# STRONG SOLUTIONS OF SDES WITH CRITICAL DISCONTINUITIES IN DIFFUSION COEFFICIENTS

S. E. BOUTIAH AND D. KINZEBULATOV

ABSTRACT. We prove strong existence for Itô SDEs with diffusion coefficients that can introduce strong attraction to a submanifold. We extend and strengthen the Röckner-Zhao approach, which uses Malliavin calculus to establish compactness of the approximating solutions in Wiener-Sobolev space. At least when the diffusion coefficients are sufficiently regular in time, this provides an alternative to Krylov's recent proof of strong existence via analysis of the Itô-Duhamel series.

## 1. INTRODUCTION AND MAIN RESULT

1.1. **Introduction.** This paper, which continues [KM1], concerns strong existence for the Itô SDE

$$dX_t = \sigma(t, X_t)\, dW_t, \quad X_0 = x \in \mathbb{R}^d, \tag{1.1}$$

where $\{W_s\}_{s\geq 0}$ is a $d$-dimensional Brownian motion ($d \geq 3$) with diffusion coefficients that are non-degenerate, i.e. $\sigma$ is bounded on $\mathbb{R}^{1+d}$ and there exists $\kappa > 0$ such that, for every $t \in \mathbb{R}$,

$$\sigma(t, \cdot)\sigma(t, \cdot)^\top \geq \kappa I \quad \text{a.e. on } \mathbb{R}^d, \tag{1.2}$$

but can have critical discontinuities. By a critical discontinuity of $\sigma$ we mean a discontinuity that can "trap" the solution, for example,

$$\sigma(x) = \frac{1}{\sqrt{1+c}}\left[I + \left(\sqrt{1+c} - 1\right)\frac{x \otimes x}{|x|^2}\right], \quad 0 \neq x \in \mathbb{R}^d, \quad c > -1. \tag{1.3}$$

Here $I$ is the $d \times d$ identity matrix, $\sigma(0) = 0$. This is a well known-example of a diffusion coefficient that can make $X_t$ arrive at the origin and stay there a.s. if $c > d-2$, see [DB] and [B, V.3, Sect.3]. This attraction phenomenon can be seen at the level of the two-sided estimates on the transition density $p(t, x, y)$ of (1.1) with $\sigma$ given by (1.3), which are no longer purely Gaussian even for $c < d-2$ but involve an extra polynomial factor $|y|^{-\theta} \vee 1$ with power $\theta = \theta(d, c)$ that quantifies the strength of the attraction to the origin introduced by $\sigma$ [MNS]. Such attraction mechanisms occur in some physical models, for instance, in some models of turbulent transport the multiplicative diffusion may cause clustering, collision, coalescence, or sticky behaviour, see [GC, GV].

See (1.6) for the definition of our class of $\sigma$ that includes (1.3).

The study of related problem of weak and strong well-posedness for SDEs with *additive noise and singular drift* has undergone dramatic progress in the past decade; more on this below. However, less is known about critical discontinuities of the diffusion coefficients. Compared to singular drifts, the former present a different set of difficulties. In particular, while the problem of weak existence for (1.1)

---

2020 *Mathematics Subject Classification.* 60H10 (primary), 60H07, 60J60, 35R05 (secondary).

*Key words and phrases.* Stochastic differential equations, strong solutions, Malliavin calculus, discontinuous diffusion coefficients, singular drifts.

The research of D.K. is supported by Discovery grant RGPIN-2024-04236 of the Natural Sciences and Engineering Research Council of Canada.

is straightforward and is settled using a simple tightness estimate obtained from the BDG inequalities, the problem of strong existence is a different matter, as illustrated by the classical example of Krylov-Zvonkin (Example 1). Recently, Krylov introduced an approach to constructing a conditionally unique strong solution of Itô SDE (1.1) – that can also contain a singular drift as in [KM1] – as the sum of the Itô-Duhamel series. Briefly, he defines $X_t$ via the *explicit formula*

$$f(X_t) = P^{0,t}f(x) + \sum_{n=1}^{\infty} \int_{\Delta_n(t)} P^{0,t_n} Q_{k_n}^{t_{n-1},t_n} \cdots Q_{k_1}^{0,t_1} f(x)\, dW_{t_1}^{k_1} \cdots dW_{t_n}^{k_n},$$

obtained by iterating Itô's formula

$$f(X_t) = P^{0,t}f(x) + \int_0^t Q_k^{r,t} f(X_r)\, dW_r,$$

where $Q_k^{r,t}f(x) := \sigma^{ik}(t,x)\nabla_i P^{r,t}f(x)$ and $P^{r,t}$ is the propagator for SDE (1.1). Krylov's proof that this formal series actually converges to a strong solution to SDE (1.1) uses his much earlier result with Veretennikov [KrV] and his new functional-analytic arguments, see [Kr1] and [Kr2, Kr3] regarding the time-homogeneous case $\sigma = \sigma(x)$ and the time-inhomogeneous case $\sigma = \sigma(t,x)$, respectively.

The principal objective of the present work is to provide an alternative to Krylov's proof of strong existence for (1.1) based on the approach to constructing strong solutions via compactness on the Wiener-Sobolev space due to [RZ] and its strengthening in [KM1]. Namely, in [RZ], Michael Röckner and Guohuan Zhao considered the SDE with additive noise

$$X_t = x + \int_r^t b(s, X_s)\, ds + W_t - W_r, \quad t \in [r,T], \tag{1.4}$$

with drift $b$ in the Ladyzhenskaya-Prodi-Serrin class

$$|b| \in L^q(\mathbb{R}, L^p(\mathbb{R}^d)), \quad \frac{d}{p} + \frac{2}{q} \leq 1, \quad d \leq p \leq \infty, \;\; 2 \leq q \leq \infty. \tag{1.5}$$

Building on earlier works of Bailly-Saussereau [BS], Meyer-Brandis–Proske [MP], Mohammed-Nilsen-Proske [MNP] and Rezakhanlou [R], they constructed a strong solution to (1.4), (1.5) by showing that the sequence of strong solutions of the SDEs with regularized drifts is compact in the Wiener-Sobolev space. They furthermore showed that thus constructed strong solution is conditionally pathwise unique – here and below, conditional on a reasonable occupation estimate – using Cherny's "dual to the Yamada-Watanabe principle" [C], exploiting the fact that (1.4) has constant diffusion coefficients. Shortly after, [KM1] strengthened the PDE part of [RZ], which allowed the authors to establish strong existence and conditional pathwise uniqueness for (1.4) for form-bounded drifts that, unlike (1.5), also cover critical-order singularities that can introduce e.g. strong attraction phenomena (see Definition 2 and examples thereafter).

The present paper thus continues [RZ, KM1] and replaces the functional-analytic arguments of Krylov by algebraic calculations involving Malliavin derivatives of solutions of the approximating SDEs, in order to keep track of how strong solutions $X_t^{x,n}$ of the approximating SDEs

$$dX_t^{n,x} = \sigma_n(t, X_t^{n,x})\, dW_t, \quad X_0^{n,x} = x \in \mathbb{R}^d,$$

with smooth $\sigma_n$, driven the same Browninan motion, depend on Brownian paths. The heavy lifting is still done by PDE estimates (Propositions 2 and 3), but only as a priori estimates, i.e. the PDEs have smooth coefficients, but the constants in the estimates do not depend on the smoothness. Up to the use of the Malliavin calculus, the proofs below arguably use only elementary probabilistic arguments.

– At the PDE level we strengthen [RZ]:

(a) Proposition 3, the crucial simplex estimate in the approach of Röckner-Zhao needed to verify compactness on the Wiener-Sobolev space, is proved along the lines of [KM1] – that was the authors' main observation in that paper – rather than [RZ].

(b) Proposition 2 provides gradient bounds for solutions of the backward Kolmogorov equation. These bounds are nontrivial because of the possible blow-up phenomena and are proved using the approach of [KS, K1].
We also use these gradient bounds to run Moser iterations in order to construct the Feller propagator.

– At the probabilistic level we extend [RZ]:

(c) In the case of multiplicative noise, the calculations involving the Malliavin derivatives of solutions of the approximating SDEs are more involved; it is at this step that we use the additional time-regularity assumption (**H**). Let us also add that iterating the estimate of Proposition 3 in order to prove Proposition 4 (i.e. the actual verification of compactness on the Wiener-Sobolev space), as was done in [RZ] and [KM1] for singular drifts, is rather difficult since now one needs to iterate Itô's integrals, and simply using the BDG inequality does not allow to control the convergence of the geometric series (in the one-dimensional case there is, however, a result of Carlen-Krée [CK] that provides the sought control on the iterated Itô integrals). Instead, to obtain Proposition 4 from Proposition 3, we appeal to a stochastic Gronwall-type estimate.

Our main interest in this work is the spatial singularities of the diffusion coefficients $\sigma$; although our $\sigma$ can depend on time $t$, Krylov in [Kr2] can handle stronger singularities in $t$ (naturally, at the moments when the drift is more regular in space). In the time-homogeneous case, our class of diffusion coefficients is somewhat larger than the class considered by Krylov – the difference is e.g. the Chang-Wilson-Wolff class, see examples below – at least comparing it with how Krylov's result and proof are written down. Let us also add that Krylov also proves other results related to SDE (1.1), such as conditional pathwise uniqueness, that we do not attempt to discuss in the present paper.

The class of time-homogeneous matrix fields $a = \sigma\sigma^\top$, where $\sigma = \sigma(x)$ is in our class (1.6), was treated earlier in [KS5] in the context of the weak existence/unique Feller semigroup theory of the SDE

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t, \quad X_0 = x \in \mathbb{R}^d$$

having additionally form-bounded drift $b \in \mathbf{F}_\delta$ (Definition 2). The present paper is of course quite different, but it uses gradient bounds (Proposition 2) and employs the construction of the Feller semigroup via Moser-type iterations (Theorem 1($ii$)) similar to those in [KS5].

Commenting further on the weak solutions (for simplicity, still for time-homogeneous $a$), we mention the detailed weak solution theory of

$$dX_t = \sqrt{a(X_t)}dW_t, \quad X_0 = x, \quad d \geq 2,$$

due to Krylov, with bounded, symmetric $a \geq \kappa I$ in the VMO class, that is, satisfying

$$\omega_a(r) \to 0 \quad \text{as } r \downarrow 0,$$

where

$$\omega_a(r) := \sup_{x\in\mathbb{R}^d, 0<\rho\le r} \frac{1}{|B_\rho|}\int_{B_\rho(x)} |a(y)-a_{B_\rho(x)}|dy, \quad a_B := \frac{1}{|B|}\int_B a.$$

The VMO is a broad condition that, unlike the class (1.6), does not impose any conditions on the derivatives of $a$. Although, on the other hand, there are matrices that satisfy (1.6) but are not in VMO, see examples before Theorem 1. Under these condition on $a$, Krylov proved, among many other results, the uniqueness in law. Recently, he also included a Morrey class singular drift in this SDE, see details in the end of Section 2.

In what follows, $\sigma_j : \mathbb{R}^{1+d} \to \mathbb{R}^d$ denotes the $j$-th column of the diffusion matrix $\sigma$.

1.2. **Main result.** In Theorem 1 we consider the class of diffusion matrices $\sigma$ having form-bounded derivatives:

$$|\nabla\sigma| := \sqrt{\sum_{i,j=1}^d |\nabla_i \sigma_j|^2} \in \mathbf{F}_\nu, \tag{1.6}$$

i.e. $|\nabla\sigma| \in L^2_{\rm loc}(\mathbb{R}^{1+d})$ and for a.e. $t\in\mathbb{R}$

$$\||\nabla\sigma(t,\cdot)|\varphi\|_2^2 \le \nu\|\nabla\varphi\|_2^2 + c_\nu\|\varphi\|_2^2 \quad \forall\,\varphi\in C_c^\infty(\mathbb{R}^d)$$

for some constant $c_\nu$ – the value of this constant is not so important for us in this paper, unlike the value of the constant $\nu$ ("form-bound of $|\nabla\sigma|$"), which measures the magnitude of the singularities of $|\nabla\sigma|$.

**Examples.** 1. We have

$$|\nabla\sigma| \in L^d(\mathbb{R}^d) \quad \Rightarrow \quad |\nabla\sigma|\in\mathbf{F}_\nu$$

with $\nu$ that can be chosen arbitrarily small at the expense of increasing $c_\nu$, as can be easily seen by applying Hölder inequality the Sobolev embedding theorem; we postpone the details until the examples in the next section.

2. More generally, if $|\nabla\sigma| \in L^{d,\infty}(\mathbb{R}^d)$ (weak $L^d$ space), then $|\nabla\sigma|\in\mathbf{F}_\nu$ with $\nu \approx \|\nabla\sigma\|_{L^{d,\infty}(\mathbb{R}^d)}$.

3. A still broader example is provided by matrices $\sigma$ with finite Morrey norm $\|\nabla\sigma\|_{M_{2+\varepsilon}}$ or finite Chang-Wilson-Wolff norm, see (2.4), (2.5).

4. We now switch to examples of concrete diffusion matrices. Consider the attracting diffusion matrix from the introduction:

$$\sigma(x) = \frac{1}{\sqrt{1+c}}\left[I + \left(\sqrt{1+c}-1\right)\frac{x\otimes x}{|x|^2}\right], \quad x\in\mathbb{R}^d.$$

A simple calculation using Hardy's inequality

$$\langle |x|^{-2},\varphi^2\rangle \le \frac{4}{(d-2)^2}\langle|\nabla\varphi|^2\rangle, \quad \varphi\in C_c^\infty(\mathbb{R}^d),$$

shows that

$$|\nabla\sigma|\in\mathbf{F}_\nu \quad \text{ with } \nu = 8\frac{(d-1)}{(d-2)^2}\left(\frac{\sqrt{1+c}-1}{\sqrt{1+c}}\right)^2.$$

5. An example of a rapidly oscillating diffusion matrix from [KS5]:

$$\sigma(x) = \mathrm{diag}(1 + c\sin(\alpha\log|x|), 1, \dots, 1)$$

for fixed $0<c<1$. Indeed, the uniform ellipticity condition $\sigma\sigma^\top \ge (1-c)^2 I$ is satisfied, and

$$|\nabla\sigma|^2 = \frac{c^2\alpha^2\cos^2(\alpha\log|x|)}{|x|^2} \le \frac{c^2\alpha^2}{|x|^2},$$

so the Hardy inequality implies the form-boundedness of $|\nabla\sigma|^2$. This example shows that the smallness of the form-bound (take $\alpha$ small) does not imply that the diffusion matrix is close to a constant matrix, as long as one keeps the amplitude of the oscillations $c$ close to 1. Let us also note that this matrix $\sigma$ is not in the VMO class of diffusion coefficients. The class has received a considerable attention in the literature on SDEs and Kolmogorov equations, more on this in the end of this section.

6. Rotating anisotropy. Define the rotation the matrix in the first two coordinates:

$$R_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \oplus I_{d-2},$$

where $I_{d-2}$ is the identity matrix of dimension $d-2$. Fix a constant diagonal matrix $D := \operatorname{diag}(\lambda_1, \dots, \lambda_d)$, $\lambda_i > 0$, and define

$$\sigma(x) := R_{\varepsilon \log|x|} D.$$

It follows that $\sigma\sigma^\top \geq (\min_i \lambda_i^2) I$ and

$$|\nabla\sigma|^2 = \frac{\varepsilon^2(\lambda_1^2 + \lambda_2^2)}{|x|^2},$$

i.e. by Hardy's inequality, $|\nabla\sigma|$ is form-bounded.

We could also write $\mathbb{R}^d = \mathbb{R}^m \oplus \mathbb{R}^{d-m}$, $m \geq 3$, and make the logarithm dependent on the first $m$ coordinates of $x$. In this case, to show the form-boundedness of $|\nabla\sigma|$, we can refer to the Hardy inequality in the first $m$ variables. Alternatively, one can verify the Morrey class condition (example 3), albeit at the expense of having much less explicit expression for the form-bound.

See also examples after Theorem 2.

More generally, we can consider a series of matrices from the previous examples – properly normalized to ensure the convergence of the series to a uniformly elliptic matrix – with their points of discontinuity constituting a dense subset of $\mathbb{R}^d$.

We now state the main result of the paper. Set

$$\rho(x) \equiv \rho_\theta(x) := (1 + \theta|x|^2)^{-\frac{d}{2}-1}, \quad x \in \mathbb{R}^d,$$

where $\theta > 0$. We have

$$|\nabla\rho| \leq C\sqrt{\theta}\rho, \qquad \frac{|\nabla\rho|^2}{\rho} \leq C\theta\rho. \tag{1.7}$$

We will be applying (1.7) to $\rho$ with $\theta$ chosen sufficiently small, in order to control the occurrences of the derivatives of $\rho$.

Let $L^p_\rho(\mathbb{R}^d)$ denote the $L^p$ space with respect to measure $\rho(x)dx$. Denote by $\rho_y$, $y \in \mathbb{R}^d$, the translation $\rho_y(x) := \rho(x-y)$.

Let $x \in \mathbb{R}^d$. We consider SDE

$$X^x_{r,t} = x + \int_r^t \sigma(s, X^x_{r,s})\, dW_s, \quad r \leq t \leq T \tag{1.8}$$

on a fixed complete probability space $(\Omega, \{\mathcal{F}^W_t\}_{0\leq t\leq T}, \mathbf{P})$ that supports a Brownian motion $\{W_t\}_{t\geq 0}$, where $\{\mathcal{F}^W_t\}_{0\leq t\leq T}$ is the filtration of $W$ on $[0,T]$ after completion.

