum_{i=1}^k \mathbf{E}\left(\mathbf{1}_{\tau=t_i} \left| \int_{t_i}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right|^q \right).
$$

Define $\kappa_n^+(t_i) := \kappa_n(t_i) + 1/n$. For each of the summands on the right-hand side, we have

$$
\mathbf{E}\left(\mathbf{1}_{\tau=t_i} \left| \int_{t_i}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right|^q \right)
$$
$$
(61) \quad \lesssim \mathbf{E}\left(\mathbf{1}_{\tau=t_i} \left| \int_{\kappa_n^+(t_i)}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr \right|^q \right) + n^{-q} \mathbf{P}(\tau = t_i),
$$

using $\|g_r\|_{\mathbb{B}}, \|f_r\|_{\mathbb{B}} \le 1, r \in [0,1]$, in a trivial way. If $1 - t_i \le 3n^{-1}$, then we have similarly the trivial estimate

$$(62) \qquad \mathbf{E}\left(\mathbf{1}_{\tau = t_i}\left|\int_{\kappa_n^+(t_i)}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr\right|^q\right) \lesssim n^{-q}\mathbf{P}(\tau = t_i).$$

If $1 - t_i > 3n^{-1}$ we can write

$$(63) \qquad \mathbf{E}\left(\mathbf{1}_{\tau = t_i}\left|\int_{\kappa_n^+(t_i)}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr\right|^q\right) = \mathbf{E}(\mathbf{1}_{\tau = t_i} G(B_{\kappa_n^+(t_i)})),$$

where

$$G(x) := \mathbf{E}\left|\int_{\kappa_n^+(t_i)}^1 g_r(B_r - B_{\kappa_n^+(t_i)} + x)\right.$$

$$\left. \times \big(f_r(B_r - B_{\kappa_n^+(t_i)} + x) - f_r(B_{\kappa_n(r)} - B_{\kappa_n^+(t_i)} + x)\big)\, dr\right|^q$$

$$= \mathbf{E}\left|\int_0^{1 - \kappa_n^+(t_i)} g_{r + \kappa_n^+(t_i)}(B_r + x)\big(f_{r + \kappa_n^+(t_i)}(B_r + x) - f_{r + \kappa_n^+(t_i)}(B_{\kappa_n(r)} + x)\big)\, dr\right|^q.$$

Hence, by (58), we conclude that in the case $1 - t_i > 3n^{-1}$ we have that

$$\mathbf{E}\left(\mathbf{1}_{\tau = t_i}\left|\int_{\kappa_n^+(t_i)}^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr\right|^q\right) = \mathbf{E}(\mathbf{1}_{\tau = t_i} G(B_{\kappa_n^+(t_i)}))$$

$$\lesssim \big(n^{-(1+\alpha)/2 + \varepsilon}\big)^q \mathbf{P}(\tau = t_i).$$

Putting the above inequality together with (62), (61), and (60), gives

$$(64) \qquad \left\|\int_\tau^1 g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr\right\|_{L_q(\Omega)} \lesssim n^{-(1+\alpha)/2 + \varepsilon}.$$

Recall that $\tau \le 1$ was a simple stopping time. It is well known that an arbitrary stopping time can be approximated by simple ones (e.g., one can take $\tau_\ell = \kappa_\ell^+(\tau)$ and let $\ell \to \infty$). Therefore, a standard approximation argument shows that (64) holds for all stopping times which are bounded by 1. Moreover, the above combined with (58) implies that for all such stopping times τ, we have

$$\left\|\int_0^\tau g_r(B_r)\big(f_r(B_r) - f_r(B_{\kappa_n(r)})\big)\, dr\right\|_{L_q(\Omega)} \lesssim n^{-(1+\alpha)/2 + \varepsilon}.$$

The claimed bound (57) then follows by Lenglart's inequality (see, e.g., [38], Proposition IV.4.7). $\square$

COROLLARY 3.5. *Let $\alpha \in (0,1)$, $p \ge 2$, $\varepsilon \in (0, 1/2)$, $\alpha' \in (1 - 2\varepsilon, 1)$, and $m \ge d$ such that $m > p$. Let Assumption 1.4 hold and let X^n be the solution of (2). Then, for all $f \in \mathbb{B}([0,1], \dot{W}_m^\alpha(\mathbb{R}^d)) \cap \mathbb{B}([0,1] \times \mathbb{R}^d)$, $g \in \mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))$, and $n \in \mathbb{N}$ one has the bound*

$$\left\|\int_0^\cdot g_r(X_r^n)\big(f_r(X_r^n) - f_r(X_{\kappa_n(r)}^n)\big)\, dr\right\|_{L_p(\Omega; \mathbb{B}[0,1])}$$

$$(65)$$

$$\le N\|g\|_{\mathbb{B}([0,1], C^{\alpha'}(\mathbb{R}^d))}\left(\sup_{r \in [0,1]} [f_r]_{\dot{W}_m^\alpha(\mathbb{R}^d)} + \|f\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)}\right) n^{-(1+\alpha)/2 + \varepsilon},$$

where N is a constant depending only on d, p, α, m, and ε.

The proof is a simple application of Girsanov's theorem and works just like the proof of Corollary 3.2, so we omit repeating the details.