**Definition 1.** *A family $\{X^x_{r,t} \mid 0 \leq r \leq t \leq T\}$ is a family of strong solutions if, for every $r \in [0,T[$ and $x \in \mathbb{R}^d$, the process $t \mapsto X^x_{r,t}$, $r \in [r,T]$, is continuous and adapted to $\mathcal{F}^W_{r,t} := \sigma(W_s - W_r \mid r \leq s \leq t)$ (after completion), and*

$$\mathbf{P}\bigg(X^x_{r,t} = x + \int_r^t \sigma(s, X^x_{r,s})\, dW_s \text{ for every } t \in [r,T]\bigg) = 1$$

**Theorem 1.** *Let $d \geq 3$. Assume that the diffusion coefficients $\sigma \in [L^\infty(\mathbb{R}^{1+d})]^{d\times d}$ satisfy the uniform ellipticity condition: for every $t \in \mathbb{R}$*

$$\sigma(t,\cdot)\sigma(t,\cdot)^\top \geq \kappa I \quad \text{a.e. on } \mathbb{R}^d \text{ for some fixed } \kappa > 0$$

*and have form-bounded derivatives, i.e. condition (1.6) holds.*

*We also impose the following auxiliary condition on the time-dependence of $\sigma$:*

($\mathbf{H}$)

$$\sup_{y\in\mathbb{R}^d} \|\sigma(t,\cdot) - \sigma(s,\cdot)\|_{L^2_{\rho_y}(\mathbb{R}^d)} \leq C|t-s|^\gamma \quad t,s \in \mathbb{R}, |t-s| \leq 1$$

*with the Hölder continuity exponent $\gamma \in ]0,1]$.*

*Let*

$$\sigma_n \in \big[L^\infty(\mathbb{R}, C^\infty_b(\mathbb{R}^d))\big]^{d\times d} \cap [C^\gamma_{unif,\mathrm{loc}}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^{d\times d}$$

*be a sequence of bounded matrix fields having bounded derivatives in the spatial variables,*

$$\|\sigma_n\|_{L^\infty(\mathbb{R}^{1+d})} \leq \|\sigma\|_{L^\infty(\mathbb{R}^{1+d})}, \quad \sigma_n\sigma_n^\top \geq \frac{\kappa}{4} I$$

*such that*

$$\sigma_n \to \sigma \quad \text{in } [L^\infty_{\mathrm{loc}}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^{d\times d},$$

*and that satisfy (1.6) and ($\mathbf{H}$) uniformly in $n$, i.e. with the same parameters $\nu$, $c_\nu$ and $C$, $\gamma$. It is not difficult to see that the spatial mollification of $\sigma$ produces such $\{\sigma_n\}$, see Appendix B. By the classical theory, for each $n$, there exists a pathwise unique strong solution $X^{x,n}_{r,t}$ to the approximating equation*

$$X^{x,n}_{r,t} = x + \int_r^t \sigma_n(s, X^{x,n}_{r,s})\, dW_s, \tag{1.9}$$

*considered on the same (fixed) complete probability space $(\Omega, \{\mathcal{F}^W_t\}_{0\leq t\leq T}, \mathcal{F}^W, \mathbf{P})$.*

*Then, if the form-bound $\nu$ of $|\nabla\sigma|$ in (1.6) is smaller than a certain constant that only depends on the dimension $d$, the ellipticity constant $\kappa$ and $\|\sigma\|_{L^\infty(\mathbb{R}^{1+d})}$, the following are true:*

*(i) There exists a subsequence $\{\sigma_{n_k}\}$ and a continuous random field $X$ such that, for every $N \geq 1$,*

$$\sup_{0\leq r\leq t\leq T, |x|\leq N} |X^{x,n_k}_{r,t} - X^x_{r,t}| \to 0 \quad \mathbf{P}-a.s.,$$

*where $X^x_{r,t}$ is a strong solution of SDE (1.8) on $(\Omega, \{\mathcal{F}^W_t\}_{0\leq t\leq T}, \mathcal{F}^W, \mathbf{P})$.*

*(ii) The operators*

$$P^{r,t}_\sigma f(x) := \mathbf{E}[f(X^x_{r,t})], \quad f \in C_\infty, \quad r,t \in [0,T], r < t,$$

*constitute a backward Feller propagator on $C_\infty$, and, furthermore, if $\{\sigma_n\}$ additionally satisfies*

$$\nabla a_n \to \nabla a \quad \text{in } L^2_{\mathrm{loc}}(\mathbb{R}, L^2_\rho(\mathbb{R}^d)),$$

*where $a = \sigma\sigma^\top$, $a_n = \sigma_n(\sigma_n)^\top$ (once again, mollifying $\sigma$ in the spatial variables produces this convergence, see Appendix B), then*

$$P^{r,t}_{\sigma} = s\text{-}C_\infty\text{-}\lim_n P^{r,t}_{\sigma_n} \quad \textit{loc. uniformly in } r,t \in [0,T],\ r<t.$$

*There is no need to pass to a subsequence of $\{\sigma_n\}$. In this sense, the law of $X^x_{r,t}$ does not depend on the choice of the subsequence in $(i)$.*

**Example 1.** The celebrated Krylov-Zvonkin's counterexample to strong existence shows that the form-bound $\nu$ of $|\nabla\sigma|$ in (1.6) cannot be greater than 12 in dimension $d=4$; this demonstrates that the smallness of $\nu$ assumed in Theorem 1 is necessary for strong existence (or rather than $\nu$ cannot be too large). Namely, in $\mathbb{R}^4$, consider SDE

$$dX_t = \sigma(X_t)dW_t, \quad X_0 = 0, \tag{1.10}$$

with diffusion matrix

$$\sigma(x) = \frac{1}{|x|}\begin{pmatrix} x_1 & -x_2 & -x_3 & -x_4 \\ x_2 & x_1 & -x_4 & x_3 \\ x_3 & x_4 & x_1 & -x_2 \\ x_4 & -x_3 & x_2 & x_1 \end{pmatrix},$$

define $\sigma(0)$ arbitrarily. Then $\sigma\sigma^\top = I$, i.e. in this case the ellipticity constant $\kappa = 1$. On the other hand, one shows, using Lévy's characterization of Brownian motion, that

$$B_t := \int_0^t \sigma^\top(W_s)dW_s$$

is a Brownian motion. We have $W_t = \int_0^t \sigma(W_s)dB_s$, hence $t \mapsto (B_t, W_t)$ is a weak solution to SDE (1.10). At the same time, $(-B_t, W_t)$ is also a weak solution to the same SDE. Therefore, pathwise uniqueness fails. One can prove that every solution to this SDE is Brownian, so there is uniqueness in law. If we had strong existence, then the uniqueness in law would imply pathwise uniqueness, a contradiction. See details in [KrZ].

We can evaluate the form-bound $\nu$ of $|\nabla\sigma|$. We have

$$|\nabla\sigma|^2 = \sum_{i,j,k=1}^4 |\nabla_k \sigma_{ij}|^2 = \frac{12}{|x|^2}$$

since each entry of the form $\frac{x_l}{|x|}$ in $\sigma$ occurs exactly four times, and $\sum_{l=1}^4 |\nabla \frac{x_l}{|x|}|^2 = \frac{3}{|x|^2}$. Now, invoking the Hardy inequality on $\mathbb{R}^4$ with sharp constant, i.e. $\||x|^{-1}\varphi\|_2^2 \le \|\nabla\varphi\|_2^2$, $\varphi \in C_c^\infty(\mathbb{R}^d)$, we obtain that the form-bound $\nu$ of $|\nabla\sigma|$ is exactly 12.

1.3. **About the proof of Theorem 1 and its possible extension.** 1. In the next Theorem 2 we consider the SDE with additive noise and singular (form-bounded) drift. The task of combining such drift and the diffusion coefficients as in Theorem 1 now seems to be achievable with some additional work, given that the principal estimates – Proposition 3 and [KM1, Prop. 1] – are proved within a compatible framework. We plan to write down the details elsewhere.

2. The proof of Theorem 1(i) follows the compactness approach of Röckner and Zhao. Proposition 3, together with the stochastic Gronwall-type estimate, yields Proposition 4, which provides uniform estimates for the spatial and Malliavin derivatives of the approximating solutions. These estimates

verify the hypotheses of the Wiener-Sobolev compactness criterion in Lemma 3.1. A diagonal argument then gives, after passing to a subsequence, convergence on a fixed countable dense subset of $\Delta_2(T) \times \mathbb{R}^d$.

The increment estimates derived from Proposition 4, together with the Kolmogorov-Chentsov theorem, yield continuity of the limiting random field and stochastic equicontinuity of the approximating random fields. A finite-net argument and the Borel-Cantelli lemma then upgrades the convergence on the dense set to a.s. local uniform convergence on $\Delta_2(T) \times \mathbb{R}^d$.

It remains to pass to the limit in the approximating SDEs. Proposition 2 yields the occupation estimate used in Lemma 5.1. The construction of the limiting occupation measure then allows us to prove convergence of the stochastic integrals and to show that the limit found at the previous step is a strong solution.

At this last step, we could instead appeal to Krylov's classical occupation-time estimate, which applies to bounded uniformly nondegenerate diffusion coefficients, i.e. without requiring the form-boundedness condition on their derivatives. Since our eventual goal is to add a singular drift to SDE (1.1), we use the longer argument based on Proposition 2 which is more amenable for extensions in our context.

The proof of Theorem 1(*ii*) uses PDE arguments only. The gradient bounds of Proposition 4 allow us to run Moser iterations for the differences of the approximating Kolmogorov equations. This yields strong convergence in $C_\infty$ of the corresponding Feller propagators, locally uniformly in the time parameters, and gives the backward Feller propagator.

**Acknowledgements.** We are deeply grateful to Kodjo Raphaël Madou for valuable discussions and for his participation in the early stages of this work.

## 2. Remarks on additive noise and singular drift

Let $d \geq 3$. Consider SDE

$$dX_t = b(t, X_t)dt + dW_t, \quad X_0 = x \in \mathbb{R}^d, \tag{2.1}$$

having form-bounded drift $b$:

**Definition 2.** *A Borel measurable vector field $b : \mathbb{R}^{1+d} \to \mathbb{R}^d$ with $|b| \in L^2_{\rm loc}(\mathbb{R}^{1+d})$ is said to be form-bounded if there exist constant $\delta > 0$ such that the following quadratic form inequality holds for a.e. $t \in \mathbb{R}$:*

$$\|b(t,\cdot)\varphi\|_2^2 \leq \delta \|\nabla \varphi\|_2^2 + g_\delta(t)\|\varphi\|_2^2 \qquad \forall\, \varphi \in C_c^\infty(\mathbb{R}^d) \tag{2.2}$$

*for some function $g_\delta \in L^1_{\rm loc}(\mathbb{R})$.*

*This is written as $b \in \mathbf{F}_{\delta,g}$.*

The function $g_\delta$ is required to handle drifts that can be locally unbounded in time, cf. examples below.

**Examples.** We mention the following examples/sub-classes of form-bounded drifts defined in elementary terms; there are some repetitions with the examples before Theorem 1.

1. The "spatial endpoint" of the Ladyzhenskaya-Prodi-Serrin class (1.5):

$$b \in L^\infty([0,\infty[, L^d(\mathbb{R}^d) + L^\infty(\mathbb{R}^d)) \quad \Rightarrow \quad b \in \mathbf{F}_{\delta,g_\delta}$$

for appropriate $\delta$ and bounded $g_\delta$; let us omit $\mathbb{R}^d$ in $L^p(\mathbb{R}^d)$ from now on. Indeed, writing $b = b_1 + b_2$, where $b_1 \in L^\infty(\mathbb{R}, L^d)$, $b_2 \in L^\infty(\mathbb{R}, L^\infty)$, we have for a.e. $t \in \mathbb{R}$,

$$\|b(t,\cdot)\varphi\|_2^2 \leq (1+\varepsilon)\|b_1(t,\cdot)\|_{L^d}^2 \|\varphi\|^2_{L^{\frac{2d}{d-2}}} + (1+\varepsilon^{-1})\|b_2(t,\cdot)\|^2_{L^\infty}\|\varphi\|^2_{L^2} \qquad (\varepsilon > 0)$$

(apply the Sobolev embedding)

$$\leq C_S(1+\varepsilon)\|b_1(t,\cdot)\|^2_{L^d}\|\nabla\varphi\|^2_{L^2} + (1+\varepsilon^{-1})\|b_2(t,\cdot)\|^2_{L^\infty}\|\varphi\|^2_{L^2},$$

so $b \in \mathbf{F}_{\delta, g_\delta}$ with $\delta := \sup_{t\in\mathbb{R}} C_S(1+\varepsilon)\|b_1(t,\cdot)\|^2_{L^d}$ and $g_\delta(t) = (1+\varepsilon^{-1})\|b_2(t,\cdot)\|^2_{L^\infty}$.

Let us note that if $b$ is time-homogeneous, then, by considering cutoff of $b$, we can find, for every $\varepsilon > 0$, vector fields $b_1$ and $b_2$ as above such that $\|b_1\|_{L^d} < \varepsilon$, and so the form-bound $\delta$ of $b$ is, in fact, arbitrarily small; this comes at the expense of having $\|b_2\|_{L^\infty}$ large.

2. Recall that a Borel measurable function $h : \mathbb{R}^d \to \mathbb{R}$ is in $L^{d,\infty}$ if $\|h\|_{L^{d,\infty}} := \sup_{s>0} s|\{x \in \mathbb{R}^d : |h(x)| > s\}|^{1/d} < \infty$. By the Strichartz inequality with sharp constants [KPS, Prop. 2.5, 2.6, Cor. 2.9], if, say, $b = b(x)$ with $|b| \in L^{d,\infty}$, then

$$\begin{aligned} b \in \mathbf{F}_{\delta_1,0} \quad \text{ with } \sqrt{\delta_1} &= \||b|(\lambda-\Delta)^{-\frac12}\|_{2\to2} \\ &\leq \|b\|_{d,\infty}\Omega_d^{-\frac1d}\||x|^{-1}(\lambda-\Delta)^{-\frac12}\|_{2\to2} \\ &\leq \|b\|_{d,\infty}\Omega_d^{-\frac1d}2^{-1}\frac{\Gamma(\frac{d-2}{4})}{\Gamma(\frac{d+2}{4})} = \|b\|_{d,\infty}\Omega_d^{-\frac1d}\frac{2}{d-2}, \end{aligned}$$

where $\Omega_d = \pi^{\frac d2}\Gamma(\frac d2 + 1)$ is the volume of the unit ball in $\mathbb{R}^d$.

3. Extending the previous example, we obtain that

$$|b| \in L^\infty(\mathbb{R}, L^{d,\infty}(\mathbb{R}^d)) \quad \Rightarrow \quad b \in \mathbf{F}_{\delta,0},$$

with $\delta \approx \sup_{t\in\mathbb{R}} \|b(t,\cdot)\|^2_{L^{d,\infty}(\mathbb{R}^d)}$. This class contains the class $|b| \in L^\infty(\mathbb{R}, L^d(\mathbb{R}^d))$ from the first example. The set-theoretic difference between the latter and $L^\infty(\mathbb{R}, L^{d,\infty}(\mathbb{R}^d))$ contains e.g. the attracting drift

$$b(x) := -\frac{d-2}{2}\sqrt{\delta}\frac{x}{|x|^2}, \tag{2.3}$$

that, by the Hardy inequality or by the calculation in 2, is in $\mathbf{F}_{\delta,0}$. This drift provides a counterexample to the weak existence if the “strength of attraction to the origin” $\delta$ is too large, see Section 2.1 for details.

The counterexample to weak existence discussed in that section already shows that there is an important distinction between the class $|b| \in L^\infty(\mathbb{R}, L^d(\mathbb{R}^d))$ and the class of form-bounded drifts $b \in \mathbf{F}_\delta$. Namely, the former does not cover such strong attraction/blow-up phenomena, despite being critical under the diffusive scaling.

4. The weighted Hardy inequality of Hoffmann-Ostenhof–Laptev [HL]. Fix

$$0 \leq \Phi \in L^s(S^{d-1}) \quad \text{ for some } s \geq \frac{2(d-2)^2}{2(d-1)} + 1,$$

where $S^{d-1}$ is the unit sphere in $\mathbb{R}^d$. If

$$|b(x)|^2 \leq \delta\frac{(d-2)^2}{4}c\frac{\Phi(x/|x|)}{|x|^2}, \qquad \text{where } c := \frac{|S^{d-1}|^{\frac1s}}{\|\Phi\|_{L^s(S^{d-1})}},$$

then $b \in \mathbf{F}_\delta$ with $g_\delta = 0$.

5. Let $K \in \mathbf{F}_\kappa(\mathbb{R}^d)$. Then the many-particle drift $b=(b_1,\dots,b_N):\mathbb{R}^{dN} \to \mathbb{R}^{dN}$ with components defined by

$$b_i(x^1,\dots,x^N) := \frac{1}{N}\sum_{j=1, j\neq i}^N K(x^i-x^j), \quad x^i, x^j \in \mathbb{R}^d$$

is in $\mathbf{F}_\delta(\mathbb{R}^{dN})$ with $\delta = \frac{(N-1)^2}{N^2}\kappa$.

6. Fix a small $\varepsilon>0$. If $\sup_{t\in\mathbb{R}}\|b(t,\cdot)\|_{M_{2+\varepsilon}}<\infty$, then $b \in \mathbf{F}_\delta$ with $\delta$ proportional to the supremum, and $g_\delta=0$. Here

$$\|f\|_{M_{2+\varepsilon}} := \sup_{r>0, x\in\mathbb{R}^d} r\biggl(\frac{1}{r^d}\int_{B_r(x)}|f(x)|^{2+\varepsilon}dx\biggr)^{\frac{1}{2+\varepsilon}} \tag{2.4}$$

is the Morrey norm. The inclusion is a consequence of the Adams inequality [A, Theorem 7.3].

We note that, for each $0<\varepsilon<\varepsilon_0$, there exist vector fields $b=b(x)$ in the Morrey class $M_{2+\varepsilon}$ such that $|b| \not\in L^{2+\varepsilon_0}_{\rm loc}(\mathbb{R}^d)$.

7. Chang-Wilson-Wolff [CWW] identified a larger than $\cup_{\varepsilon>0}M_{2+\varepsilon}$ sub-class of $\mathbf{F}_\delta \cap \{b=b(x)\}$. That is, they demonstrated that if $|b| \in L^2_{\rm loc}(\mathbb{R}^d)$ and

$$\sup_{r>0,x\in\mathbb{R}^d}\frac{1}{|B_r|}\int_{B_r(x)}|b(y)|^2\,r^2\xi\bigl(|b(y)|^2\,r^2\bigr)dy<\infty, \tag{2.5}$$

where $\xi:\mathbb{R}_+ \to [1,\infty[$ is a fixed increasing function such that

$$\int_1^\infty \frac{ds}{s\xi(s)}<\infty,$$

then $b \in \mathbf{F}_\delta$ with appropriate $\delta$. For instance, one can take $\xi(s)=1+(\log^+ s)^{1+\epsilon}$ or $\xi(s)=1+\log^+ s(\log\log^+ s)^{1+\epsilon}$ for some $\epsilon>0$ (but not $\xi(s)=1+\log^+ s$).

8. In Theorem 2 we impose a more restrictive condition on $g_\delta$, namely, it must be in $L^{1+\varepsilon}(\mathbb{R})$ for some arbitrarily small but fixed $\varepsilon>0$. For instance,

$$|b(t,x)| \leq |t-t_0|^{-\frac{1}{2}+\varepsilon_1} \quad \text{for some } \varepsilon_1>0, \text{ for } t_0 \text{ fixed},$$

satisfies this requirement. Note that, due to this $\varepsilon>0$, the Ladyzhenskaya-Prodi-Serrin class of drifts (1.5) handled by [RZ] is not entirely contained in the class of form-bounded drifts of Theorem 2 unless $q=\infty$ in (1.5), i.e. one focuses on the maximal admissible spatial singularities of the drift.

The following theorem strengthens the result in [RZ] and yields strong existence/conditional uniqueness for drifts that can introduce strong attraction phenomena, e.g. as in example 3.

The authors of [RZ], in turn, strengthened Krylov-Röckner [KrR], i.e. replaced the sub-critical Ladyzhenskaya-Prodi-Serrin class ($<1$) with the critical one ($=1$), see (1.5), with the exception of the pathwise uniqueness since, in the setting of [KrR], one can prove unconditional pathwise uniqueness, see Zhang [Z].

Let us also add that the strong solutions to SDE (2.1) with $b$ in the Ladyzhenskaya-Prodi-Serrin class (1.5), but only for a.e. initial point $x \in \mathbb{R}^d$ – possibly excluding the singular set of the drift – were constructed earlier in [BFGM] via the regularity theory of the stochastic transport/continuity equations.[1]

[1]We note that the authors of [KSS] extended the relevant parts of [BFGM] to time-homogeneous form-bounded drifts, so, strictly speaking, the extension of [RZ] to form-bounded drifts in [KM1] was consistent with what was done earlier.

In Theorem 2 we work on a finite time interval, so, without loss of generality, we impose from now on the global $L^1$ condition on $g_\delta$.

**Theorem 2** ([KM1], but with a few technical conditions on the drift removed, see remark below)**.** *Let $d \geq 3$, $T > 0$. Assume that $b \in \mathbf{F}_{\delta,g}$ and that there exists $\varepsilon > 0$ such that the function $g_\delta$ from (2.2) additionally satisfies*

$$g_\delta \in L^{1+\varepsilon}(\mathbb{R}).$$

*Let $\{b_n\} \subset (C^\infty \cap L^\infty)(\mathbb{R}^{1+d}, \mathbb{R}^d)$ be a sequence of vector fields such that*

$$b_n \to b \quad \text{in } [L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^d,$$

*are form-bounded with the same form-bound $\delta$ having $\sup_n \|g_{\delta,n}\|_{L^{1+\varepsilon}(\mathbb{R})} < \infty$. (For example, the standard mollifier in $\mathbb{R}^{1+d}$ applied to $b$ yields $b_n$, see Appendix A for details.)*

*If the form-bound $\delta$ of drift $b$ is sufficiently small, then the following are true:*

*(i) There exists a subsequence $\{b_{n_k}\}$ such that for any initial time $r \in [0, T[$ and every initial point $x \in \mathbb{R}^d$ the approximating strong solutions $\{X^{x,n_k}_{r,t}\}_{k=1}^\infty$ to*

$$X^{x,n}_{r,t} = x + \int_r^t b_n(s, X^{x,n}_{r,s})ds + W_t - W_r,$$

*considered on a fixed complete probability space $(\Omega, \{\mathcal{F}^W_t\}_{0\leq t\leq T}, \mathcal{F}^W, \mathbf{P})$, converge for every $t \in [r, T]$ $\mathbf{P}$-a.s. to a strong solution $X^x_{r,t}$ of SDE*

$$X^x_{r,t} = x + \int_r^t b(s, X^x_{r,s})ds + W_t - W_r, \quad t \in [r, T].$$

*(ii) The constructed strong solution $X^x_{r,t}$ satisfies the following Krylov-type bound. For every $q \in (d, \delta^{-1/2})$ and any vector field $f \in \mathbf{F}_\beta$, $\beta < \infty$, there exists a constant $C$, depending only on $d$, $q$, $\delta$, $g_\delta$, $\beta$, $g_\beta$ and $T$, such that*

$$\mathbf{E}\int_r^T \left|fh\left(\tau, X^x_{r,\tau}\right)\right| d\tau \leq C \left\| f|h|^{q/2} \right\|^{2/q}_{L^2([r,T], L^2_\rho(\mathbb{R}^d))} \quad \text{for all } h \in C_c\left([r,T] \times \mathbb{R}^d\right). \tag{2.6}$$

*(iii) The constructed solution $X^x_{r,t}$ is pathwise unique among strong solutions to (2.1) that satisfy (2.6) for some $q \in (d, \delta^{-1/2})$, with $|f| = 1$ and with $|f| = |b|$.*

*(iv) The operators*

$$T^{r,t}_b f(x) := \mathbf{E}[f(X^x_{r,t})], \quad f \in C_\infty, \tag{2.7}$$

*constitute a backward Feller propagator on $C_\infty$, and, furthermore, we have*

$$T^{r,t}_b = s\text{-}C_\infty\text{-}\lim_n T^{r,t}_{b_n} \quad \text{loc. uniformly in } r, t \in [0, T]$$

*regardless of the choice of the sequence $\{b_n\}$ as long as it satisfies the preceeding uniform form-boundedness assumptions. In this sense, the law of $X^x_{r,t}$ does not depend on the choice of the subsequence in $(i)$.*

**Remark 1.** Assertion $(iv)$ was proved in [K1]. Its proof uses the analogue of the gradient bounds of Proposition 2 on the solution to Cauchy problem for the Kolmogorov equation $(\partial_t - \frac{1}{2}\Delta + b \cdot \nabla u) = F$, $u|_{t=0} = f$.

The proofs of the Krylov-type bound $(ii)$ and the conditional pathwise uniqueness $(iii)$ remain unchanged. The Krylov-type bound was proved in [KM2]. The proof of the conditional pathwise uniqueness follows the idea of Röckner and Zhao and uses:

– Chernyi's theorem [C], which can be viewed as a dual of the Yamada-Watanabe principle; its application here exploits the fact that the diffusion coefficients in (2.1) are constant;

– conditional weak uniqueness for (2.1) established in [KM2].

The proof of the strong existence $(i)$ repeats the proof in [KM1] with one essential addition. The simplex estimate [KM1, Prop. 1] (= the drift analogue of Proposition 3 in the proof of Theorem 1) imposes the compact support assumption on the drift $b$ (i.e. function $f_i$ in the estimate below) and assumes that the function $g_\delta$ in the form-boundedness condition on $b$ is identically zero[2]. In Section 8 we state and prove [KM1, Prop. 1] in complete generality, as is needed to obtain assertion $(i)$.

2.1. **Weak solutions.** If one is interested in the weak solution theory of SDE (2.1), then one can go beyond the setting of Theorem 2 in several directions. Let us focus for simplicity on $b = b(x)$, although the results cited below are, in general, valid for time-inhomogeneous drifts as well; we also ignore the behaviour of the drift at infinity and focus on admissible local singularities.

1. Theorem 2 requires the form-bound of $b$ to be smaller than a certain dimension-dependent constant. One can, however, prove weak existence/strong Markov property (in fact, one obtains a strongly continuous Feller semigroup), plus some conditional weak uniqueness results, for $b \in \mathbf{F}_\delta$, $\delta < 4$ [KS4, KS3, KV]. One thus arrives at completely dimension-independent conditions on the drift, which is important for applications to particle systems [K4]. The value $\delta = 4$ corresponds to the first critical threshold in the following table, that is, until the PDE methods work. Specifically, consider SDE

$$dX_t = -\frac{d-2}{2}\sqrt{\delta}\frac{X_t}{|X_t|^2}dt + \sqrt{2}W_t, \quad X_0 = 0,$$

with drift set to zero at the origin. We have the following tetrachotomy:

TABLE 1. Weak solvability for $b(x) = -\frac{d-2}{2}\sqrt{\delta}\frac{x}{|x|^2}$

| Values of $\delta$ | Proof of weak existence |
|---|---|
| $[0, 4[$ | Via PDEs with (2.3) viewed as a form-bounded drift [KS4, KS3] |
| $[0, 4\big(\frac{d-1}{d-2}\big)^2[$ | Tightness via SDE exploiting special form of (2.3) [FJ] |
| $[4\big(\frac{d-1}{d-2}\big)^2, 4\big(\frac{d}{d-2}\big)^2[$ | Via $X_t \mapsto \|X_t\|^2 X_t$ or via distributional drift, using special form of (2.3) [FJ, T] |
| $[4\big(\frac{d}{d-2}\big)^2, \infty[$ | No weak solution |

The nonexistence of a weak solution is due to the strong attraction to the origin, see the detailed discussion of the counterexample for $\delta \geq 4\big(\frac{d}{d-2}\big)^2$ in [BFGM, FJ]; in fact, even started away from the origin, if $\delta$ surpasses the threshold $4\big(\frac{d}{d-2}\big)^2$, the solution arrives at the origin with positive probability and stays there; the proof is an application of the theory of Bessel processes.

2. There is weak solution theory to SDE

$$dX_t = b(X_t)dt + \sqrt{2}W_t, \quad X_0 = x \in \mathbb{R}^d, \quad d \geq 2,$$

with the same level of detail as in Theorem 2 for a larger class of weakly form-bounded drifts [KS6, K3, KY]: $|b| \in L^1_{\rm loc}(\mathbb{R}^d)$ and

$$\langle |b|, \varphi^2\rangle \leq \delta \langle |(\lambda - \Delta)^{\frac{1}{4}}\varphi|^2\rangle \text{ for all } \varphi \in C_c^\infty(\mathbb{R}^d),$$

[2] These simplifying assumptions do not affect the class of admissible local singularities of $b$, which was the authors' main interest in [KM1].

for some sufficiently small $\delta > 0$, for some $\lambda > 0$, or for the broad subclass of drifts having sufficiently small Morrey norm

$$\|b\|_{M_{1+\varepsilon}} = \sup_{x\in\mathbb{R}^d, r>0} r\left(\frac{1}{|B_r(x)|}\int_{B_r(x)} |b|^{1+\varepsilon}\right)^{\frac{1}{1+\varepsilon}}$$

for some fixed $0 < \varepsilon \leq 1$ (the smaller it is, the broader is the class). For example, the many-particle drift in the finite-particle approximation of the celebrated Keller-Segel model of chemotaxis, i.e. $b = (b^1, \dots, b^N) : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$,

$$b^i(x) = \frac{\kappa}{N}\sum_{j=1, j\neq i}^{N} \frac{x^i - x^j}{|x^i - x^j|^2}, \quad x = (x^1, \dots, x^N) \in \mathbb{R}^{2N},$$

belongs to these classes, but does not belong to $\mathbf{F}_\delta$ since its square is not locally summable, or, one could also say, because there is no non-trivial Hardy inequality in $\mathbb{R}^2$. In this setting, additionally having discontinuous diffusion coefficients seems to be impossible, at least within the existing approach of [KS6, K3, KY].

3. At the other end of the range (or, perhaps, convex body) of possible results, there are recent results of Krylov on detailed weak solution theory of

$$dX_t = b(X_t)dt + \sqrt{a(X_t)}dW_t, \quad X_0 = x, \quad d \geq 2,$$

with bounded, symmetric $a \geq \kappa I$ in the VMO class (see the introduction) and drift satisfying, for some $q_b > d_0(d,\kappa) > \frac{d}{2}$,

$$b \in M_{q_b} \text{ with sufficiently small Morrey norm.}$$

Under these assumptions, Krylov proves existence and conditional weak uniqueness in Krylov class (=reasonable occupation time estimate). The conditions on the drift $b$ are more restrictive than in the results described in the previous paragraph. Indeed, to deal with the VMO coefficients (in presence of a singular drift) one needs to establish regularity of the second derivatives of solutions of the corresponding Kolmogorov backward equation; see [BKRS, D, Kr3] and the references therein for the relevant regularity theory. The classes of form-bounded or weakly form-bounded drifts, on the other hand, destroy any non-trivial estimate on the second derivatives, but these are not needed when the diffusion coefficients in the SDE are constant or satisfy (1.6); in small dimensions the situation is better, see more detailed discussion in [K2].

## 3. Notations and auxiliary results

### 3.1. Notations.

- Let $\mathcal{B}(X,Y)$ denote the space of bounded linear operators between Banach spaces $X \to Y$, endowed with the operator norm $\|\cdot\|_{X\to Y}$, and set $\mathcal{B}(X) := \mathcal{B}(X,X)$.
- We write $T = s\text{-}Y\text{-}\lim_n T_n$, for $T, T_n \in \mathcal{B}(X,Y)$ if

$$\lim_n \|Tf - T_n f\|_Y = 0 \quad \text{for every } f \in X.$$

- Let $[X]^d$ and $[X]^{d\times m}$ denote the spaces of $d$-vectors and $d \times m$-matrices with entries in $X$.
- Put

$$\langle f, g\rangle = \langle fg\rangle := \int_{\mathbb{R}^d} f(x)g(x)dx.$$

Given a weight $\rho$ on $\mathbb{R}^d$, we write

$$\langle f,g\rangle_\rho=\langle fg\rangle_\rho=\int_{\mathbb{R}^d} f(x)g(x)\rho(x)dx.$$

- Given a matrix $A=(a_{ij})\in\mathbb{R}^{d\times d}$, we denote by $|A|$ its Euclidean norm, i.e. $|A|^2:=\sum_{i,j=1}^d a_{ij}^2$.
- Put $a\wedge b:=\min\{a,b\}$, $a\vee b:=\max\{a,b\}$.
- Set $\nabla_i:=\partial_{x_i}$.
- Let $L^p=L^p(\mathbb{R}^d,dx)$, $W^{1,p}=W^{1,p}(\mathbb{R}^d,dx)$ denote the usual Lebesgue and Sobolev spaces, respectively.
- $C_\infty:=\{f\in C_b(\mathbb{R}^d)\mid \lim_{|x|\to\infty f(x)=0}\}$ endowed with the sup-norm.
- For any $p>1$, we use $p'$ to denote its conjugate $p/(p-1)$; if $p=1$, then $p'=\infty$.
- For a given matrix field $a=(a_{ij})_{i,j=1}^d:\mathbb{R}^d\to\mathbb{R}^{d\times d}$, we define its row divergence vector field $\nabla a:\mathbb{R}^d\to\mathbb{R}^d$ by the formula

$$(\nabla a)_j:=\sum_{i=1}^d\nabla_i a_{ij}.$$

- Given $T_0<T_1$, define the temporal simplex

$$\Delta_n(T_0,T_1):=\{(t_1,\dots,t_n)\in\mathbb{R}^n:T_0\leq t_1\leq\dots\leq t_n\leq T_1\},\qquad \Delta_n(T):=\Delta_n(0,T).$$

- Denote by $\mathbf{E}_{\mathcal{F}_t}$ the conditional expectation with respect to the $\sigma$-algebra $\mathcal{F}_t$.
- Recall that a family of operators $\{P^{r,t}\}_{0\leq r\leq t\leq T}\subset\mathcal{B}(C_\infty)$ is a backward Feller propagator if
  1°) $P^{r,s}P^{s,t}=P^{r,t}$ for every $s\in[r,t]$ (reproduction property).
  2°) $P^{r,r}=\mathrm{Id}$ (identity operator),
  3°) $\|P^{r,t}f\|_\infty\leq\|f\|_\infty$, $P^{r,t}[C_\infty^+]\subset C_\infty^+$,
  4°) for every $f\in C_\infty$, the function $(r,t)\mapsto P^{r,t}f$, $0\leq r\leq t\leq T$, is continuous in $C_\infty$.

  The last property also follows from the separate continuity: for each fixed $r$ the function $[r,T]\ni t\mapsto P^{r,t}f$ is continuous, and for each fixed $t$ the function $[0,t]\ni r\mapsto P^{r,t}f$ is continuous [GvC, Theorem 2.1].

### 3.2. Compactness for $L^2$ random vector fields.

**Lemma 3.1** ([RZ, Lemma 3.1]). *Let $U\subset\mathbb{R}^d$ be a bounded smooth domain. Let $\{F_n=F_n(x,\omega)\}$ be a sequence of random fields in $L^2(U\times\Omega,dx\times\mathbf{P})$. Assume that*

$$\mathbf{E}\int_U\big(|F_n(x)|^2+|\nabla_x F_n(x)|^2\big)dx\leq K \tag{3.1}$$

$$\mathbf{E}\int_U\int_0^T|D_sF_n(x)|^2\,ds\,dx\leq K, \tag{3.2}$$

$$\mathbf{E}\int_U\int_0^T\int_0^T\frac{|D_sF_n(x)-D_{s'}F_n(x)|^2}{|s-s'|^{1+2\beta}}\,ds\,ds'\,dx\leq K \tag{3.3}$$

*for some $\beta>0$ and $K<\infty$ independent of $n$. Then the sequence $\{F_n\}$ is relatively compact in $L^2(U\times\Omega)$.*

We refer to [RZ] for the discussion of the history of this type of results. Regarding the definition of the Malliavin derivative and its properties, see [H, N].

## 4. PDE estimates

In this section, the entries of $\sigma$ and of the other functions have bounded derivatives in the spatial variables, uniformly in $t$, in order to justify the manipulations with the equations; however, the constants in the estimates will not depend on the smoothness of $\sigma(t,\cdot)$. In the next section we will apply our PDE estimates to matrices $\sigma_n$ from the statement of Theorem 1, and so the constants in the resulting estimates will be independent of $n$.

4.1. **Contractivity and gradient bounds for the Kolmogorov equation.** The following simple observation makes the proofs of the results of this section – Propositions 1 and 2 – more transparent. Since the matrix of diffusion coefficients $\sigma$ is bounded, the conditions of Theorem 1 imply that

$$(\nabla_r a_{il})_{i=1}^d \in \mathbf{F}_{\gamma_{rl}} \quad \text{for all } r,l, \qquad \nabla a \in \mathbf{F}_{\delta_a} \tag{4.1}$$

with $\gamma_{rl} \leq C\nu$, $\delta_a \leq C\nu$. It therefore suffices to show that, if $\gamma_{rl}$, $\delta_a$ are sufficiently small, then the assertions of Propositions 1, 2 hold.