4. An intermediate stability estimate. In this section we consider two "approximating solutions" to the main SDE (1). More precisely, we assume that we are given adapted continuous processes $X, \bar{X}, Y, \bar{Y}$, all of them with initial condition x_0, such that

$$dX_t = b(\bar{X}_t) \, dt + \sigma(\bar{X}_t) \, dB_t,$$

$$dY_t = b(\bar{Y}_t) \, dt + \sigma(\bar{Y}_t) \, dB_t,$$

and such that the laws of X_t and Y_t are absolutely continuous with respect to the Lebesgue measure for all $t > 0$. We furthermore denote $\hat{X} = X - \bar{X}$, $\hat{Y} = Y - \bar{Y}$. To relate with Euler–Maruyama scheme (2), one may think of $\bar{X}, \bar{Y}$ respectively as X and $X^n_{\kappa_n}$. This kind of reformulation of the error analysis is inspired by [18, 20, 25], and in some vague sense, replaces the "regularisation lemma" step from [7]. Since the coefficients b and σ are bounded, we have that there exists N depending only on d, $\|b\|_{\mathbb{B}}$, $\|\sigma\|_{\mathbb{B}}$, such that

(66) $\qquad \mathbf{E} \exp(\|X.\|_{\mathbb{B}([0,1])}) \leq N \exp(x_0), \qquad \mathbf{E} \exp(\|Y.\|_{\mathbb{B}([0,1])}) \leq N \exp(x_0).$

Estimating the difference of the drifts is done via a PDE method, similar to, for example, [5, 9, 30, 34]. First, for $K \in (0, \infty)$, we introduce the truncation $b_K = b\mathbf{1}_{|x| \leq K}$. The reason for this truncation is to enforce the right-hand side of the PDE below to be in $L_p(\mathbb{R}^d)$ with $p < \infty$, since Schauder estimates fail in the endpoint $p = \infty$ case. For $\ell \in \{1, \dots, d\}$, and $\theta > 0$ to be chosen later, let us consider the equation

(67)
$$\partial_t u^\ell + \frac{1}{2}(\sigma \sigma^*) \cdot \nabla^2 u^\ell + b \cdot \nabla u^\ell - \theta u^\ell = b_K^\ell \quad \text{in } (0, 1) \times \mathbb{R}^d$$

$$u^\ell = 0 \quad \text{on } \{1\} \times \mathbb{R}^d.$$

By u we denote the $\mathbb{R}^d$-valued function whose coordinates are $u^1, \dots, u^d$. Note that by a change of time variable $t \leftrightarrow 1 - t$, the estimates in Lemma 2.10 also apply for the backward equation (67).

For a function f let us denote by $\mathcal{M}f$ its Hardy–Littlewood maximal function (see, e.g., [1] for a brief introduction), that is,

(68)
$$\mathcal{M}f(x) := \sup_{r > 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} f(y) \, dy, \quad x \in \mathbb{R}^d.$$

Given the objects above and $p \geq 2$, introduce the increasing process

(69)
$$A_t = t + \int_0^t \left| (\mathcal{M}|\nabla(\nabla u \sigma)|)(s, X_s) + (\mathcal{M}|\nabla(\nabla u \sigma)|)(s, Y_s) \right|^p ds.$$

LEMMA 4.1. *Assume the above setting and fix $p \geq 2$, $m \in \mathbb{N}$. Then there exist constants N, θ depending only on $d, p, \|b\|_{\mathbb{B}}$, and $\|\sigma\|_{\mathbb{B}}$ (but not on K and m) such that*

(70)
$$\mathbf{E} \sup_{t \in [0,1]} |X_t - Y_t|^p \leq N \left(\mathbf{P}\left(A_1 \geq \frac{m}{2N} \right) \right)^{1/2} + N^m \mathcal{R},$$

where

$$\mathcal{R} = \left(e^{-K} + \mathbf{1}_{\nabla \sigma \neq 0} \sum_{U = X, Y} \sup_{t \in [0,1]} (\mathbf{E}(|\hat{U}_t|^{2p}))^{1/2} \left(1 + \mathbf{E} \int_0^1 |\nabla^2 u(s, U_s)|^{2p} ds \right)^{1/2} \right.$$

(71)
$$+ \sum_{U = X, Y} \mathbf{E} \sup_{t \in [0,1]} \left| \int_0^t (b(U_s) - b(\bar{U}_s)) \, ds \right|^p$$

$$\left. + \left| \int_0^t (b(U_s) - b(\bar{U}_s)) \nabla u(s, U_s) \, ds \right|^p \right),$$

and u is a solution to (67).

PROOF. Let τ_1, τ_2 be stopping times with $0 \le \tau_1 \le \tau_2 \le T$. Denote $Z_t = |X_t - Y_t|^p$. Notice that by (66), and Markov's inequality one has

$$\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t b(X_s) - b_K(X_s)\, ds \right|^p \lesssim \mathbf{P}\Big(\sup_{t \in [0,1]} |X_t| \ge K \Big) \lesssim e^{-K},$$

and similarly for Y. Therefore, repeated application of the triangle inequality yields

$$
\begin{aligned}
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} Z_t &\lesssim \mathbf{E} Z_{\tau_1} + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t b(\bar{X}_s) - b(\bar{Y}_s)\, ds \right|^p \\
&\quad + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t \sigma(\bar{X}_s) - \sigma(\bar{Y}_s)\, dB_s \right|^p \\
&\lesssim \mathbf{E} Z_{\tau_1} + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t b(X_s) - b(Y_s)\, ds \right|^p \\
&\quad + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t \sigma(\bar{X}_s) - \sigma(\bar{Y}_s)\, dB_s \right|^p + \mathscr{R} \\
&\lesssim \mathbf{E} Z_{\tau_1} + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t b_K(X_s) - b_K(Y_s)\, ds \right|^p \\
&\quad + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t \sigma(\bar{X}_s) - \sigma(\bar{Y}_s)\, dB_s \right|^p + \mathscr{R}.
\end{aligned}
\tag{72}
$$