**Proposition 1.** *Under the conditions of Theorem 1, for every $p>1$, provided that the form-bound $\nu$ of the derivatives of $\sigma$ is sufficiently small, and $\theta$ is fixed sufficiently small, the solution to the terminal-value problem*

$$\begin{cases} (\partial_t + \frac{1}{2} a \cdot \nabla^2) u = 0 & \text{in } ]0,T[\times \mathbb{R}^d, \\ u|_{t=T} = f \in C_c^\infty(\mathbb{R}^d), \end{cases} \tag{4.2}$$

*where $a=\sigma\sigma^\top$, satisfies*

$$\|u(t)\|_{L^p_\rho} \leq e^{\omega(T-t)} \|f\|_{L^p_\rho}, \quad t \in [0,T],$$

*where $\omega=\omega(c_\delta,\theta)$. The same estimate holds in the non-weighted case $\rho=1$, i.e. $\theta=0$.*

*Proof.* We will only need $\nabla a \in \mathbf{F}_{\delta_a}$, see the remark in the beginning of Section 4.1. The proof given below is standard, but we nevertheless include the details for the reader's convenience.

It will be convenient to re-denote $u(T-t,\cdot)$, $a(T-t,\cdot)$ by $u(t,\cdot)$, $a(t,\cdot)$, respectively, and carry out the proof for the initial-value problem

$$\begin{cases} \partial_t u - \frac{1}{2} a \cdot \nabla^2 u = 0, \\ u|_{t=0} = f. \end{cases} \tag{4.3}$$

Let us rewrite the equation in the divergence form

$$\partial_t u - \frac{1}{2}\nabla \cdot a \cdot \nabla u + \frac{1}{2}(\nabla a)\cdot \nabla u = 0.$$

We now multiply this equation by $\rho|u|^{p-2}u$ and, after integrating by parts, obtain,

$$\begin{aligned} &\frac{1}{p}\frac{d}{dt}\langle |u|^p \rho\rangle + \frac{2(p-1)}{p^2}\langle a\cdot \nabla |u|^{p/2}, \nabla |u|^{p/2}\rho\rangle \\ &\quad = -\frac{1}{p}\langle \nabla a \cdot \nabla |u|^{\frac{p}{2}}, |u|^{\frac{p}{2}}\rho\rangle - \frac{1}{p}\langle a\cdot\nabla |u|^{\frac{p}{2}}, |u|^{\frac{p}{2}}\nabla\rho\rangle. \end{aligned}$$

Multiplying by $p$ and using the ellipticity of $a$, we obtain

$$\begin{aligned} &\frac{d}{dt}\langle |u|^p\rho\rangle + \frac{2(p-1)}{p}\kappa\langle |\nabla |u|^{p/2}|^2\rho\rangle \\ &\quad \leq |\langle \nabla a\cdot \nabla |u|^{\frac{p}{2}}, |u|^{\frac{p}{2}}\rho\rangle| + |\langle a\cdot\nabla |u|^{\frac{p}{2}}, |u|^{\frac{p}{2}}\nabla\rho\rangle| =: I_1 + I_2. \end{aligned} \tag{4.4}$$

Let us handle $I_2$ first. By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
I_2 &\leq \alpha\langle a\cdot\nabla|u|^{\frac{p}{2}},(\nabla|u|^{\frac{p}{2}})\rho\rangle+\frac{1}{4\alpha}\langle a\cdot\nabla\rho,\frac{\nabla\rho}{\rho}|u|^p\rangle \\
&\leq \alpha\|a\|_\infty\langle|\nabla|u|^{\frac{p}{2}}|^2\rho\rangle+\frac{\|a\|_\infty}{4\alpha}\langle|\nabla\rho|^2,|u|^p\rangle \\
&\text{(apply (1.7))} \\
&\leq \alpha\|a\|_\infty\langle|\nabla|u|^{\frac{p}{2}}|^2\rangle+\frac{\|a\|_\infty}{4\alpha}C\theta\langle|u|^p\rho\rangle,
\end{aligned}
$$

where $\|a\|_\infty\leq\|\sigma\|^2_\infty$. Next,

$$
\begin{aligned}
I_1 &\leq \beta\langle|\nabla a|^2,|u|^p\rho\rangle+\frac{1}{4\beta}\langle|\nabla|u|^{\frac{p}{2}}|^2\rho\rangle \\
&\text{(use } \nabla a\in\mathbf{F}_{\delta_a}) \\
&\leq \beta\Big(\delta_a\langle|\nabla(|u|^{\frac{p}{2}}\sqrt{\rho})|^2\rangle+c_{\delta_a}\langle|u|^p\rho\rangle\Big)+\frac{1}{4\beta}\langle|\nabla|u|^{\frac{p}{2}}|^2\rho\rangle.
\end{aligned}
$$

The term $\langle|\nabla(|u|^{\frac{p}{2}}\sqrt{\rho})|^2\rangle$ is bounded from above by the sum of $\langle|\nabla|u|^{\frac{p}{2}}|^2\rho\rangle$ and, via (1.7), $\langle|u|^p\rho\rangle$. Applying these estimates in (4.4) and optimizing in $\alpha$ and $\beta$, we arrive at the inequality

$$
\frac{d}{dt}\langle|u(t)|^p\rho\rangle+\left[\frac{2(p-1)}{p}\kappa-\text{contribution from } I_1,I_2\right]\langle|\nabla|u|^{p/2}|^2\rho\rangle\leq\omega\langle|u|^p\rho\rangle.
$$

Provided that $\delta_a$ is sufficiently small, we can select $\theta$ sufficiently small so that the expression in the square brackets, i.e. the multiple of the energy term, is positive. Discarding the positive energy term, we arrive at the sought bound. $\square$

**Proposition 2** (A priori gradient bounds). *Under the conditions of Theorem 1, for every $q>d$, provided that the form-bound $\nu$ of the derivatives of $\sigma$ is sufficiently small, and $\theta$ is fixed sufficiently small, the solution to*

$$
\begin{cases}
(\partial_t+\frac{1}{2}a\cdot\nabla^2)v=-|F| & \text{in } ]0,T[\times\mathbb{R}^d, \\
v|_{t=T}=f\in C_c^\infty(\mathbb{R}^d),
\end{cases}
\tag{4.5}
$$

*where $F\in\mathbf{F}_\mu$ for some $\mu<\infty$ (here, additionally, bounded and smooth), satisfies*

$$
\begin{aligned}
&\|v\|^q_{L^\infty([0,T],L^q_\rho)}+\|\nabla v\|^q_{L^\infty([0,T],L^q_\rho)}+\|\nabla|\nabla v|^{\frac{q}{2}}\|^2_{L^2([0,T],L^2_\rho)} \\
&\qquad\qquad\leq C\big(\|F\|^2_{L^2([0,T],L^2_\rho)}+\|\nabla f\|^q_{L^q_\rho}+\|f\|^q_{L^q_\rho}\big).
\end{aligned}
$$

*with constant $C=C(T,\theta,d,q,\kappa,\|\sigma\|_\infty,\nu,c_\nu,\mu,c_\mu)$ independent of the smoothness of $\sigma$, $f$ or $F$. In particular, by the Sobolev embedding theorem, for every $z\in\mathbb{Z}^d$*

$$
\|v\|^q_{L^\infty([0,T],B_{10}(z))}\leq C_z\big(\|F\|^2_{L^2([0,T],L^2_\rho)}+\|\nabla f\|^q_{L^q_\rho}+\|f\|^q_{L^q_\rho}\big).
\tag{4.6}
$$

*The same estimates hold in the non-weighted case $\rho=1$, in which case, in the last estimate, one has simply $\|v\|_{L^\infty([0,T],\mathbb{R}^d)}$.*

We prove Proposition 2 in Section 6; see Section 1.3 for the discussion of the role of this proposition in the proof of Theorem 1.

4.2. **Simplex estimate.** Since $\sigma$ is assumed to be smooth in this section, by classical theory there exists a unique strong solution $X_t^x$ to SDE

$$X_t^x = x + \int_0^t \sigma(r, X_r^x)\, dW_r. \tag{4.7}$$

Let $0 \leq T_0 \leq T_1 \leq T$.

At the risk of annoying the reader, we will carry out the proof of the next proposition in a greater generality that is needed for the proof of Theorem 1. Namely, we will assume that the functions $g_\nu$ appearing in the form-boundedness condition on $\nabla\sigma(t,\cdot)$ can be locally unbounded. For the proof of Theorem 1 we only need $g_\nu$ to be bounded, and thus can replace them with constant $c_\nu$. However, keeping in mind possible extensions of Theorem 1 and the importance of the next proposition, we opt for greater generality. (We will also need this generality when we discuss Theorem 2 dealing with singular drift.)

**Proposition 3.** *Let $f_i \in \mathbf{F}_\nu$ $(i \geq 1)$ with the same*

$$g_\nu \in L^{1+\varepsilon}(\mathbb{R})$$

*for some fixed $\varepsilon > 0$. Assuming that the constant $\theta$ in the definition of weight $\rho$ is fixed sufficiently small, there exist positive constants $C_0$, $K$ (independent of the smoothness of $\sigma$ and $f_i$) such that, for every $n \geq 2$, for all $T_0 < T_1$ satisfying $0 \leq T_0 \leq T_1 \leq T$,*

$$\left\| \mathbf{E} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n |f_i(t_i, X_{t_i}^x)|^2 dt_1 \dots dt_n \right\|_{L^2_\rho(\mathbb{R}^d)}^2 \leq C_0 K^{n-2} (T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}}. \tag{4.8}$$

*Moreover, $K$ can be made as small as needed by assuming that the form-bound $\nu$ of $f_i$ is sufficiently small.*

*If $g_\nu \in L^\infty(\mathbb{R})$, then*

$$\left\| \mathbf{E} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n |f_i(t_i, X_{t_i}^x)|^2 dt_1 \dots dt_n \right\|_{L^2_\rho(\mathbb{R}^d)}^2 \leq C_0 K^{n-2} (T_1 - T_0). \tag{4.9}$$

**Corollary 1.** *For every integer $n \geq 2$,*

$$\int_{\mathbb{R}^d} \left[ \mathbf{E} \left( \int_0^t |\nabla\sigma(r, X_r^x)|^2 ds \right)^n \right]^2 \rho(x) dx \leq C_0 (n!)^2 K^{n-2} t^{\frac{\varepsilon}{1+\varepsilon}}.$$

*If, moreover, $g_\nu \in L^\infty(\mathbb{R})$, then $t^{\frac{\varepsilon}{1+\varepsilon}}$ is replaced with $t$.*

*Proof of Corollary 1.* This follows from

$$\begin{aligned} &\int_{\mathbb{R}^d} \left[ \mathbf{E} \left( \int_0^t |\nabla\sigma(r, X_r^x)|^2 ds \right)^n \right]^2 \rho(x) dx \\ &\quad = \int_{\mathbb{R}^d} \left[ n! \mathbf{E} \int_{\Delta_n(0,t)} \prod_{i=1}^n |\nabla\sigma(r_i, X_{r_i}^x)|^2 dr_1, \cdots, dr_n \right]^2 \rho(x) dx. \end{aligned}$$

□

**Remark 2.** The assertions of Proposition 3 and Corollary 1 remain valid if we replace weight $\rho$ by its translation $\rho_y$, for any $y \in \mathbb{R}^d$ fixed, defined by $\rho_y(x) := \rho(x-y)$. In particular, the constants in these estimates do not depend on $y$ since the form-boundedness condition on $f_i$ and $b$ does not involve any special points.

*Proof of Proposition 3.* Put $u_{n+1}=1$ and define consecutively

$$h_k = f_k^2 u_{k+1}, \quad k=1,\dots,n,$$

where $u_k$ solves the terminal-value problem on $[T_0,T_1]$

$$\partial_t u_k + \frac{1}{2} a\cdot\nabla^2 u_k + h_k = 0, \quad u_k(T_1,\cdot)=0. \tag{4.10}$$

Then, repeating the proof of [RZ, Lemma 4.2], we obtain

$$\mathbf{E}_{\mathcal{F}^W_{T_0}} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n \bigg| f_i(t_i, X^x_{t_i})\bigg|^2 dt_1 \dots dt_n = u_1(T_0, X^x_{T_0}).$$

Therefore,

$$\int_{\mathbb{R}^d} \bigg| \mathbf{E} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n |f_i(t_i, X^x_{t_i})|^2 dt_1 \dots dt_n \bigg|^2 dx = \int_{\mathbb{R}^d} |\mathbf{E} u_1(T_0, X^x_{T_0})|^2 dx.$$

The goal is to control $u_1$ in terms of $u_2$ (Step 1), $u_2$ in terms of $u_3$ (Step 2), and so on, until $u_n$ (Step $n$). Note that $u_n$ can be estimated explicitly because $u_{n+1}=1$.

1. Let us first consider the special case:

$$g_{\delta_a} = g_\nu = 0 \text{ in conditions } \nabla a \in \mathbf{F}_{\delta_a},\ f_i \in \mathbf{F}_\nu, \text{ and } \rho \equiv 1.$$

The latter forces us to impose an additional *temporary* condition on the $f_i$:

$$\text{all } f_i \text{ have supports in } [0,T]\times B_R \text{ for some fixed } R \text{ independent of } n. \tag{A}$$

Below we will exclude this condition using non-constant (i.e. polynomially vanishing at infinity) weight $\rho$.

**Remark 3.** This proposition plays a crucial role in the proof of Theorem 1. Specifically, we will be applying it to $f_i = \nabla_k \sigma_{il}$. If we put condition (A) in the assumptions of Proposition 3, this will require us to assume that the matrix of diffusion coefficients has form $\sigma = \sigma_{\rm const} + \sigma_c$, where $\sigma_{\rm const}$ is a constant matrix and $\sigma_c$ has compact support. Since we are interested in describing admissible local discontinuities of $\sigma$, this would not be a deal breaking extra condition on $\sigma$, but with a few additional efforts we will avoid it by means of the weight $\rho$.

What we will obtain for $u_2,\dots,u_n$ are the following two inequalities

$$\int_{T_0}^{T_1} \langle |\nabla u_2|^2\rangle \leq K^{n-2} \int_{T_0}^{T_1} \langle |\nabla u_n|^2\rangle \tag{4.11}$$

where $0<K<1$ can be made as small as needed by assuming that the form-bounds $\delta_a$ and $\nu$ are sufficiently small, and

$$\int_{T_0}^{T_1} \langle |\nabla u_n(s)|^2\rangle ds \leq C(T_1-T_0) \tag{4.12}$$

whose proof is basically the energy inequality for (4.10). The relationship between $u_1$ and $u_2$ is different: we need to prove inequality

$$\int_{\mathbb{R}^d} |\mathbf{E} u_1(T_0, X^x_{T_0})|^2 dx \leq \int_{T_0}^{T_1} \langle |\nabla u_2|^2\rangle. \tag{4.13}$$

Proposition 3 for $g_{\delta_a} = g_\nu = 0$ and $\rho\equiv 1$ follows right away from (4.11)-(4.13).

Let us prove (4.11)-(4.13). Recall that the action of the row-divergence operator $\nabla$ on $a = \sigma\sigma^\top$ is defined as the vector field $\nabla a : \mathbb{R}^{1+d} \to \mathbb{R}^d$ with components $(\nabla a)_k := \sum_{i=1}^d (\nabla_i a_{ik})$. The boundedness of $\sigma$ and the hypothesis of form-boundedness of $|\nabla\sigma|$ of Theorem 1 ensure that

$$\nabla a \in \mathbf{F}_{\delta_a} \text{ for } \delta_a = C\nu. \tag{4.14}$$

Step 1 (proof of (4.13)). Set $v(t,x) := \mathbf{E}_{X_t=x} u_1(T_0, X^x_{T_0})$. The function $v$ satisfies the terminal-value problem on $[0,T_0]$

$$\partial_t v + \frac{1}{2} a \cdot \nabla^2 v = 0, \quad v(T_0,\cdot) = u_1(T_0,\cdot).$$

We need to bound $\langle |v(0,\cdot)|^2 \rangle$, where $v(0,\cdot) = \mathbf{E}_{X_0=x} u_1(T_0, X^x_{T_0}) \equiv \mathbf{E} u_1(T_0, X^x_{T_0})$. So, we define

$$\tilde{u}(t,\cdot) := \begin{cases} v(t,\cdot) & t \in [0,T_0] \\ u_1(t,\cdot) & t \in [T_0,T_1]. \end{cases}$$

Then $\tilde{u}$ satisfies on the whole interval $[0,T_1]$ the terminal-value problem

$$\partial_t \tilde{u} + \frac{1}{2} a \cdot \nabla^2 \tilde{u} + \tilde{h} = 0, \quad \tilde{u}(T_1,\cdot) = 0, \tag{4.15}$$

where

$$\tilde{h}(t,\cdot) = \begin{cases} 0 & t \in [0,T_0], \\ h_1(t,\cdot) & t \in ]T_0,T_1]. \end{cases}$$

So, what we need to bound is $\langle |\tilde{u}(0,\cdot)|^2 \rangle$. To this end, we multiply the equation in (4.15) by $\tilde{u}$ and integrate over $[0,T_1] \times \mathbb{R}^d$:

$$\frac{1}{2}\langle |\tilde{u}(0,\cdot)|^2 \rangle - \frac{1}{2}\int_0^{T_1} \langle a \cdot \nabla^2 \tilde{u}, \tilde{u} \rangle = \int_0^{T_1} \langle \tilde{h}, \tilde{u} \rangle. \tag{4.16}$$

We now use our hypothesis on $a$:

$$\begin{aligned}
-\int_0^{T_1} \langle a \cdot \nabla^2 \tilde{u}, \tilde{u} \rangle &= -\int_0^{T_1} \langle \nabla \cdot a \cdot \nabla \tilde{u}, \tilde{u} \rangle + \int_0^{T_1} \langle (\nabla a) \cdot \nabla \tilde{u}, \tilde{u} \rangle \\
&\quad \text{(integrate by parts and use } a \geq \kappa I \text{ in the first term,} \\
&\quad \text{apply Cauchy-Schwarz in the second term)} \\
&\geq \kappa \int_0^{T_1} \langle |\nabla \tilde{u}|^2 \rangle - \left( \frac{1}{2\sqrt{\delta_a}} \int_0^{T_1} \langle |\nabla a|^2, \tilde{u}^2 \rangle + \frac{\sqrt{\delta_a}}{2} \int_0^{T_1} \langle |\nabla \tilde{u}|^2 \rangle \right) \\
&\quad \text{(use (4.14), where, recall, for now } g_{\delta_a} = 0) \\
&\geq (\kappa - \sqrt{\delta_a}) \int_0^{T_1} \langle |\nabla \tilde{u}|^2 \rangle. 
\end{aligned} \tag{4.17}$$

Next, we estimate the RHS of (4.16):

$$\begin{aligned}\int_0^{T_1}\langle\tilde h,\tilde u\rangle &= \int_{T_0}^{T_1}\langle h_1,u_1\rangle = \int_{T_0}^{T_1}\langle f_1^2u_2,u_1\rangle \\ &\leq \beta\int_{T_0}^{T_1}\langle f_1^2,u_1^2\rangle + \frac{1}{4\beta}\int_{T_0}^{T_1}\langle f_1^2,u_2^2\rangle \\ &\text{(use } f_i\in\mathbf{F}_\nu) \\ &\leq \beta\nu\int_{T_0}^{T_1}\langle|\nabla u_1|^2\rangle + \frac{\nu}{4\beta}\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle \\ &\leq \beta\nu\int_0^{T_1}\langle|\nabla \tilde u|^2\rangle + \frac{\nu}{4\beta}\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle. \end{aligned} \tag{4.18}$$

Applying these estimates in (4.16), we arrive at

$$\frac12\langle|\tilde u(0,\cdot)|^2\rangle + C_1\int_0^{T_1}\langle|\nabla\tilde u|^2\rangle \leq C_2\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle, \tag{4.19}$$

where $C_1=\frac12(\kappa-\sqrt{\delta_a})-\beta\nu$, $C_2=\frac{\nu}{4\beta}$. So, $0<\frac{C_2}{C_1}<1$ can be made arbitrarily small by assuming that the form-bound $\delta_a$ and $\nu$ are sufficiently small. So, we get

$$\int_{\mathbb{R}^d}\mathbf{E}|u_1(T_0,X^x_{T_0})|^2dx \leq 2C_2\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle,$$

i.e. we have proved (4.13). (At this step we actually did not need the term with $\nabla\tilde u$ in (4.19), but since we will be repeating this calculation below, we kept it.)

Step 2 (proof of (4.11)). Now, we repeat the same procedure for $u_2$ in place of $\tilde u$. That is, we multiply equation (4.10) (for $k=2$) by $u_2$ and integrate over $[T_0,T_1]\times\mathbb{R}^d$ to obtain

$$\frac12\langle u_2^2(T_0)\rangle + \frac12(\kappa-\sqrt{\delta_a})\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle = \int_{T_0}^{T_1}\langle h_2,u_2\rangle,$$

where

$$\begin{aligned}\int_{T_0}^{T_1}\langle h_2,u_2\rangle ds &= \int_{T_0}^{T_1}\langle f_2^2u_3,u_2\rangle \\ &\leq \beta\int_{T_0}^{T_1}\langle f_2^2,u_3^2\rangle + \frac{1}{4\beta}\int_{T_0}^{T_1}\langle f_2^2,u_2^2\rangle \\ &\text{(we are using } f_i\in\mathbf{F}_\nu) \\ &\leq \beta\nu\int_{T_0}^{T_1}\langle|\nabla u_2(s)|^2\rangle + \frac{\nu}{4\beta}\int_{T_0}^{T_1}\langle|\nabla u_3(s)|^2\rangle. \end{aligned}$$

Thus, we arrive at

$$\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle \leq \frac{C_2}{C_1}\int_{T_0}^{T_1}\langle|\nabla u_3|^2\rangle.$$

If $n>3$, we repeat this $n-3$ more times:

$$\int_{T_0}^{T_1}\langle|\nabla u_2|^2\rangle \leq \left(\frac{C_2}{C_1}\right)^{n-2}\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle,$$

i.e. we have (4.11).

Step 3 (proof of (4.12)). Finally, we estimate $\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle$. Arguing as above, we have (recall that $u_{n+1}=1$)

$$\begin{aligned}\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle \le C_3\int_{T_0}^{T_1}\langle f_n^2,u_n\rangle \\ &\text{(we apply the Cauchy-Schwarz inequality)}\\ &\le C_4\int_{T_0}^{T_1}\langle f_n^2\rangle+\frac{1}{2}\int_{T_0}^{T_1}\langle f_n^2,u_n^2\rangle\\ &\le C_4\int_{T_0}^{T_1}\langle f_n^2\rangle+\frac{\nu}{2}\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle\\ &\text{(using assumption $(A)$,}\\ &\text{we apply $f_n\in\mathbf{F}_\nu$ in } \int_{T_0}^{T_1}\langle f_n^2\varphi^2\rangle\ge\int_{T_0}^{T_1}\langle f_n^2\rangle \text{ for a smooth } \varphi\ge\mathbf{1}_{B_R(0)})\\ &\le C_5(T_1-T_0)+\frac{\nu}{2}\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle.\end{aligned}$$

Thus, $(1-\frac{\nu}{2})\int_{T_0}^{T_1}\langle|\nabla u_n|^2\rangle\le C_5(T_1-T_0)$, i.e. since $\nu$ is small, we obtain (4.12).

2. We now explain what needs to be modified in the previous calculations to handle the case $g_{\delta_a},g_\nu\in L^{1+\varepsilon}(\mathbb{R})$, but still assuming for now $\rho\equiv1$.

Every time we were applying the form-boundedness condition on $\nabla a$ or $f_i$, we were getting terms with bad sign that needed to be absorbed by the dispersion term in the equation, cf. (4.17). Now, with $g_{\delta_a},g_\nu\neq0$, we will be getting additional bad terms $\int_0^{T_1}g_{\delta_a}\langle\tilde{u}^2\rangle$ and $\int_{T_0}^{T_1}g_\nu\langle\tilde{u}^2\rangle$. These terms, however, can be dealt with simply by introducing in the equation a time-dependent exponential factor. In detail, the following are the extensions of Steps 1-3 from above:

Step 1': Put

$$F(t):=\lambda\int_0^t\big[g_{\delta_a}(s)+g_\nu(s)\big]ds,$$

where $\lambda$ is to be fixed sufficiently large (depending on the values of $\delta_a$ and $\nu$). We multiply (4.15) by $e^F$:

$$\partial_t(e^F\tilde{u})-F'e^F\tilde{u}+\frac{1}{2}a\cdot\nabla^2e^F\tilde{u}+e^F\tilde{h}=0,\quad e^{F(T_1)}\tilde{u}(T_1,\cdot)=0.$$

After multiplying the previous equation by $\tilde{u}$ and integrating over $[0,T_1]\times\mathbb{R}^d$, one sees that the second term from the left

$$\int_0^{T_1}\langle F'e^F\tilde{u}^2\rangle=\lambda\int_0^{T_1}(g_{\delta_a}+g_\nu)e^F\langle\tilde{u}^2\rangle$$

will absorb the new terms $\frac{1}{2\sqrt{\delta_a}}\int_0^{T_1}\langle g_{\delta_a}e^F\tilde{u}^2\rangle$ and $\beta\int_0^{T_1}\langle g_\nu e^F\tilde{u}^2\rangle$ that now appear in (4.17) and (4.18) as long as we fix $\lambda\ge\frac{1}{2\sqrt{\delta_a}}+\beta$. This will give us, instead of (4.19), the following estimate (note that $e^{F(0)}=1$):

$$\frac{1}{2}\langle|\tilde{u}(0,\cdot)|^2\rangle+C_1\int_0^{T_1}e^F\langle|\nabla\tilde{u}|^2\rangle\le\frac{\nu}{4\beta}\int_{T_0}^{T_1}e^F\langle|\nabla u_2|^2\rangle+\frac{1}{4\beta}\int_{T_0}^{T_1}e^Fg_\nu\langle u_2^2\rangle,$$

where $C_1 = \frac{1}{2}(\kappa - \sqrt{\delta_a}) - \beta\nu$ is as before. Hence

$$\mathbf{E}u_1(T_0, X^x_{T_0}) \leq C\biggl(\int_{T_0}^{T_1} e^F\langle|\nabla u_2|^2\rangle + \int_{T_0}^{T_1} e^F g_\nu\langle u_2^2\rangle\biggr). \tag{4.20}$$

Step 2': Set

$$G(t) := \lambda\int_{T_0}^{t}\big(g_{\delta_a}(s) + g_\nu(s)\big)ds.$$

We multiply (4.10) for $k=2$ by $e^G$:

$$\partial_t(e^G u_2) - G' e^G u_2 + \frac{1}{2}a\cdot\nabla^2 e^G u_2 + e^G h_2 = 0, \quad u_2(T_1,\cdot) = 0.$$

Next, we multiply it by $u_2$ and integrate over $[T_0, T_1]\times\mathbb{R}^d$, obtaining (note that $e^{G(T_0)} = 1$)

$$\frac{1}{2}\langle u_2^2(T_0)\rangle + (\frac{\lambda}{2} - \beta)\int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle + (\kappa - \sqrt{\delta_a} - \beta\nu)\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle \\ \leq \frac{1}{4\beta}\biggl(\nu\int_{T_0}^{T_1} e^G\langle|\nabla u_3|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_3^2\rangle\biggr).$$

Since we are going to iterate this inequality ($u_3$ and $u_4$, then $u_4$ and $u_5$, etc), we need to be more precise when specifying the constants, including $\lambda$. Discarding the first (positive) term and assuming that $\nu \leq 1$, we get

$$(\frac{\lambda}{2} - \beta)\int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle + (\kappa - \sqrt{\delta_a} - \beta\nu)\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle \\ \leq \frac{1}{4\beta}\biggl(\int_{T_0}^{T_1} e^G\langle|\nabla u_3|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_3^2\rangle\biggr).$$

So, we take $C_2 := \frac{1}{4\beta}$, $C_1 := \kappa - \sqrt{\delta_a} - \beta\nu$ and fix $\frac{\lambda}{2} > \beta$ so that $\frac{\lambda}{2} - \beta = C_1$. Note that by selecting $\beta$ sufficiently large we can make $C_2$ arbitrarily small. Accordingly, we will have to select $\nu$ small to have $C_1$, say, $C_1 \geq \frac{\kappa}{2}$ and have $0 < \frac{C_2}{C_1} < 1$. This gives us inequality

$$\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle + \int_{T_0}^{T} g_\nu e^G\langle u_2^2\rangle \\ \leq \frac{C_2}{C_1}\biggl(\int_{T_0}^{T_1} e^G\langle|\nabla u_3|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_3^2\rangle\biggr)$$

which we can iterate (since the selection of the parameters $\beta$, $\lambda$, $\delta_a$, $\nu$ does not depend on $n$), obtaining

$$\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle + \int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle \\ \leq \Bigl(\frac{C_2}{C_1}\Bigr)^{n-2}\biggl(\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_n^2\rangle\biggr). \tag{4.21}$$

Step 3': Finally, we estimate $\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu \langle u_n^2\rangle$. Arguing as above, we obtain (recall that $u_{n+1}=1$)

$$\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_n^2\rangle \leq C_3\int_{T_0}^{T_1}\langle e^G f_n^2, u_n\rangle$$

(we are applying Cauchy-Schwarz inequality)

$$\begin{aligned}
&\leq C_4\int_{T_0}^{T_1}\langle e^G f_n^2\rangle + \frac{1}{2}\int_{T_0}^{T_1}\langle f_n^2, e^G u_n^2\rangle \\
&\leq C_4\int_{T_0}^{T_1}\langle e^G f_n^2\rangle + \frac{\nu}{2}\int_{T_0}^{T_1}\langle e^G|\nabla u_n|^2\rangle + \frac{1}{2}\int_{T_0}^{T_1} g_\nu e^G\langle u_n^2\rangle \\
&\leq C_4\int_{T_0}^{T_1}\langle e^G f_n^2\rangle + \frac{1}{2}\bigg(\int_{T_0}^{T_1}\langle e^G|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} g_\nu e^G\langle u_n^2\rangle\bigg).
\end{aligned}$$

Thus,

$$\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_n^2\rangle \leq 2C_4\int_{T_0}^{T_1}\langle e^G f_n^2\rangle.$$

We are now using the compact support assumption (A), i.e. we apply $f_n\in\mathbf{F}_\nu$ in $\int_{T_0}^{T_1}\langle f_n^2\varphi^2\rangle \geq \int_{T_0}^{T_1}\langle f_n^2\rangle$ for a smooth $\varphi\geq \mathbf{1}_{B_R(0)}$, obtaining

$$\begin{aligned}
\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle + \int_{T_0}^{T_1} e^G g_\nu\langle u_n^2\rangle &\leq C_5\int_{T_0}^{T_1}(1+g_\nu) \\
&\leq C_6\big[T_1-T_0+(T_1-T_0)^{\frac{\varepsilon}{1+\varepsilon}}\big], \qquad (4.22)
\end{aligned}$$

where at the last step we have applied $g_\nu\in L^{1+\varepsilon}_{\rm loc}(\mathbb{R})$.

Now, combining (4.20), (4.21) and (4.22), and noting that both $F$, $G$ are non-negative continuous functions that are, thus, bounded on $[T_0,T_1]$, we obtain the assertion of Proposition 3 in the case $\rho\equiv 1$.

3. We finally pass to the weighted space $L^2_\rho(\mathbb{R}^d)$ and remove assumption (A). We write

$$\langle h\rangle_\rho := \int_{\mathbb{R}^d} h(x)\rho(x)\,dx.$$

We first note that the weight $\rho$ works very well together with the form-boundedness condition. Namely, for every fixed $\eta>0$, applying the form-boundedness of $f_i$ to test function $\sqrt{\rho}\,u$ yields

$$\begin{aligned}
\langle f_i^2, u^2\rangle_\rho &\leq \nu\langle|\nabla(\sqrt{\rho}\,u)|^2\rangle + g_\nu\langle u^2\rangle_\rho \\
&\leq (1+\eta)\nu\langle|\nabla u|^2\rangle_\rho + (g_\nu + C_\eta\nu\theta)\langle u^2\rangle_\rho.
\end{aligned}$$

Similarly,

$$\langle|\nabla a|^2, u^2\rangle_\rho \leq (1+\eta)\delta_a\langle|\nabla u|^2\rangle_\rho + (g_{\delta_a}+C_\eta\delta_a\theta)\langle u^2\rangle_\rho.$$

Also,

$$-\langle a\cdot\nabla^2 u, u\rangle_\rho = \langle a\nabla u,\nabla u\rangle_\rho + \langle(\nabla a)\cdot\nabla u, u\rangle_\rho + \int_{\mathbb{R}^d} a\nabla u\cdot\nabla\rho\,u\,dx.$$

Since the matrix field $a$ is bounded and $\frac{|\nabla\rho|^2}{\rho^2}\leq C\theta$, the last term satisfies

$$\bigg|\int_{\mathbb{R}^d} a\nabla u\cdot\nabla\rho\,u\,dx\bigg| \leq \eta\langle|\nabla u|^2\rangle_\rho + C_\eta\theta\langle u^2\rangle_\rho.$$

Thus, after first fixing $\eta > 0$ sufficiently small, the terms containing $\nabla\rho$ only produce an arbitrarily small perturbation of the energy term $\langle|\nabla u|^2\rangle_\rho$, as well as an additional term of the form $C_\eta\theta\langle u^2\rangle_\rho$. Now, we define

$$F(t) := \lambda\int_0^t \big[g_{\delta_a}(s) + g_\nu(s) + C_\eta\theta\big]\,ds, \quad G(t) := \lambda\int_{T_0}^t \big[g_{\delta_a}(s) + g_\nu(s) + C_\eta\theta\big]\,ds.$$

Multiplying the equations by $e^F\widetilde{u}\,\rho$ and $e^G u_k\,\rho$, respectively, repeat Steps $1'$-$3'$. The previous estimates show that, provided $\eta$ and then $\theta$ are chosen sufficiently small, the constants $C_1$ and $C_2$ are changed only by arbitrarily small values. In particular, they can still be chosen so that $0 < \frac{C_2}{C_1} < 1$, and $C_2/C_1$ can still be made arbitrarily small by requiring the form-bounds $\delta_a$ and $\nu$ to be sufficiently small. We therefore obtain

$$\langle|\tilde{u}(0)|^2\rangle_\rho \leq C\left(\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle_\rho dt + \int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle_\rho\right)$$

and

$$\begin{aligned}\int_{T_0}^{T_1} & e^G\langle|\nabla u_2|^2\rangle_\rho\,dt + \int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle_\rho dt \\ &\leq \left(\frac{C_2}{C_1}\right)^{n-2}\left[\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle_\rho dt + \int_{T_0}^{T_1} g_\nu e^G\langle u_n^2\rangle_\rho\,dt\right].\end{aligned}$$

It remains to estimate the last expression. The weighted version of Step $3'$ gives

$$\int_{T_0}^{T_1} e^G\langle|\nabla u_n|^2\rangle_\rho dt + \int_{T_0}^{T_1} g_\nu e^G\langle u_n^2\rangle_\rho dt \leq C\int_{T_0}^{T_1} e^G\langle f_n^2\rangle_\rho\,dt.$$

Here we no longer need the compact support assumption of $f_i$. Namely, applying the form-boundedness inequality for $f_n$ to $\sqrt{\rho}$, we obtain

$$\begin{aligned}\langle f_n^2(t)\rangle_\rho &\leq \nu\langle|\nabla\sqrt{\rho}|^2\rangle + g_\nu(t)\langle\rho\rangle \\ &\leq C\nu\theta + Cg_\nu(t),\end{aligned}$$

where we used $\langle\rho\rangle < \infty$ and $|\nabla\sqrt{\rho}|^2 = \frac{|\nabla\rho|^2}{4\rho} \leq C\theta\rho$. Since $G$ is non-negative and bounded on $[T_0, T_1]$, Hölder's inequality now yields

$$\begin{aligned}\int_{T_0}^{T_1} e^G\langle f_n^2\rangle_\rho\,dt &\leq C\left[\nu\theta(T_1 - T_0) + \int_{T_0}^{T_1} g_\nu(t)dt\right] \\ &\leq C(T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}},\end{aligned}$$

where $T$ is fixed and $g_\nu \in L^{1+\varepsilon}_{\rm loc}(\mathbb{R})$.

Combining the previous estimates, we obtain

$$\left\|\mathbf{E}\int_{\Delta_n(T_0,T_1)}\prod_{i=1}^n |f_i(t_i, X_{t_i}^{\cdot})|^2 dt_1\ldots dt_n\right\|^2_{L^2_\rho(\mathbb{R}^d)} \leq C_0K^{n-2}(T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}},$$

with $K = C_2/C_1$. The proof of Proposition 3 is completed. □

## 5. Proof of Theorem 1(i): strong existence

**Proposition 4.** *Let $X_t^x = X_{0,t}^x$ be the strong solution to (4.7). Then, for every integer $p \geq 2$, there exist constants $K_1$, $K_2$ independent of the smoothness of $\sigma$ such that*

(*i*)

$$\sup_{y\in\mathbb{R}^d} \|\nabla X_t^x - I\|_{L^{2p}(B_1(y),\, L^p(\Omega))} \leq K_1 t^{\frac{1}{4p}} \tag{5.1}$$

*for all* $0 \leq t \leq T$;

(*ii*)

$$\sup_{y\in\mathbb{R}^d} \|D_s X_t^x - \sigma(s, X_s^x)\|_{L^{2p}(B_1(y),\, L^p(\Omega))} \leq K_1 |t-s|^{\frac{1}{4p}} \tag{5.2}$$

*for a.e.* $s \in [0,T]$ *and* $0 \leq s \leq t \leq T$;

(*iii*)

$$\sup_{y\in\mathbb{R}^d} \|D_s X_t^x - D_{s'} X_t^x\|_{L^{2p}(B_1(y),\, L^p(\Omega))} \leq K_2 \big(|s'-s|^{\frac{1}{8p}} + |s'-s|^{\frac{\gamma}{2p}}\big) \tag{5.3}$$

*for a.e.* $s, s' \in [0,T]$ *and* $0 \leq s, s' \leq t \leq T$.