By the Burkholder–Davis–Gundy inequality and the Lipschitz continuity of σ we have

$$
\begin{aligned}
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t \sigma(\bar{X}_s) - \sigma(\bar{Y}_s)\, dB_s \right|^p &\lesssim \mathbf{E}\left(\int_{\tau_1}^{\tau_2} \big(\sigma(\bar{X}_s) - \sigma(\bar{Y}_s)\big)^2\, ds \right)^{p/2} \\
&\lesssim \mathbf{E} \int_{\tau_1}^{\tau_2} Z_s\, ds + \mathbf{1}_{\nabla \sigma \ne 0} \sup_{t \in [0,1]} \mathbf{E}\big(|\hat{X}_t|^p + |\hat{Y}_t|^p \big).
\end{aligned}
$$

Therefore we arrive at

$$
(73) \quad \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} Z_t \lesssim \mathbf{E} Z_{\tau_1} + \mathbf{E} \int_{\tau_1}^{\tau_2} Z_s\, ds + \mathbf{E} \sup_{t \in [\tau_1, \tau_2]} \left| \int_{\tau_1}^t b_K(X_s) - b_K(Y_s)\, ds \right|^p + \mathscr{R}.
$$

The integral involving b_K is treated via a PDE method. Although u^ℓ is not spatially twice continuously differentiable, one has $u^\ell \in W_q^{1,2}([0,1] \times \mathbb{R}^d)$ for all $q < \infty$ by Lemma 2.10. Therefore, Itô's formula can be applied (see, e.g., [23], Theorem 1, p. 122). Hence for any $\ell \in \{1, \ldots, d\}$ and $U \in \{X, Y\}$, on $\{t \ge \tau_1\}$ we have from (67) and Itô formula that

$$
\begin{aligned}
u^\ell(t, U_t) - u^\ell(\tau_1, U_{\tau_1}) &= \int_{\tau_1}^t \big(\theta u^\ell(s, U_s) + b_K^\ell(U_s)\big)\, ds + \int_{\tau_1}^t \nabla u^\ell(s, U_s) \sigma(U_s)\, dB_s \\
&\quad + \mathscr{E}_1^{U,\ell}(t) + \mathscr{E}_2^{U,\ell}(t) + \mathscr{E}_3^{U,\ell}(t),
\end{aligned}
$$

where

$$\mathscr{E}_1^{U,\ell}(t) := \int_{\tau_1}^t \nabla u^\ell(s, U_s) \cdot \big(b(\bar{U}_s) - b(U_s)\big)\, ds,$$

$$\mathscr{E}_2^{U,\ell}(t) := \int_{\tau_1}^t \nabla^2 u^\ell(s, U_s) \cdot \big((\sigma\sigma^*)(\bar{U}_s) - (\sigma\sigma^*)(U_s)\big)\, ds,$$

$$\mathscr{E}_3^{U,\ell}(t) := \int_{\tau_1}^t \nabla u^\ell(s, U_s)\big(\sigma(\bar{U}_s) - \sigma(U_s)\big)\, dB_s.$$

It follows that

$$
\int_{\tau_1}^{t} b_K(X_s) - b_K(Y_s)\, ds = u(t, X_t) - u(t, Y_t) - u(\tau_1, X_{\tau_1}) + u(\tau_1, Y_{\tau_1})
$$

(74)
$$
- \int_{\tau_1}^{t} \theta\big(u(s, X_s) - u(s, Y_s)\big)\, ds
$$

$$
- \int_{\tau_1}^{t} \big(\nabla u(s, X_s)\sigma(X_s) - \nabla u(s, Y_s)\sigma(Y_s)\big)\, dB_s
$$

$$
- \mathscr{E}_1^X(t) - \mathscr{E}_2^X(t) - \mathscr{E}_3^X(t) + \mathscr{E}_1^Y(t) + \mathscr{E}_2^Y(t) + \mathscr{E}_3^Y(t).
$$

For the first couple of terms in (74) we apply (29), keeping in mind that b_K (playing the role of f therein) has its L_∞ norm bounded by $\|b\|_{\mathbb{B}}$, independently of K. Therefore, we have $\|\nabla u\|_{\mathbb{B}([0,1]\times\mathbb{R}^d)} \leq N\theta^{-1/2}$, and so

$$
\sup_{t\in[\tau_1,\tau_2]} \big|u(t, X_t) - u(t, Y_t) - u(\tau_1, X_{\tau_1}) + u(\tau_1, Y_{\tau_1})\big|
$$

(75)
$$
+ \theta \int_{\tau_1}^{t} \big|u(s, X_s) - u(s, Y_s)\big|\, ds
$$

$$
\lesssim \theta^{-1/2} \sup_{t\in[\tau_1,\tau_2]} |X_t - Y_t| + \theta^{1/2} \int_{\tau_1}^{t} |X_s - Y_s|\, ds.
$$