*Proof of Proposition 4.* We will be applying Corollary 1. Without loss of generality, $y = 0$. (In the general case $y \in \mathbb{R}^d$ we would have to apply Corollary 1 with the translated weight $\rho_y$, see Remark 2.)

(*i*) We will consider initial time $0 \leq s < T$, i.e. $X_{s,s}^x = x$; this will allow us to reuse the proof of (*i*) in the proof of (*ii*). Since $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is smooth with bounded derivatives, we can differentiate (4.7) in $x$:

$$\nabla X_{s,t}^x - I = \int_s^t \nabla\sigma(r, X_r^x) \cdot \nabla X_{s,r}^x dW_r, \tag{5.4}$$

where $\nabla\sigma(r,x) \equiv \nabla_x \sigma(r,x) = (\partial_j \sigma_{ik}(r,x))_{1\leq i,j,k\leq d}$ and, in coordinates,

$$(\nabla_x\sigma)(r, X_r^x) \cdot \nabla_x X_{s,r}^x := \left(\sum_{l=1}^d \partial_l \sigma_{ik}(r, X_r^x)\, \partial_j X_{s,r}^{x,l}\right)_{1\leq i,j,k\leq d}.$$

So, with the Brownian motion, we have

$$(\nabla_x\sigma)(r, X_r^x) \cdot \nabla_x X_{s,r}^x \, dW_r = \left(\sum_{k=1}^d \sum_{l=1}^d \partial_l \sigma_{ik}(r, X_r^x)\, \partial_j X_{s,r}^{x,l} dW_r^k\right)_{1\leq i,j\leq d}.$$

Put for brevity

$$J_{s,t}^x := \nabla_x X_{s,t}^x, \quad J_{s,s}^x = I,$$

We estimate

$$\begin{aligned}
\|J^x_{s,t}-I\|^{2p}_{L^{2p}((B_1(0),L^p(\Omega))} &= \int_{B_1(0)} \big[\mathbf{E}|J^x_{s,t}-I|^p\big]^2 dx \\
&\text{(we use (5.4) and then apply the BDG inequality)} \\
&\leq C\int_{B_1(0)}\bigg[\mathbf{E}\bigg(\int_s^t |\nabla\sigma(r,X^x_{s,r})|^2|J^x_{s,r}|^2 dr\bigg)^{\frac{p}{2}}\bigg]^2 dx \\
&\leq C\int_{B_1(0)}\bigg[\mathbf{E}\sup_{s\leq r\leq t}|J^x_{s,r}|^p\bigg(\int_s^t |\nabla\sigma(r,X^x_{s,r})|^2 dr\bigg)^{\frac{p}{2}}\bigg]^2 dx \\
&\text{(use Cauchy-Schwarz twice)} \\
&\leq C\int_{B_1(0)}\mathbf{E}\sup_{s\leq r\leq t}|J^x_{s,r}|^{2p}\,\mathbf{E}\bigg(\int_s^t |\nabla\sigma(r,X^x_{s,r})|^2 dr\bigg)^{p} dx \\
&\leq C\bigg(\int_{B_1(0)}\bigg[\mathbf{E}\sup_{s\leq r\leq t}|J^x_{s,r}|^{2p}\bigg]^2 dx\bigg)^{\frac{1}{2}}\bigg(\int_{B_1(0)}\bigg[\mathbf{E}\bigg(\int_s^t |\nabla\sigma(r,X^x_{s,r})|^2 dr\bigg)^{p}\bigg]^2 dx\bigg)^{\frac{1}{2}} \\
&\text{(apply Corollary 1 the the second factor)} \\
&\leq C\bigg(\int_{B_1(0)}\bigg[\mathbf{E}\sup_{s\leq r\leq t}|J^x_{s,r}|^{2p}\bigg]^2 dx\bigg)^{\frac{1}{2}} C_1 t^{\frac{1}{2}},
\end{aligned}$$

where $C_1 := C_0^{\frac{1}{2}} p! K^{\frac{p-2}{2}}$ (since $p$ is fixed, the growth of $C_1$ in $p$ plays no role here). This yields assertion $(i)$ once we prove that the first factor is bounded from above by a constant independent of the smoothness of $\sigma$. To this end, we will use the following stochastic Gronwall-type estimate:

*Claim* 1.

$$\bigg[\mathbf{E}\sup_{s\leq r\leq t}|J^x_{s,r}|^{2p}\bigg]^2 \leq C_p \mathbf{E} e^{c_p\int_s^t |\nabla\sigma(r,X^x_{s,r})|^2 dr}.$$

*Proof of Claim 1.* Put for brevity $J_t := J^x_{s,t}$ and $V_t := |J_t|^2$, where $x \in B_1(0)$ is fixed. Our goal is to estimate $\mathbf{E}\sup_{s\leq r\leq t}|V_r|^p$. Identity (5.4) becomes $dJ_t = \sum_{l=1}^d A^l_t J_t dW^l_t$, where $A^l_t = (\nabla_x \sigma_{\cdot l})(t, X^x_{s,t})$ (a $d\times d$ matrix field). By Itô's formula,

$$dV_t = 2\sum_{l=1}^d J_t \cdot \Big(A^l_t J_t\Big)\, dW^l_t + \sum_{l=1}^d |A^l_t J_t|^2 dt, \tag{5.5}$$

where $\cdot$ is the Euclidean inner product in $\mathbb{R}^{d\times d}$, i.e. in coordinates,

$$\begin{aligned}
J_t\cdot\Big(A^l_t J_t\Big)\, dW^l_t &= \sum_{i,j,q=1}^d (J_t)_{ij}(A^l_t)_{iq}(J_t)_{qj}\, dW^l_t \\
&= \sum_{i,j,q=1}^d \partial_j X^{x,i}_{s,t}\partial_q\sigma_{il}(t,X^x_{s,t})\partial_j X^{x,q}_{s,t} dW^l_t.
\end{aligned}$$

Since $\sigma$ is smooth, $J_t$ is invertible (for the proof, if needed, see Appendix C) and so $V_t > 0$. Therefore, the identity (5.5) can be rewritten as

$$dV_t = q_t V_t dt + V_t dM_t,$$

where

$$q_t := \frac{\sum_{l=1}^d |A_t^l J_t|^2}{|J_t|^2} \le |\nabla\sigma(t, X_{s,t}^x)|^2, \tag{5.6}$$

and

$$dM_t := c_t dW_t, \quad c_t = (c_t^1, \dots, c_t^d) \quad \text{with } c_t^l := 2\frac{J_t \cdot A_t^l J_t}{|J_t|^2}, \quad |c_t|^2 \le 4|\nabla\sigma(t, X_{s,t}^x)|^2. \tag{5.7}$$

Since $\nabla_x\sigma$ is bounded, cf. the beginning of this section, and thus so is $|c|$, the Itô integral $M_t = \int_s^t c_r \cdot dW_r = \sum_{l=1}^d \int_s^t c_r^l \, dW_r^l$, $t \le T$, is a continuous square-integrable martingale. Applying Itô's formula to $\log V_t$,

$$\begin{aligned} d\log V_t &= \frac{1}{V_t} dV_t - \frac{1}{2V_t^2} d[V]_t \\ &= \frac{1}{V_t}\left(q_t V_t \, dt + V_t dM_t\right) - \frac{1}{2V_t^2} V_t^2 d[M]_t \\ &= q_t dt + dM_t - \frac{1}{2}\, d[M]_t. \end{aligned}$$

Then, we obtain the following representation for $V_t$:

$$V_t = V_s e^{M_t - M_s - \frac{1}{2}([M]_t - [M]_s) + \int_s^t q_r dr},$$

Hence

$$V_t^p = V_s^p e^{p(M_t - M_s) - \frac{p}{2}([M]_t - [M]_s) + p\int_s^t q_r dr},$$

which we rewrite, for a $\lambda > 0$ to be chosen, as

$$V_t^p = V_s^p (Z_t^\lambda)^{\frac{p}{\lambda}} e^{\frac{1}{2}(p\lambda - p)([M]_t - [M]_s) + p\int_s^t q_r dr},$$

where $Z_t^\lambda = e^{\lambda(M_t - M_s) - \frac{\lambda^2}{2}([M]_t - [M]_s)}$. The process $Z^\lambda$ is a positive continuous martingale, $Z_s^\lambda = 1$ – Novikov's condition is trivially satisfied for $M_t$ due to $d[M]_t = |c_t|^2 dt$, the estimates in (5.6), (5.7) and the boundedness $|\nabla\sigma|$. Also by (5.6), (5.7),

$$V_t^p \le V_s^p (Z_t^\lambda)^{\frac{p}{\lambda}} e^{c_p \int_s^t |\nabla\sigma(r, X_{s,r}^x)|^2 dr}.$$

It easily follows that

$$\sup_{s \le r \le t} V_r^p \le V_s^p \left(\sup_{s \le r \le t} Z_r^\lambda\right)^{\frac{p}{\lambda}} e^{c_p \int_s^t |\nabla\sigma(r, X_{s,r}^x)|^2 dr}.$$

Recalling that $J_s = I$, that $V_s$ is a dimension-dependent constant, and applying the Cauchy-Schwarz inequality, we obtain

$$\mathbf{E} \sup_{s \le r \le t} V_r^p \le C\left(\mathbf{E}\left(\sup_{s \le r \le t} Z_r^\lambda\right)^{\frac{2p}{\lambda}}\right)^{\frac{1}{2}} \left(\mathbf{E} e^{2c_p \int_s^t |\nabla\sigma(r, X_{s,r}^x)|^2 dr}\right)^{\frac{1}{2}}. \tag{5.8}$$

Since $Z^\lambda$ is a positive continuous martingale with $Z_s^\lambda = 1$, Doob's weak maximal inequality yields

$$\mathbf{P}\left(\sup_{s \le r \le t} Z_r^\lambda \ge a\right) \le \frac{\mathbf{E} Z_s^\lambda}{a} = \frac{1}{a}, \qquad a > 0.$$

Furthermore, clearly, $\mathbf{P}\left(\sup_{s\leq r\leq t} Z_r^\lambda \geq a\right) \leq 1\wedge\frac{1}{a}$. Now, applying the Cavalieri principle to $\sup_{s\leq r\leq t} Z_r^\lambda$ gives

$$
\begin{aligned}
\mathbf{E}\biggl(\sup_{s\leq r\leq t} Z_r^\lambda\biggr)^{\frac{2p}{\lambda}} &= \frac{2p}{\lambda}\int_0^\infty a^{\frac{2p}{\lambda}-1}\mathbf{P}\left(\sup_{s\leq r\leq t} Z_r^\lambda \geq a\right)da \\
&\leq \frac{2p}{\lambda}\int_0^1 a^{\frac{2p}{\lambda}-1}da + \frac{2p}{\lambda}\int_1^\infty a^{\frac{2p}{\lambda}-2}\,da \\
&= 1 + \frac{\frac{2p}{\lambda}}{1-\frac{2p}{\lambda}} \\
&= \frac{1}{1-\frac{2p}{\lambda}}.
\end{aligned}
$$

This, and the previous estimate (5.8), yield Claim 1, provided that $\lambda$ is fixed sufficiently large. $\square$

Armed with Claim 1, we can now show that $\mathbf{E}\sup_{s\leq r\leq t}|J_{s,r}^x|^{2p}$ is bounded by a constant independent of smoothness of $\sigma$. Namely, we bound the right-hand side of the estimate in Claim 1 as follows. Write

$$
\mathbf{E}e^{c_p\int_s^t|\nabla\sigma(r,X_{s,r})|^2dr} = \sum_{n=0}^\infty \frac{c_p^n}{n!}\mathbf{E}\biggl(\int_s^t|\nabla\sigma(r,X_{s,r})|^2dr\biggr)^n.
$$

Then, using Cauchy-Schwarz,

$$
\begin{aligned}
\int_{B_1(0)}\mathbf{E}e^{c_p\int_s^t|\nabla\sigma(r,X_{s,r})|^2dr}dx &\leq \sum_{n=0}^\infty\frac{c_p^n}{n!}|B_1(0)|^{\frac{1}{2}}\biggl(\int_{B_1(0)}\biggl[\mathbf{E}\biggl(\int_s^t|\nabla\sigma(r,X_{s,r})|^2dr\biggr)^n\biggr]^2dx\biggr)^{\frac{1}{2}} \\
&\qquad \text{(apply Corollary 1 to the } n\geq 2 \text{ terms)} \\
&\leq (\text{term } n=0) + (\text{term } n=1) + \sum_{n=2}^\infty\frac{c_p^n}{n!}|B_1(0)|C_0^{\frac{1}{2}}n!K^{\frac{n-2}{2}}t^{\frac{1}{2}}.
\end{aligned}
$$

The term corresponding to $n=0$ is, clearly, finite and independent of the smoothness of $\sigma$. The term $n=1$ is bounded by the term for $n=2$ by applying elementary inequality $\alpha\leq 1+\alpha^2$ for all $\alpha\geq 0$. Regarding the terms corresponding to $n\geq 2$, i.e. the convergence of the series, it only remains to note that, by Corollary 1, the constant $K$ can be made as small as needed assuming that the form-bound $\nu$ is sufficiently small.

Thus, $\mathbf{E}\sup_{s\leq r\leq t}|J_{s,r}^x|^{2p}dx$ is bounded by a constant independent of the smoothness of $\sigma$, from which assertion $(i)$ follows upon selecting everywhere above $s=0$.

$(ii)$ The Malliavin derivative $D_sX_t^x$, $0\leq s<t\leq T$, satisfies SDE of the same type as (5.4):

$$
D_sX_t^x = \sigma(s,X_s^x) + \int_s^t\nabla\sigma(\tau,X_\tau^x)\cdot D_sX_\tau^x dW_\tau.
$$

So, we can repeat the proof of $(i)$ word by word, except that now the analogue of Claim 1 takes form

$$
\biggl[\mathbf{E}\sup_{s\leq r\leq t}|D_sX_r^x|^{2p}\biggr]^2 \leq C_p\|\sigma\|_\infty^{4p}\mathbf{E}e^{c_p\int_s^t|\nabla\sigma(r,X_{s,r}^x)|^2dr}
$$

since the initial value for $D_sX_r^x$ at $r=s$ is $\sigma(s,X_s^x)$.

$(iii)$ Set $\hat{J}_{s,t}^x := J_{s,t}^{X_s^x}$. Let $0\leq s<s'\leq t\leq T$. We note the standard identities

$$
D_sX_t^x = \hat{J}_{s',t}^xD_sX_{s'}^x
$$

and

$$D_{s'}X_t^x = \hat{J}_{s',t}\sigma(s', X_{s'}^x). \tag{5.9}$$

Hence

$$D_sX_t^x - D_{s'}X_t^x = \hat{J}_{s',t}^x\big(D_sX_{s'}^x - \sigma(s', X_{s'}^x)\big).$$

Apply the Cauchy-Schwarz inequality to the right-hand side,

$$\begin{aligned}&\|\hat{J}_{s',t}^x\big(D_sX_{s'}^x - \sigma(s', X_{s'}^x)\big)\|_{L^{2p}(B_1(0),L^p(\Omega))}\\ &\qquad\leq \|\hat{J}_{s',t}^x\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}\|D_sX_{s'}^x - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}.\end{aligned} \tag{5.10}$$

1. The norm of the first factor $\|\hat{J}_{s',t}^x\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}$ is bounded from above by a constant independent of the smoothness of $\sigma$, which follows, via (5.9), from the uniform ellipticity condition $\sigma\sigma^\top \geq \kappa > 0$, and thus $\|\sigma^{-1}\|_\infty \leq \kappa^{-\frac{1}{2}}$, and the bound on $\|D_{s'}X_t^x\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}$ stemming from assertion $(ii)$ with $p$ replaced by $2p$.