Note that we have $\mathbf{E}\sup_{t\in[\tau_1,\tau_2]} Z_t < \infty$. Therefore, if we combine (73), (74), (75), and choose θ to be large enough, we get

$$
\mathbf{E}\sup_{t\in[\tau_1,\tau_2]} Z_t \lesssim \mathbf{E} Z_{\tau_1} + \mathbf{E}\int_{\tau_1}^{\tau_2} Z_s\, ds + \mathscr{R}
$$

(76)
$$
+ \mathbf{E}\sup_{t\in[\tau_1,\tau_2]}\left|\int_{\tau_1}^{t} \nabla u(s, X_s)\sigma(X_s) - \nabla u(s, Y_s)\sigma(Y_s)\, dB_s\right|^p
$$

$$
+ \sum_{i=1,2,3;\, U=X,Y} \mathbf{E}\sup_{t\in[\tau_1,\tau_2]} |\mathscr{E}_i^U(t)|^p.
$$

In the sequel, we suppress the time argument from u whenever there is no danger of confusion. The next term to deal with is the stochastic integral, which requires some care due to the lack of Lipschitz continuity of ∇u. We argue as in [5], using the Hardy–Littlewood maximal function (see (68)). One then has the following d such that for all $f \in W^1_{1,\mathrm{loc}}(\mathbb{R}^d)$, for almost all $x, y \in \mathbb{R}^d$:

(77)
$$
|f(x) - f(y)| \leq N|x - y|\big(\mathcal{M}|\nabla f|(x) + \mathcal{M}|\nabla f|(y)\big).
$$

Recall also the Hardy–Littlewood maximal inequality ([1]),

(78)
$$
\|\mathcal{M}f\|_{L_p(\mathbb{R}^d)} \leq N\|f\|_{L_p(\mathbb{R}^d)},
$$

for all $f \in L_p(\mathbb{R}^d)$, $p \in (1,\infty)$, where N depends only on d and p. Since by assumption the laws of X_t and Y_t are absolutely continuous with respect to the Lebesgue measure for all $t > 0$, (77) holds with X_t and Y_t in place of x and y, $d\mathbf{P} \otimes dt$-almost surely. Combining this with the Burkholder–Davis–Gundy inequality, we have

$$
\mathbf{E}\sup_{t\in[\tau_1,\tau_2]}\left|\int_{\tau_1}^{t} \nabla u(X_s)\sigma(X_s) - \nabla u(Y_s)\sigma(Y_s)\, dB_s\right|^p
$$

$$
\lesssim \mathbf{E}\left|\int_{\tau_1}^{\tau_2} |\nabla u(X_s)\sigma(X_s) - \nabla u(Y_s)\sigma(Y_s)|^2\, ds\right|^{p/2}
$$

(79)
$$
\lesssim \mathbf{E} \int_{\tau_1}^{\tau_2} |X_s - Y_s|^p |(\mathcal{M}|\nabla(\nabla u \, \sigma)|(X_s) + (\mathcal{M}|\nabla(\nabla u \, \sigma)|(Y_s)|^p \, ds
$$
$$
\leq \mathbf{E} \int_{\tau_1}^{\tau_2} Z_s \, dA_s,
$$

where in the last equality, we used the definition (69) of the process A. We now move on to bounding the terms $\mathscr{E}_i^U$. The bound

(80)
$$
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} |\mathscr{E}_1^U(t)|^p \leq \mathscr{R}
$$

is immediate. Next, by the Lipschitz continuity of σ and Hölder's inequality, we have

(81)
$$
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} |\mathscr{E}_2^U(t)|^p \lesssim \mathbf{1}_{\nabla \sigma \neq 0} \int_0^1 \mathbf{E}|\nabla^2 u^\ell(U_s)|^p |\hat{U}_s|^p \, ds \leq \mathscr{R}.
$$

Finally, by the Burkholder–Davis–Gundy and Jensen's inequalities and the Lipschitz continuity of σ, we have

(82)
$$
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} |\mathscr{E}_3^U(t)|^p \lesssim \mathbf{1}_{\nabla \sigma \neq 0} \int_0^1 \|\nabla u\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \mathbf{E}|\hat{U}_s|^p \, ds \leq \mathscr{R}
$$

since $\|\nabla u\|_{\mathbb{B}([0,1] \times \mathbb{R}^d)} \lesssim 1$. We can now combine (76), (79), (80), (81), and (82) altogether to get

$$
\mathbf{E} \sup_{t \in [\tau_1, \tau_2]} Z_t \leq N \mathbf{E} Z_{\tau_1} + N \mathbf{E} \int_{\tau_1}^{\tau_2} Z_s \, dA_s + N \mathscr{R}.
$$

This brings us to the setting of Lemma 2.11. From (32), we therefore obtain

$$
\mathbf{E} \sup_{t \in [0,1]} Z_t \leq N^m \mathscr{R} + N \Big(\mathbf{E} \sup_{t \in [0,1]} Z_t^2 \Big)^{1/2} \Big(\mathbf{P}\Big(A_1 \geq \frac{m}{2N}\Big) \Big)^{1/2}.
$$

Since by (66), $\mathbf{E} \sup_{t \in [0,1]} Z_t^2 \leq N$, this is precisely the claimed bound. □

5. Proofs of the main results. First we recall the following estimate on the density of the Euler–Maruyama scheme due to Gyöngy and Krylov [14], Theorem 4.2. We remark that while therein this bound is proved for the $b = 0$ case, the general case follows immediately by means of Girsanov's theorem.