2. Next, we estimate the second factor in (5.10):

$$\begin{aligned}&\|D_sX_{s'}^x - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}\\ &\qquad\leq \|D_sX_{s'}^x - \sigma(s, X_s^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))} + \|\sigma(s, X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}.\end{aligned} \tag{5.11}$$

(a) The first term in (5.11) has the Malliavin derivative and $\sigma$ evaluated at the same time $s$. So, this term can be estimated by applying assertion $(ii)$ with $p$ replaced by $2p$:

$$\|D_sX_{s'}^x - \sigma(s, X_s^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))} \leq K_1|s-s'|^{\frac{1}{8p}}.$$

(b) Let us handle the second term in (5.11):

$$\begin{aligned}&\|\sigma(s, X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}^{4p}\\ &\qquad\leq \|\sigma(s, X_s^x) - \sigma(s', X_s^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}^{4p} + \|\sigma(s', X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}^{4p}\\ &\qquad\leq 2^{4p-2}\|\sigma\|_\infty^{4p-2}\big\langle\big(\mathbf{E}|\sigma(s, X_s^x) - \sigma(s', X_s^x)|\big)^2\big\rangle_{x\in B_1(0)} + \|\sigma(s', X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0),L^{2p}(\Omega))}^{4p}\\ &\qquad=: I_1 + I_2.\end{aligned}$$

b.1 We have

$$\begin{aligned}\text{(constant times) } I_1 \leq C_0\big\langle\big(\mathbf{E}|\sigma(s, X_s^x) - \sigma(s', X_s^x)|\big)^2\rho(x)\big\rangle_{x\in\mathbb{R}^d} &= \|P^{0,s}|\sigma(s,\cdot) - \sigma(s',\cdot)|\|_{L_\rho^2(\mathbb{R}^d)}^2\\ &\quad\text{(apply Proposition 1))}\\ &\leq C\|\sigma(s,\cdot) - \sigma(s',\cdot)\|_{L_\rho^2(\mathbb{R}^d)}^2\\ &\quad\text{(use } (\mathbf{H}) \text{ in Theorem 1)}\\ &\leq C_2|s-s'|^{2\gamma},\end{aligned}$$

where the constant $C$ does not depend on the smoothness of $\sigma$.

b.2 Next, let us estimate $I_2$. Once again, we are looking for an upper bound in terms of a positive power of $|s-s'|$, with constants that do not depend on the smoothness of $\sigma$. Denote by $\sigma_\eta$,

$\eta \downarrow 0$, the standard mollifier in the spatial variable applied to $\sigma$. We write

$$
\begin{aligned}
&\|\sigma(s', X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} \\
&\qquad \leq \|\sigma(s', X_s^x) - \sigma_\eta(s', X_s^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} && (K_1) \\
&\qquad + \|\sigma_\eta(s', X_s^x) - \sigma_\eta(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} && (K_2) \\
&\qquad + \|\sigma_\eta(s', X_{s'}^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))}. && (K_3)
\end{aligned}
$$

The terms ($K_1$) and ($K_3$) are estimated by essentially arguing as in (b.1), e.g. for ($K_1$):

$$
\begin{aligned}
\|\sigma(s', X_s^x) - \sigma_\eta(s', X_s^x)\|^{4p}_{L^{4p}(B_1(0), L^{2p}(\Omega))} &\leq 2^{4p-2}\|\sigma\|_\infty^{4p-2}\big\langle \big(\mathbf{E}|\sigma(s', X_s^x) - \sigma_\eta(s', X_s^x)|\big)^2\big\rangle_{x \in B_1(0)} \\
&= 2^{4p-2}\|\sigma\|_\infty^{4p-2}\|P^{0,s}|\sigma(s', \cdot) - \sigma_\eta(s', \cdot)|\|^2_{L^2(B_1(0))} \\
&\leq C\|P^{0,s}|\sigma(s', \cdot) - \sigma_\eta(s', \cdot)|\|^2_{L^2_\rho(\mathbb{R}^d))} \\
&\text{(use Proposition 1)} \\
&\leq C\|\sigma(s', \cdot) - \sigma_\eta(s', \cdot)\|^2_{L^2_\rho(\mathbb{R}^d)} \\
&\text{(use Lemma B.1)} \\
&\leq C_1\eta.
\end{aligned}
$$

Therefore,

$$(K_1) \leq C_1\eta^{\frac{1}{2p}}.$$

Similarly,

$$(K_3) \leq C_1\eta^{\frac{1}{2p}}.$$

Let us deal with ($K_2$). We have

$$|\sigma_\eta(s', X_s^x) - \sigma_\eta(s', X_{s'}^x)| \leq \|\nabla\sigma_\eta(s', \cdot)\|_{L^\infty(\mathbb{R}^d)}|X_s^x - X_{s'}^x|.$$

Hence, using $\|\nabla\sigma_\eta(s', \cdot)\|_{L^\infty(\mathbb{R}^d)} \leq C_1\eta^{-1}\|\sigma(s')\|_{L^\infty(\mathbb{R}^d)} \leq C\eta^{-1}$, we obtain

$$
\begin{aligned}
\mathbf{E}|\sigma_\eta(s', X_s^x) - \sigma_\eta(s', X_{s'}^x)|^{2p} &\leq C^{2p}\eta^{-2p}\mathbf{E}|X_s^x - X_{s'}^x|^{2p} \\
&= C^{2p}\eta^{-2p}\mathbf{E}\left|\int_s^{s'} \sigma(r, X_r^x)dW_r\right|^{2p} \quad \text{(apply BDG inequality)} \\
&\leq C_2\eta^{-2p}|s' - s|^p.
\end{aligned}
$$

We arrive at the following estimate:

$$(K_2) \leq C_3\eta^{-1}|s' - s|^{\frac{1}{2}}.$$

Combining the estimates on ($K_1$)-($K_3$), we obtain

$$I_2^{\frac{1}{4p}} \leq 2C_1\eta^{\frac{1}{2p}} + C_2\eta^{-1}|s' - s|^{\frac{1}{2}}.$$

The latter, and the estimate on $I_1$, allow us to bound the second term in (5.11) as follows:

$$
\begin{aligned}
\|\sigma(s, X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} &\leq I_1^{\frac{1}{4p}} + I_2^{\frac{1}{4p}} \\
&\leq C\big(|s' - s|^{\frac{\gamma}{2p}} + \eta^{\frac{1}{2p}} + \eta^{-1}|s' - s|^{\frac{1}{2}}\big).
\end{aligned}
$$

The left-hand side does not depend on $\eta$, so it is for us to select. We take $\eta = |s'-s|^{\frac{1}{4}}$, thus arriving at

$$\begin{aligned}\|\sigma(s, X_s^x) - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} &\le C(|s'-s|^{\frac{\gamma}{2p}} + |s'-s|^{\frac{1}{8p}} + |s'-s|^{\frac{1}{4}}) \\ &\quad \text{(use the fact that we work on a finite time interval)} \\ &\le C_1(|s'-s|^{\frac{\gamma}{2p}} + |s'-s|^{\frac{1}{8p}}).\end{aligned}$$

Finally, returning to (5.11), we can estimate

$$\|D_s X_{s'}^x - \sigma(s', X_{s'}^x)\|_{L^{4p}(B_1(0), L^{2p}(\Omega))} \le K_1|s'-s|^{\frac{1}{8p}} + C_1(|s'-s|^{\frac{\gamma}{2p}} + |s'-s|^{\frac{1}{8p}}).$$

It remains to use this in (5.10) to end the proof of $(iii)$. □

Armed with Propositions 2 and 4, we can now prove Theorem 1$(i)$, i.e. strong existence. We follow the approach of [RZ]. Since $\sigma_m$ are smooth, by classical theory there exists a unique continuous random field $X^m : \Delta_2(T) \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ such that

$$X_{s,t}^{x,m} = x + \int_s^t \sigma_m(r, X_{s,r}^{x,m}) dW_r, \quad 0 \le s \le t \le T, \quad x \in \mathbb{R}^d. \tag{5.12}$$

**Remark 4.** We will use the follow standard properties of the approximating SDEs which are obtained from the Picard construction and pathwise uniqueness. For every $m$, the continuous version of the random field $\{X_{r,t}^{x,m} \mid 0 \le r \le t \le T, x \in \mathbb{R}^d\}$ may be chosen so that, on an event of probability one,

$$X_{r,t}^{x,m} = X_{s,t}^{X_{r,s}^{x,m},m}, \quad 0 \le r \le s \le t \le T, \quad x \in \mathbb{R}^d. \tag{5.13}$$

Moreover, for every $r \le t$ and $x \in \mathbb{R}^d$, $X_{r,t}^{x,m}$ is measurable with respect to $\mathcal{F}_{r,t}^W = \sigma(W_u - W_r \mid r \le u \le t)$ after completion. In particular, for fixed $s \le t$, the random map

$$\mathbb{R}^d \ni y \longmapsto X_{s,t}^{y,m}$$

is measurable with respect to $\mathcal{F}_{s,t}^W$ as a $[C_{\mathrm{loc}}(\mathbb{R}^d)]^d$-valued random variable and is independent of $\mathcal{F}_s^W$.

**Step 1.** We make a few preliminary observations. By the standard Burkholder-Davis-Gundy inequality and the uniform (in $m$) boundedness of $\sigma_m$ on $\mathbb{R}^d$,

$$\begin{aligned}\mathbf{E}|X_{s,t_2}^{x,m} - X_{s,t_1}^{x,m}|^r &= \mathbf{E}\left|\int_{t_1}^{t_2} \sigma_m(t, X_{s,t}^{x,m})\, dW_t\right|^r \\ &\le C_r \mathbf{E}\left|\int_{t_1}^{t_2} |\sigma_m(t, X_{s,t}^{x,m})|^2\, dt\right|^{\frac{r}{2}} \le C(t_2 - t_1)^{\frac{r}{2}}.\end{aligned} \tag{5.14}$$

In particular, for $0 \le s \le t \le T$,

$$\sup_m \sup_{x \in \mathbb{R}^d} \mathbf{E}|X_{s,t}^{x,m} - x|^r \le C(t-s)^{\frac{r}{2}}. \tag{5.15}$$

Next, one has

$$\sup_m \sup_{0 \le s \le t \le T} \sup_{y \in \mathbb{R}^d} \mathbf{E} \int_{B_1(y)} |\nabla_x X_{s,t}^{x,m}|^r dx < \infty. \tag{5.16}$$

Indeed,

$$\mathbf{E}\int_{B_1(y)} |\nabla_x X_{s,t}^{x,m}|^r dx \le \left(\int_{B_1(y)} \left(\mathbf{E}|\nabla_x X_{s,t}^{x,m}|^r\right)^2 dx\right)^{\frac{1}{2}} |B_1(y)|^{\frac{1}{2}} = C_d \|\nabla X_{s,t}^{\cdot,m}\|_{L^{2r}(B_1(y), L^r(\Omega))}^r,$$

so it remains to apply Proposition 4$(i)$.

Let now $r > d$. (At Steps 4 and 5 we will have to fix $r$ even larger).

**Step 2.** Let us compare solutions with different initial points, but the same initial time. Applying the Sobolev embedding theorem to (5.16), we obtain the following bound on the Hölder seminorm

$$\sup_m \sup_{0\leq s\leq t\leq T} \sup_{y\in\mathbb{R}^d} \mathbf{E}[X^{\cdot,m}_{s,t}]^r_{C^{1-\frac{d}{r}}(B_1(y))} < \infty,$$

Therefore, for all $0 \leq s \leq t \leq T$, $x, y \in \mathbb{R}^d$ with $|x-y| \leq 1$,

$$\mathbf{E}|X^{x,m}_{s,t} - X^{y,m}_{s,t}|^r \leq C|x-y|^{r-d}, \quad r-d>0.$$

For $|x-y| > 1$, we estimate, using (5.15)

$$\mathbf{E}|X^{x,m}_{s,t} - X^{y,m}_{s,t}|^r \leq C_r(|x-y|^r + \mathbf{E}|X^{x,m}_{s,t} - x|^r + \mathbf{E}|X^{y,m}_{s,t} - y|^r) \leq C|x-y|^r.$$

Combining the last two bounds, we obtain, for all $0 \leq s \leq t \leq T$, $x, y \in \mathbb{R}^d$,

$$\mathbf{E}|X^{x,m}_{s,t} - X^{y,m}_{s,t}|^r \leq C(|x-y|^{r-d} + |x-y|^r). \tag{5.17}$$

**Step 3.** Next, we compare solutions with the same initial points, but different initial times. That is, assume that $0 \leq s_1 \leq s_2 \leq t$. By the flow property of the approximating solutions (see Remark 4),

$$X^{x,m}_{s_1,t} = X^{X^{x,m}_{s_1,s_2},m}_{s_2,t}.$$

Moreover, the map $y \mapsto X^{y,m}_{s_2,t}$ is independent of $\mathcal{F}^W_{s_2}$, while $X^{x,m}_{s_1,s_2}$ is $\mathcal{F}^W_{s_2}$-measurable. Therefore,

$$\mathbf{E}\left|X^{x,m}_{s_1,t} - X^{x,m}_{s_2,t}\right|^r = \mathbf{E}\big[\mathbf{E}|X^{y,m}_{s_2,t} - X^{x,m}_{s_2,t}|^r\big|_{y=X^{x,m}_{s_1,s_2}}\big].$$

Hence

$$\begin{aligned}
\mathbf{E}|X^{x,m}_{s_1,t} - X^{x,m}_{s_2,t}|^r &\leq C_r\bigg(\mathbf{E}\Big|\int_{s_1}^{s_2} \sigma_m(s, X^{x,m}_{s_1,s})\,dW_s\Big|^r + \mathbf{E}\Big|\int_{s_2}^{t} [\sigma_m(s, X^{x,m}_{s_1,s}) - \sigma_m(s, X^{x,m}_{s_2,s})]\,dW_s\Big|^r\bigg) \\
&\quad \text{(apply (5.14) to the first term)} \qquad (5.18) \\
&\leq C_r\bigg((s_2-s_1)^{\frac{r}{2}} + \mathbf{E}\Big|\int_{s_2}^{t} \big[\sigma_m(s, X^{X^{x,m}_{s_1,s_2}}_{s_2,s}) - \sigma_m(s, X^{x,m}_{s_2,s})\big]\,dW_s\Big|^r\bigg) \\
&\quad \text{(reuse the SDE in the second term)} \qquad (5.19) \\
&\leq C_r\bigg(|s_2-s_1|^{\frac{r}{2}} + \mathbf{E}\big|X^{X^{x,m}_{s_1,s_2}}_{s_2,t} - X^{x,m}_{s_2,t}\big|^r\bigg) \qquad (5.20) \\
&= C_r\bigg(|s_2-s_1|^{\frac{r}{2}} + \mathbf{E}\big[\mathbf{E}|X^{y,m}_{s_2,t} - X^{x,m}_{s_2,t}|^r_{y=X^{x,m}_{s_1,s_2}}\big]\bigg) \\
&\quad \text{(use (5.17) in the second term)} \qquad (5.21) \\
&\leq C_r\bigg(|s_2-s_1|^{\frac{r}{2}} + C\mathbf{E}\big(|X^{x,m}_{s_1,s_2} - x|^{r-d} + |X^{x,m}_{s_1,s_2} - x|^r\big)\bigg) \quad \text{(use again (5.14))} \\
&\leq C_{1,r}(s_2-s_1)^{\frac{r-d}{2}}
\end{aligned}$$

(there is additional $(s_2-s_1)^{\frac{r}{2}}$ term, but, since we work on a finite time interval, it is dominated by $(s_2-s_1)^{\frac{r-d}{2}}$).

To summarize: estimates (5.14), (5.17), (5.18) yield, for all $r > d$,

$$\mathbf{E}|X^{x,m}_{s_1,t_1} - X^{y,m}_{s_2,t_2}|^r \leq C_{1,r}(|t_2-t_1|^{\frac{r}{2}} + |x-y|^{r-d} + |x-y|^r + |s_2-s_1|^{\frac{r-d}{2}}) \tag{5.22}$$

for all $(s_i, t_i) \in \Delta_2(T)$, $i = 1, 2$.

**Step 4.** Proposition 4($i$)-($iii$) yields

*Claim* 2. *For fixed* $y \in \mathbb{R}^d$ *and* $0 \le r < t \le T$, *the random fields* $X^{x,m}_{r,t}$, $x \in B_1(y)$ *verify conditions of Lemma 3.1 where we take* $F_m(x) := X^{x,m}_{r,t}$, $U := B_1(y)$.

For the reader's convenience, we included the proof of Claim 2 in Appendix D.

Claim 2 and Lemma 3.1 yield that, for every fixed $(s,t) \in \mathbb{Q}^2 \cap \Delta_2(T)$ and every integer $N \ge 1$, the sequence

$$B_N \times \Omega \ni (x,\omega) \mapsto X^{x,m}_{s,t}(\omega), \quad m = 1, 2, \dots$$

has a convergent subsequence in $L^2(B_N(0) \times \Omega)$ (i.e. we cover ball $B_N(0)$ by a finite number of balls of radius 1 and pass to a convergent subsequence finitely many times). Further, since $\mathbb{Q}^2 \cap \Delta_2(T)$ is countable, a diagonal argument yields a subsequence of $\{X^{\cdot,m}_{s,t}\}_{m\ge1}$ (without loss of generality, still denoted by $\{X^{\cdot,m}_{s,t}\}$), and random fields $Y_{s,t} \in L^2_{\mathrm{loc}}(\mathbb{R}^d \times \Omega)$, $(s,t) \in \mathbb{Q}^2 \cap \Delta_2(T)$ such that, for every $N$ and all $(s,t) \in \mathbb{Q}^2 \cap \Delta_2(T)$

$$X^{\cdot,m}_{s,t} \to Y_{s,t} \quad \text{in } L^2(B_N \times \Omega) \text{ as } m \to \infty.$$

Passing to a further subsequence, we may additionally assume that

$$\sum_{m=1}^{\infty} \|X^{\cdot,m}_{s,t} - Y_{s,t}\|^2_{L^2(B_N\times\Omega)} < \infty, \quad (s,t) \in \mathbb{Q}^2 \cap \Delta_2(T), \quad N \ge 1.$$

It follows that there exists a Borel measurable set $G \subset \mathbb{R}^d$ such that its complement $\mathbb{R}^d - G$ is of measure zero and such that

$$\sum_{m=1}^{\infty} \mathbf{E}\left|X^{x,m}_{s,t} - Y_{s,t}(x)\right|^2 < \infty, \quad x \in G. \tag{5.23}$$

Fix a countable subset $D \subset G$ that is dense in $\mathbb{R}^d$. Then there exists $\Omega_0 \subset \Omega$, $\mathbf{P}(\Omega_0) = 1$, such that

$$X^{x,m}_{s,t} \to Y_{s,t}(x) \quad \text{on } \Omega_0 \text{ for all } x \in D,\ (s,t) \in \mathbb{Q}^2 \cap \Delta_2(T). \tag{5.24}$$

Now, define the sought limiting process, so far on a countable dense set of times and starting points:

$$X^x_{s,t} := Y_{s,t}(x) \quad \text{for } (s,t) \in \mathbb{Q}^2 \cap \Delta_2(T), \quad s < t, \quad x \in D, \qquad \text{and } X^x_{s,s} := x.$$

Clearly, (5.23) yields convergence $X^{x,m}_{s,t} \to X^x_{s,t}$ in $L^2(\Omega)$ for all $(s,t) \in \mathbb{Q}^2 \cap \Delta_2(T)$. Interpolating with (5.15), one has

$$X^{x,m}_{s,t} \to X^x_{s,t} \quad \text{in } L^r(\Omega) \quad \forall r \ge 2.$$

Next, fix an even integer $r > 3d + 4$ (this will be needed at the next step). Applying the previous convergence in (5.22), we obtain

$$\mathbf{E}\left|X^x_{s_1,t_1} - X^y_{s_2,t_2}\right|^r \le C_{1,r}\big(|t_2 - t_1|^{\frac{r}{2}} + |x-y|^{r-d} + |x-y|^r + |s_1 - s_2|^{\frac{r-d}{2}}\big) \tag{5.25}$$

for all $(s_i,t_i) \in \mathbb{Q}^2 \cap \Delta_2(T)$, $x, y \in D$.