LEMMA 5.1. *Let $p \in (1, \infty]$. Under Assumption* 1.1, *there exists N depending only on $p, d, \lambda, \|\sigma\|_{C^2}$, and $\|b\|_{\mathbb{B}}$ such that for all $G \in L_p(\mathbb{R}^d)$ and $t \in (0, 1]$*

$$
|\mathbf{E}G(X_t^n)| \leq N \|G\|_{L_p(\mathbb{R}^d)} t^{-d/(2p)}.
$$

From Lemma 5.1, using Khas'minskii's argument [16] one can get estimates for exponential moments. For the adaptation of Khas'minskii's argument for the process X^n we refer to [5], Lemma 2.3, or alternatively to [27], Lemma 5.14.

LEMMA 5.2. *Let $q > (d + 2)/2$. There exist $\beta_q, \gamma_q \in (0, \infty)$ such that for all $\mu > 0$*

$$
\mathbf{E} \exp\Big(\mu \int_0^1 |f(X_s)| + |f(X_s^n)| \, ds \Big) \leq \exp(\beta_q(1 + (\mu \|f\|_{L_q((0,1) \times \mathbb{R}^d)})^{\gamma_q})).
$$

Moreover $\lim_{q \to \infty} \gamma_q = 1$, *and there exists $\beta = \beta(\lambda, \|\sigma\|_{C^2}, \|b\|_{\mathbb{B}}, d) \in \mathbb{R}$ such that* $\lim_{q \to \infty} \beta_q = \beta$.

PROOF OF THEOREM 1.5. We apply Lemma 4.1 with X being the solution of the main SDE (1) (and so $\bar{X} = X$, $\hat{X} = 0$) and $Y = X^n$ being the solution of the approximate equation (2) (and so $\bar{Y}_t = X^n_{\kappa_n(t)}$, $\hat{Y}_t = X^n_t - X^n_{\kappa_n(t)}$). Our task is therefore to choose the parameters K, m, so that the right-hand side of (70) can be bounded by $Nn^{-p((1+\alpha)/2-\varepsilon)}$. Note that in the case of Theorem 1.5 $\mathbf{1}_{\nabla\sigma\neq 0} = 0$, eliminating one term from the right-hand side of (71).

It will be convenient to introduce two further parameters $\mu, q \in (1, \infty)$. By Markov's inequality, we have,

$$\mathbf{P}\left(A_1 \geq \frac{m}{2N}\right) \leq \exp(-\mu m/2N)\mathbf{E}\exp(\mu A_1).$$

The maximal inequality (78) and Lemma 5.2 imply that

$$\mathbf{E}\exp(\mu A_1) \leq \exp(\beta_q(1 + (\mu\||\nabla^2 u|^p\|_{L_q((0,1)\times\mathbb{R}^d)})^{\gamma_q})).$$

By Lemma 2.10 we have

$$\|u\|_{W_r^{1,2}} \leq N(r)\|b_K\|_{L_r(\mathbb{R}^d)} \leq K^{d/r},$$

and therefore, for any sufficiently large q,

$$(83) \qquad \||\nabla^2 u|^p\|_{L_q((0,1)\times\mathbb{R}^d)} \leq N(q)K^{d/q}, \qquad \|\nabla u\|_{\mathbb{B}([0,1],C^{1-\varepsilon}(\mathbb{R}^d))} \leq N(q)K^{d/q},$$

where the second inequality follows from Sobolev embedding. Finally, notice that the last terms in (71) are precisely the ones which were estimated in Corollary 3.5, with the choice $f = b$ and $g = 1$ or $g = \nabla u$. Consequently, we obtain from (70) of Lemma 4.1 that

$$\mathbf{E}\sup_{t\in[0,1]}|X_t - X^n_t|^p \leq N\exp(-\mu m/2N)\exp((\mu\beta_q N(q)K^{d/q})^{\gamma_q})$$

$$+ N^m\exp(-K) + N^m N(q)n^{-p((1+\alpha)/2-\varepsilon)}K^{d/q}.$$

Choose $K = p\ln n$ and q large enough so that $d\gamma_q/q \leq 1/2$. Then, the above bound implies

$$\mathbf{E}\sup_{t\in[0,1]}|X_t - X^n_t|^p \leq N\exp(-\mu m/2N)\exp(\mu N(\ln n)^{1/2}) + N^m n^{-p((1+\alpha)/2-2\varepsilon)}.$$

Now choose $m = \lfloor\frac{p\varepsilon\ln n}{\ln N}\rfloor$, so that $N^m \leq n^{p\varepsilon}$. Then by choosing μ sufficiently large one can achieve $\exp(-\mu m/2N) \leq n^{-\gamma}$ for any exponent γ, which then yields the required bound

$$\mathbf{E}\sup_{t\in[0,1]}|X_t - X^n_t|^p \leq Nn^{-p((1+\alpha)/2-3\varepsilon)},$$

completing the proof. $\square$

PROOF OF THEOREM 1.2. Similar to the previous proof, we apply Lemma 4.1. This time $\mathbf{1}_{\nabla\sigma\neq 0} = 1$, so we have one more term to bound. However, we are aiming only at a bound of order $Nn^{-p(1/2-\varepsilon)}$. Let $q \in (1, \infty)$ and $v \in L_q([0,1] \times \mathbb{R}^d)$. Then from Lemma 5.1 we have

$$(84) \qquad \mathbf{E}\int_0^1 |v(s, X^n_s)|\,ds \leq N(q)\int_0^1 \|v(s,\cdot)\|_{L_q(\mathbb{R}^d)}s^{-\frac{d}{2q}}\,ds$$

$$\leq N(q)\|v\|_{L_q([0,1]\times\mathbb{R}^d)}\left(\int_0^1 s^{-\frac{d}{2(q-1)}}\,ds\right)^{\frac{q-1}{q}}.$$

For every $q > d/2 + 1$ the integral is finite. Similarly, we get, for the same range of q,

$$(85) \qquad \mathbf{E}\int_{1/n}^1 |v(s, X^n_{\kappa_n(s)})|\,ds \leq N(q)\|v\|_{L_q([0,1]\times\mathbb{R}^d)}.$$

Therefore, recalling that in the context of (70) we have $\hat{X} = 0$, $\hat{Y}_t = X_t^n - X_{\kappa_n(t)}^n$, we can write

$$\sup_{t\in[0,1]} (\mathbf{E}|X_t^n - X_{\kappa_n(t)}^n|^{2p})^{1/2} \left(1 + \mathbf{E}\int_0^1 |\nabla^2 u(s, X_s^n)|^{2p}\, ds\right)^{1/2}$$

(86)
$$\leq N(q)n^{-p/2}\left(1 + \||\nabla^2 u|^{2p}\|_{L_q([0,1]\times\mathbb{R}^d)}\right)^{1/2}$$

$$\leq N(q)n^{-p/2}K^{d/(2q)},$$

where we used (83) in the last step. Concerning the remaining terms

$$\mathscr{R}_1 := \mathbf{E} \sup_{t\in[0,1]} \left|\int_0^t (b(X_s^n) - b(X_{\kappa_n(s)}^n))\, ds\right|^p,$$

$$\mathscr{R}_2 := \mathbf{E} \sup_{t\in[0,1]} \left|\int_0^t (b(X_s^n) - b(X_{\kappa_n(s)}^n))\nabla u(s, X_s^n)\, ds\right|^p,$$

we can treat $\mathscr{R}_1$ just as before, using the results of Section 3, more precisely Corollary 3.2 with $f = b$. This yields the bound $\mathscr{R}_1 \leq Nn^{-p(1/2-\varepsilon)}$. For $\mathscr{R}_2$ we argue slightly differently, since, unlike in Corollary 3.5, there is no "weight" function g in Corollary 3.2 (although it would not be too difficult to include, but we choose to have at least one of the integral estimates free of the tedium of weights). Instead we write

$$\mathscr{R}_2 \leq N(\mathscr{R}_3 + \mathscr{R}_4 + \mathscr{R}_5)$$

$$:= N\left(\mathbf{E} \sup_{t\in[0,1]} \left|\int_0^t ((b\nabla u)(s, X_s^n) - (b\nabla u)(s, X_{\kappa_n(s)}^n))\, ds\right|^p\right.$$

$$+ \mathbf{E} \sup_{t\in[1/n,1]} \left|\int_{1/n}^t (\nabla u(s, X_s^n) - \nabla u(s, X_{\kappa_n(s)}^n))b(X_{\kappa_n(s)}^n)\, ds\right|^p$$

$$\left. + \mathbf{E} \sup_{t\in[0,1/n]} \left|\int_0^t (\nabla u(s, X_s^n) - \nabla u(s, X_{\kappa_n(s)}^n))b(X_{\kappa_n(s)}^n)\, ds\right|^p\right).$$

The term $\mathscr{R}_3$ falls within the scope of Corollary 3.2 with $f = b\nabla u$, yielding the bound $\mathscr{R}_3 \leq Nn^{-p(1/2+\varepsilon)}$. The bound $\mathscr{R}_5 \leq Nn^{-p}$ is trivial. It remains to bound $\mathscr{R}_4$. Since for $s \geq 1/n$, X_s^n and $X_{\kappa_n(s)}^n$ both have densities, we can apply (77) to get

$$\mathscr{R}_4 \leq N\mathbf{E}\int_{1/n}^1 |\nabla u(s, X_s^n) - \nabla u(s, X_{\kappa_n(s)}^n))|^p\, ds$$

$$\leq N\mathbf{E}\int_{1/n}^1 |X_s^n - X_{\kappa_n(s)}^n|^p (|\mathscr{M}|\nabla^2 u(s, X_s^n)| + |\mathscr{M}|\nabla^2 u(s, X_{\kappa_n(s)}^n)|)^p\, ds$$

$$\leq N\left(\sup_{s\in[0,1]} \mathbf{E}|X_s^n - X_{\kappa_n(s)}^n|^{2p}\right)^{1/2}$$

$$\times \left(\int_{1/n}^1 \mathbf{E}(|\mathscr{M}|\nabla^2 u(s, X_s^n)| + |\mathscr{M}|\nabla^2 u(s, X_{\kappa_n(s)}^n)|)^{2p}\, ds\right)^{1/2}$$

$$\leq Nn^{-p/2}\left(\int_{1/n}^1 \mathbf{E}(|\mathscr{M}|\nabla^2 u(s, X_s^n)| + |\mathscr{M}|\nabla^2 u(s, X_{\kappa_n(s)}^n)|)^{2p}\, ds\right)^{1/2}.$$

By (78), (83), (84), and (85), the bound $\mathscr{R}_4 \leq N(q)n^{-p/2}K^{1/2q}$ follows. Combining all of the above, (70) then implies

$$\mathbf{E} \sup_{t\in[0,1]} |X_t - X_t^n|^p \leq N\mathbf{P}\left(A_1 \geq \frac{m}{2N}\right) + N^m(e^{-K} + N(q)n^{-p(1/2-\varepsilon)}K^{d/q}).$$

From here it remains to tune the parameter K, q, m, which is done similarly as in the previous proof, yielding the bound (4). □

REFERENCES

[1] AALTO, D. and KINNUNEN, J. (2009). Maximal functions in Sobolev spaces. In *Sobolev Spaces in Mathematics. I. Int. Math. Ser. (N. Y.)* **8** 25–67. Springer, New York. MR2508838 https://doi.org/10.1007/978-0-387-85648-3_2

[2] ALTMEYER, R. (2021). Approximation of occupation time functionals. *Bernoulli* **27** 2714–2739. MR4303901 https://doi.org/10.3150/21-BEJ1328

[3] ATHREYA, S., BUTKOVSKY, O., LÊ, K. and MYTNIK, L. (2022). Well-posedness of stochastic heat equation with distributional drift and skew stochastic heat equation. *Comm. Pure Appl. Math.* To appear.

[4] BAO, J., HUANG, X. and YUAN, C. (2019). Convergence rate of Euler–Maruyama scheme for SDEs with Hölder–Dini continuous drifts. *J. Theoret. Probab.* **32** 848–871. MR3959630 https://doi.org/10.1007/s10959-018-0854-9

[5] BAO, J., HUANG, X. and ZHANG, S.-Q. (2020). Convergence rate of EM algorithm for SDEs under integrability condition. ArXiv E-prints. Available at arXiv:2009.04781.

[6] BELLINGERI, C., FRIZ, P. K. and GERENCSÉR, M. (2021). Singular paths spaces and applications. *Stoch. Anal. Appl.* 1–24.

[7] BUTKOVSKY, O., DAREIOTIS, K. and GERENCSÉR, M. (2021). Approximation of SDEs: A stochastic sewing approach. *Probab. Theory Related Fields* **181** 975–1034. MR4344136 https://doi.org/10.1007/s00440-021-01080-2

[8] CATELLIER, R. and GUBINELLI, M. (2016). Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.* **126** 2323–2366. MR3505229 https://doi.org/10.1016/j.spa.2016.02.002

[9] DAREIOTIS, K. and GERENCSÉR, M. (2020). On the regularisation of the noise for the Euler–Maruyama scheme with irregular drift. *Electron. J. Probab.* **25** Paper No. 82. MR4125787 https://doi.org/10.1214/20-ejp479

[10] DE ANGELIS, T., GERMAIN, M. and ISSOGLIO, E. (2019). A numerical scheme for stochastic differential equations with distributional drift. ArXiv e-prints.

[11] FLANDOLI, F., GUBINELLI, M. and PRIOLA, E. (2010). Well-posedness of the transport equation by stochastic perturbation. *Invent. Math.* **180** 1–53. MR2593276 https://doi.org/10.1007/s00222-009-0224-4

[12] FLANDOLI, F., PRIOLA, E. and ZANCO, G. (2019). A mean-field model with discontinuous coefficients for neurons with spatial interaction. *Discrete Contin. Dyn. Syst.* **39** 3037–3067. MR3959420 https://doi.org/10.3934/dcds.2019126

[13] FRIEDMAN, A. (1964). *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Inc., Englewood Cliffs, NJ. MR0181836

[14] GYÖNGY, I. and KRYLOV, N. (1996). Existence of strong solutions for Itô's stochastic equations via approximations. *Probab. Theory Related Fields* **105** 143–158. MR1392450 https://doi.org/10.1007/BF01203833

[15] HAIRER, M., HUTZENTHALER, M. and JENTZEN, A. (2015). Loss of regularity for Kolmogorov equations. *Ann. Probab.* **43** 468–527. MR3305998 https://doi.org/10.1214/13-AOP838

[16] KHAS'MINSKIĬ, R. Z. (1959). On positive solutions of the equation $\mathcal{U} + Vu = 0$. *Theory Probab. Appl.* **4** 309–318. MR0123373 https://doi.org/10.1137/1104030

[17] ICHIBA, T., PAPATHANAKOS, V., BANNER, A., KARATZAS, I. and FERNHOLZ, R. (2011). Hybrid atlas models. *Ann. Appl. Probab.* **21** 609–644. MR2807968 https://doi.org/10.1214/10-AAP706

[18] JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. MR1617049 https://doi.org/10.1214/aop/1022855419

[19] KOHATSU-HIGA, A., MAKHLOUF, A. and NGO, H. L. (2014). Approximations of non-smooth integral type functionals of one dimensional diffusion processes. *Stochastic Process. Appl.* **124** 1881–1909. MR3170228 https://doi.org/10.1016/j.spa.2014.01.003

[20] KRYLOV, N. V. (1990). A simple proof of the existence of a solution to the Itô equation with monotone coefficients. *Teor. Veroyatn*. *Primen*. **35** 576–580. MR1091217 https://doi.org/10.1137/1135082

[21] KRYLOV, N. V. (1996). *Lectures on Elliptic and Parabolic Equations in Hölder Spaces*. *Graduate Studies in Mathematics* **12**. Amer. Math. Soc., Providence, RI. MR1406091 https://doi.org/10.1090/gsm/012

[22] KRYLOV, N. V. (2008). *Lectures on Elliptic and Parabolic Equations in Sobolev Spaces*. *Graduate Studies in Mathematics* **96**. Amer. Math. Soc., Providence, RI. MR2435520 https://doi.org/10.1090/gsm/096

[23] KRYLOV, N. V. (2009). *Controlled Diffusion Processes*. *Stochastic Modelling and Applied Probability* **14**. Springer, Berlin. Translated from the 1977 Russian original by A. B. Aries, Reprint of the 1980 edition. MR2723141

[24] KRYLOV, N. V. and RÖCKNER, M. (2005). Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields* **131** 154–196. MR2117951 https://doi.org/10.1007/s00440-004-0361-z

[25] KURTZ, T. G. and PROTTER, P. (1991). Wong–Zakai corrections, random evolutions, and simulation schemes for SDEs. In *Stochastic Analysis* 331–346. Academic Press, Boston, MA. MR1119837

[26] LÊ, K. (2020). A stochastic sewing lemma and applications. *Electron*. *J. Probab.* **25** Paper No. 38. MR4089788 https://doi.org/10.1214/20-ejp442

[27] LÊ, K. and LING, C. (2021). Taming singular stochastic differential equations: A numerical method. ArXiv preprint. Available at arXiv:2110.01343.

[28] LEOBACHER, G. and SZÖLGYENYI, M. (2017). A strong order 1/2 method for multidimensional SDEs with discontinuous drift. *Ann*. *Appl*. *Probab.* **27** 2383–2418. MR3693529 https://doi.org/10.1214/16-AAP1262

[29] LEOBACHER, G. and SZÖLGYENYI, M. (2018). Convergence of the Euler–Maruyama method for multi-dimensional SDEs with discontinuous drift and degenerate diffusion coefficient. *Numer. Math*. **138** 219–239. MR3745015 https://doi.org/10.1007/s00211-017-0903-9

[30] MENOUKEU PAMEN, O. and TAGUCHI, D. (2017). Strong rate of convergence for the Euler–Maruyama approximation of SDEs with Hölder continuous drift coefficient. *Stochastic Process*. *Appl*. **127** 2542–2559. MR3660882 https://doi.org/10.1016/j.spa.2016.11.008

[31] MIKULEVIČIUS, R. and XU, F. (2018). On the rate of convergence of strong Euler approximation for SDEs driven by Levy processes. *Stochastics* **90** 569–604. MR3784978 https://doi.org/10.1080/17442508.2017.1381095

[32] MÜLLER-GRONBACH, T. and YAROSLAVTSEVA, L. (2020). On the performance of the Euler–Maruyama scheme for SDEs with discontinuous drift coefficient. *Ann*. *Inst*. *Henri Poincaré Probab*. *Stat*. **56** 1162–1178. MR4076779 https://doi.org/10.1214/19-AIHP997

[33] MÜLLER-GRONBACH, T. and YAROSLAVTSEVA, L. (2022). Sharp lower error bounds for strong approximation of SDEs with discontinuous drift coefficient by coupling of noise. *Ann*. *Appl*. *Probab*. To appear.

[34] NEUENKIRCH, A. and SZÖLGYENYI, M. (2021). The Euler–Maruyama scheme for SDEs with irregular drift: Convergence rates via reduction to a quadrature problem. *IMA J. Numer. Anal*. **41** 1164–1196. MR4246871 https://doi.org/10.1093/imanum/draa007

[35] NGO, H.-L. and OGAWA, S. (2011). On the discrete approximation of occupation time of diffusion processes. *Electron*. *J. Stat*. **5** 1374–1393. MR2842909 https://doi.org/10.1214/11-EJS645

[36] NGO, H.-L. and TAGUCHI, D. (2016). Strong rate of convergence for the Euler–Maruyama approximation of stochastic differential equations with irregular coefficients. *Math*. *Comp*. **85** 1793–1819. MR3471108 https://doi.org/10.1090/mcom3042

[37] PAGÈS, G. (2018). *Numerical Probability*: *An introduction with applications to finance*. *Universitext*. Springer, Cham. MR3823785 https://doi.org/10.1007/978-3-319-90276-0

[38] REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Springer, Berlin. MR1725357 https://doi.org/10.1007/978-3-662-06400-9

[39] SHARDIN, A. A. and SZÖLGYENYI, M. (2016). Optimal control of an energy storage facility under a changing economic environment and partial information. *Int*. *J. Theor*. *Appl*. *Finance* **19** 1650026. MR3505500 https://doi.org/10.1142/S0219024916500266

[40] SICKEL, W. (1999). Pointwise multipliers of Lizorkin–Triebel spaces. In *The Maz'Ya Anniversary Collection*, *Vol*. 2 (*Rostock*, 1998). *Oper. Theory Adv. Appl*. **110** 295–321. Birkhäuser, Basel. MR1747900

[41] VERETENNIKOV, A. J. (1980). Strong solutions and explicit formulas for solutions of stochastic integral equations. *Mat*. *Sb*. (*N.S.*) **111** 434–452, 480. MR0568986

[42] ZVONKIN, A. K. (1974). A transformation of the phase space of a diffusion process that will remove the drift. *Mat*. *Sb*. (*N.S.*) **93** 129–149, 152. MR0336813