**Step 5.** Since $\frac{r-d}{2} > d + 2$ (= dimension of $\Delta_2(T) \times \mathbb{R}^d$), the Kolmogorov-Chentsov theorem (or, rather, its version due to Kunita [Ku, Theorem 1.4.1], see Appendix E) allows us to extend $X^x_{s,t}$ to a continuous random field on $\Delta_2(T) \times \mathbb{R}^d$, which we still denote by $X^x_{s,t}$. Let us show that, after passing to a subsequence of $\{X^m\}$ (as usual, re-denoted as $\{X^m\}$)

$$X^m \to X \quad \mathbf{P}\text{-a.s. locally uniformly on } \Delta_2(T) \times \mathbb{R}^d. \tag{5.26}$$

The rest of Step 5 is the proof of (5.26).

We will need to introduce a few notations that will save us some space later.

Set $K_N := \Delta_2(T) \times \bar{B}_N(0)$, $N \geq 1$.

Define the distance between the time-space parameters $\xi = (s,t,x)$, $\eta = (s',t',y) \in \Delta_2(T) \times \mathbb{R}^d$ by

$$d(\xi,\eta) := |s-s'| + |t-t'| + |x-y|.$$

Using this distance, we define the modulus of continuity of the approximating random field $X^m$ on $K_N$ by

$$\omega_{m,N}(\delta) := \sup_{\xi,\eta \in K_N, d(\xi,\eta) \leq \delta} |X^m(\xi) - X^m(\eta)|,$$

where, in the rest of Step 5, we write

$$X^m(s,t,x) := X^{x,m}_{s,t}.$$

We divide the rest of Step 5 into a few parts. In the first part we deal with random fields $X^m$ and in the second part with $X$.

1. Inequality (5.22) implies that, whenever $\xi, \eta \in K_N$ and $d(\xi,\eta) \leq 1$,

$$\sup_m \mathbf{E}|X^m(\xi) - X^m(\eta)|^r \leq C_{1,r} d(\xi,\eta)^{\frac{r-d}{2}}.$$

This inequality and $\sup_{\xi \in K_N} \sup_m \mathbf{E}|X^m(\xi)|^r < \infty$, which follows from (5.15), allow us to apply the Kolmogorov–Chentsov theorem of [Ku, Theorem 1.4.7] (Appendix E) that yields the tightness of the laws of $X^m|_{K_N}$ in $C(K_N)$. In turn, this yields, for every $\varepsilon > 0$,

$$\lim_{\delta \downarrow 0} \sup_m \mathbf{P}[\omega_{m,N}(\delta) > \varepsilon] = 0 \tag{5.27}$$

for every $\varepsilon > 0$ and every $N$. Indeed, the tightness of $X^m|_{K_N}$ implies that, for any small $\alpha > 0$, there is a compact set $M \subset C(K_N)$ such that the probability that a realization of $X^m|_{K_N}$ is in $M$

$$\mathbf{P}[X^m|_{K_N} \in M] \geq 1 - \alpha.$$

Since $M$ must be equicontinuous, there exists $\delta > 0$ such that $\sup_{f \in M} \sup_{\xi,\eta \in K_N, d(\xi,\eta) \leq \delta} |f(\xi) - f(\eta)| \leq \varepsilon$, and so

$$\mathbf{P}\biggl[\sup_{\xi,\eta \in K_N, d(\xi,\eta) \leq \delta} |X^m(\xi) - X^m(\eta)| \leq \varepsilon\biggr] \geq \mathbf{P}[X^m|_{K_N} \in M] \geq 1 - \alpha.$$

It follows that $\mathbf{P}[\sup_{\xi,\eta \in K_N, d(\xi,\eta) \leq \delta} |X^m(\xi) - X^m(\eta)| > \varepsilon] \leq \alpha$. Since $\alpha$ could be chosen arbitrarily small, we obtain (5.27).

2. Define the modulus of continuity of the limiting field $X$ by

$$\omega_N(\delta) := \sup_{\xi,\eta \in K_N, d(\xi,\eta) \leq \delta} |X(\xi) - X(\eta)|.$$

Since $X$ is already a.s. continuous on compact $K_N$, we have

$$\omega_N(\delta) \to 0 \quad \mathbf{P}\text{-a.s. as } \delta \downarrow 0. \tag{5.28}$$

3. Fix some $\delta > 0$ and put for brevity $\mathbb{D} := (\mathbb{Q}^2 \cap \Delta_2(T)) \times D$. Note that since $\mathbb{D} \cap K_N$ is dense in $K_N$, there exists a finite $\delta$-net $\Gamma_{N,\delta} \subset \mathbb{D} \cap K_N$. So, for every $\xi \in K_N$ there exists $q(\xi) \in \Gamma_{N,\delta}$ such that

$$d(\xi, q(\xi)) \leq \delta.$$

Now we estimate

$$|X^m(\xi) - X(\xi)| \leq |X^m(\xi) - X^m(q(\xi))| + |X^m(q(\xi)) - X(q(\xi))| + |X(q(\xi)) - X(\xi)|.$$

Taking the maximum over $\xi \in K_N$, we further obtain

$$\|X^m - X\|_{C(K_N)} \leq \omega_{m,N}(\delta) + \max_{q\in\Gamma_{N,\delta}} |X^m(q) - X(q)| + \omega_N(\delta).$$

Hence

$$\mathbf{P}\big[\|X^m - X\|_{C(K_N)} > 3\varepsilon\big] \leq \mathbf{P}\big[\omega_{m,N}(\delta) > \varepsilon\big] + \mathbf{P}\big[\max_{q\in\Gamma_{N,\delta}} |X^m(q) - X(q)| > \varepsilon\big] + \mathbf{P}\big[\omega_N(\delta) > \varepsilon\big].$$

For a fixed $\delta > 0$, the second term on the right tends to zero as $m \to \infty$ since the $\delta$-net $\Gamma_{N,\delta}$ is finite and $X^m(q) \to X(q)$ almost surely for every $q \in \mathbb{D}$. Hence, by (5.27),

$$\limsup_{m\to\infty} \mathbf{P}\big[\|X^m - X\|_{C(K_N)} > 3\varepsilon\big] \leq \sup_m \mathbf{P}\big[\omega_{m,N}(\delta) > \varepsilon\big] + \mathbf{P}\big[\omega_N(\delta) > \varepsilon\big].$$

Taking $\delta \downarrow 0$ and using (5.27) and (5.28), we arrive at

$$X^m \to X \quad \text{in probability in } C(K_N), \qquad \text{for every } N \geq 1. \tag{5.29}$$

4. We finally extract a subsequence that converges a.s. on every $K_N$. By (5.29), passing to a subsequence,

$$\mathbf{P}\left[\|X^{m_j} - X\|_{C(K_j)} > 2^{-j}\right] \leq 2^{-j}.$$

By the Borel-Cantelli lemma, on an event of probability one, $\|X^{m_j} - X\|_{C(K_j)} \leq 2^{-j}$ starting with some $j$. Hence, for every fixed $N$,

$$\sup_{(s,t)\in\Delta_2(T), |x|\leq N} \left|X^{x,m_j}_{s,t} - X^x_{s,t}\right| \to 0 \quad \mathbf{P}\text{-a.s.}$$

Thus, after passing to a subsequence, we arrive at (5.26).

**Step 6.** Next, we show that $X^x_{s,t}$, $x \in \mathbb{R}^d$, is a family of strong solutions (Definition 1) to SDE

$$X^x_{s,t} = x + \int_s^t \sigma(r, X^x_{s,r})\, dW_r, \quad s \leq t \leq T \tag{5.30}$$

In view of (5.26), and since $X^x_{s,t}$ are by construction adapted to the increment filtration of the Brownian motion, to end the proof it suffices for us to show that

$$\int_s^t \sigma_m(\tau, X^{x,m}_{s,\tau})\, dW_\tau \to \int_s^t \sigma(\tau, X^x_{s,\tau})\, dW_\tau \quad \text{in } L^2(\Omega, [C[s,T]]^d). \tag{5.31}$$

Indeed, using the BDG inequality, we have

$$
\begin{aligned}
&\mathbf{E}\sup_{s\leq t\leq T}\bigg|\int_s^t \sigma_m(\tau,X^{x,m}_{s,\tau})dW_\tau-\int_s^t\sigma(\tau,X^x_{s,\tau})dW_\tau\bigg|^2 \\
&\qquad\leq 2\mathbf{E}\sup_{s\leq t\leq T}\bigg|\int_s^t(\sigma_m-\sigma_k)(\tau,X^{x,m}_{s,\tau})dW_\tau\bigg|^2 && (5.32)\\
&\qquad+2\mathbf{E}\sup_{s\leq t\leq T}\bigg|\int_s^t\sigma_k(\tau,X^{x,m}_{s,\tau})dW_\tau-\int_s^t\sigma_k(\tau,X^x_{s,\tau})dW_\tau\bigg|^2 \\
&\qquad+2\mathbf{E}\sup_{s\leq t\leq T}\bigg|\int_s^t(\sigma_k-\sigma)(\tau,X^x_{s,\tau})dW_\tau\bigg|^2 \\
&\qquad\leq 2C\mathbf{E}\int_s^T|\sigma_m-\sigma_k|^2(\tau,X^{x,m}_{s,\tau})d\tau+2C\mathbf{E}\int_s^T\Big|\sigma_k(\tau,X^{x,m}_{s,\tau})-\sigma_k(\tau,X^x_{s,\tau})\Big|^2d\tau && (5.33)\\
&\qquad+2C\mathbf{E}\int_s^T|\sigma_k-\sigma|^2(\tau,X^x_{s,\tau})d\tau. && (5.34)
\end{aligned}
$$

We now apply the following consequence of Proposition 2. In what follows, without loss of generality, $x\in B_{10}(0)$.

**Lemma 5.1.** *Define occupation measures: for a Borel measurable set $A\subset[s,t]\times\mathbb{R}^d$, set*

$$\mu^{s,x}_m(A):=\mathbf{E}\int_s^t\mathbf{1}_A(r,X^{x,m}_{s,r})dr,\quad m=1,2,\dots$$

*and*

$$\mu^{s,x}(A):=\mathbf{E}\int_s^t\mathbf{1}_A(r,X^x_{s,r})dr.$$

*The following are true:*

- $(i)$ *For every function $0\leq F\in C_c([s,t]\times\mathbb{R}^d)\cap\mathbf{F}_\mu$,*

$$
\begin{aligned}
\int Fd\mu^{s,x}_m&=\mathbf{E}\int_s^tF(r,X^{x,m}_{s,r})dr\\
&\to\int_s^t\int_{\mathbb{R}^d}Fd\mu^{s,x}=\mathbf{E}\int_s^tF(r,X^x_{s,r})dr\ \text{as}\ m\to\infty
\end{aligned}
$$

  *and*

$$\mathbf{E}\int_s^tF(r,X^{m,x}_{s,r})dr,\ \ \mathbf{E}\int_s^tF(r,X^x_{s,r})dr\leq C\|F\|^{\frac{2}{q}}_{L^2([s,t],L^2_\rho(\mathbb{R}^d))},\tag{5.35}$$

  *where constant $C$ depends on $\mu$, but does not depend on the smoothness or the support of $F$.*
- $(ii)$ *The limiting occupation measure $\mu^{s,x}$ is absolutely continuous with respect to the weighted Lebesgue measure $dr\times\rho(x)dx$ on $[s,t]\times\mathbb{R}^d$.*
- $(iii)$ *For every function $0\leq F\in L^2([s,t]\times L^2_\rho(\mathbb{R}^d))\cap\mathbf{F}_\mu$,*

$$\mathbf{E}\int_s^tF(r,X^x_{s,r})dr\leq C\|F\|^{\frac{2}{q}}_{L^2([s,t],L^2_\rho(\mathbb{R}^d))}.$$

*Proof.* $(i)$ The first estimate in (5.35) is an immediate consequence of Proposition 2, i.e. (4.6), that we apply (on the interval $[s,t]$ rather than $[0,T]$) with initial function $f=0$. The convergence in $(i)$ follows from (5.26) via the Dominated convergence theorem. This convergence transforms the first estimate into the second estimate.

$(ii)$ Our goal is to show that $\mu^{s,x} \ll dr\times\rho(x)dx$. Let $K\subset[s,t]\times\mathbb{R}^d$ be a compact set of $dr\times\rho(x)dx$-measure zero. There exists a decreasing sequence of open sets $U_k\supset K$ such that $(dr\times\rho(x)dx)(U_k)\downarrow 0$. By Urysohn's theorem, there exists a sequence of continuous functions $\zeta_k$ having $\operatorname{sprt}\zeta_k\subset U_k$ and identically equal to 1 on $K$. Now, we estimate:

$$\begin{aligned}\mu^{s,x}(K) &\le \int \zeta_k d\mu^{s,x}\\ &\le C\|\zeta_k\|^{\frac{2}{q}}_{L^2([s,t],L^2_\rho(\mathbb{R}^d))}\downarrow 0 \quad \text{as } k\to\infty.\end{aligned}$$

Hence, $\mu^{s,x}(K)=0$. Since $\mu^{s,x}$ is a Radon measure, by the regularity $\mu^{s,x}(A)=0$ for all $dr\times\rho(x)dx$ measure zero $A$.

$(iii)$ We use $(ii)$ to pass from the second estimate in $(i)$ to $(iii)$. Let $F_n:=\eta^{\mathbb{R}^{1+d}}_{\varepsilon_n}*\mathbf{1}_nF$, where $\eta_{\epsilon_n}$ is a mollifier on $\mathbb{R}^{1+d}$ with $\epsilon_n\downarrow 0$ and $\mathbf{1}_n:=\{(r,x)\in[s,t]\times\mathbb{R}^d \mid |x|\le n, F(r,x)\le n\}$. Then $F_n\in\mathbf{F}_\mu$ with the same $c_\mu$ as $F$, see Appendix A where we show that, selecting $\varepsilon_n\downarrow 0$ sufficiently fast, we have

$$\|F_n-F\|_{L^2([s,t],L^2_\rho(\mathbb{R}^d))}\to 0,$$

hence there exists a subsequence (without loss of generality, $\{F_n\}$ itself) such that

$$F_n\to F \quad dr\times\rho(x)dx\text{-a.e. on } [s,t]\times\mathbb{R}^d.$$

By $(ii)$,

$$F_n\to F \quad \mu^{s,x}\text{-a.e. on } [s,t]\times\mathbb{R}^d.$$

We can apply Fatou's lemma in $n\to\infty$ in the second estimate in $(ii)$ applied to $F_n$ to obtain $(iii)$. $\square$

Let us now use Lemma 5.1 to send (5.33), (5.34) to zero, as $m\to\infty$, by selecting $k$ appropriately. The first term in the first term in (5.33) is bounded, upon selecting $F:=|\sigma_m-\sigma_k|^2$ in Lemma 5.1, as follows:

$$\mathbf{E}\int_s^T|\sigma_m-\sigma_k|^2(\tau,X^{x,m}_{s,\tau})d\tau\le C\|\sigma_m-\sigma_k\|^2_{L^2([0,T],L^2_\rho(\mathbb{R}^d))}.$$

The right-hand side can be made smaller than any given $\varepsilon$, assuming that $m\ge m_\varepsilon$, by fixing $k$ sufficiently large. The same holds, upon selecting $F:=|\sigma_k-\sigma|$ in Lemma 5.1, for (5.34) – select $k$ even larger, if needed, to make (5.34) smaller than given $\varepsilon$. Finally, for thus fixed $k$, the second term in (5.33) goes to zero as $m\to\infty$ by the Dominated convergence theorem. $\square$

**Remark 5.** At the last step we could appeal instead to Krylov's occupation-time estimate. See Section 1.3 for the discussion.

## 6. Proof of Proposition 2

We carry out the proof, which is a rather direct combination of the proofs of the analogous results in [K1] and [KS5], taking into account the remark on $(\nabla_r a_{il})_{i=1}^d$ and $\nabla a$ in the beginning of Section 4.1. Without loss of generality, we assume throughout the proof that the ellipticity constant $\kappa=1$.

1. We first consider the case of the right-hand side $F=0$ and the constant weight $\rho=1$. We will explain in the end what needs to be modified in the proof to include non-zero $F$ with $\rho$ vanishing polynomially at infinity.

We reverse time and carry out the proof for solutions of the corresponding initial-value problem that we write in the divergence form:

$$\begin{cases} \partial_t u - \frac{1}{2}\nabla\cdot a\cdot\nabla + \frac{1}{2}\nabla a\cdot\nabla u = 0 & \text{in } ]0,T]\times\mathbb{R}^d, \\ u|_{t=s} = f, \end{cases} \tag{6.1}$$

where $0 \leq s < T$. Here $a \mapsto \nabla a$ is the row-divergence operator defined by

$$(\nabla a)_k := \sum_{i=1}^d (\nabla_i a_{ik}), \quad k = 1,\dots,d.$$

Below we will first select $s$ close to $T$, but then will remove this constraint and will be able to take $s = 0$.

Set

$$w_r := \nabla_r u, \quad w := (w_r)_{r=1}^d \equiv \nabla u,$$

$$I_q := \sum_{r=1}^d \int_s^\tau \langle (\nabla_r w)^2 |w|^{q-2}\rangle dt, \quad J_q := \int_s^\tau \langle \big|\nabla|w|\big|^2 |w|^{q-2}\rangle dt,$$

$$I_q^a := \sum_{r=1}^d \int_s^\tau \langle (\nabla_r w\cdot a\cdot\nabla_r w)|w|^{q-2}\rangle dt, \quad J_q^a := \int_s^\tau \langle (\nabla|w|\cdot a\cdot\nabla|w|)|w|^{q-2}\rangle dt.$$

Define the commutator:

$$[F,G] := FG - GF.$$

Set $A := -\nabla\cdot a\cdot\nabla$.

We will be using the test function due to Kovalenko-Semënov [KS]:

$$\phi_r := -\partial_r(w_r|w|^{q-2}).$$

**Remark 6.** The use of this test function allows for a fine control of the admissible form-bounds. There are alternative rougher arguments that involve differentiating the equation and integrating by parts in a more straightforward way, see [K2]. One challenging part of the proof given below – not handled in the elliptic setting of [KS] – is dealing with the time derivative of the solution; we deal with it arguing as in [K1].

That is, we multiply the parabolic equation in (6.1) by $\phi_r$, integrate in space and time (from the initial time $s$ up to time $\tau$), and then sum over $r$ to get :

$$\begin{aligned}\frac{1}{q}\langle |w(\tau)|^q\rangle - \frac{1}{q}\langle |\nabla f|^q\rangle + \frac{1}{2}\int_s^\tau \langle Aw, w|w|^{q-2}\rangle dt = \\ -\frac{1}{2}\sum_{r=1}^d \int_s^\tau \langle [\nabla_r, A]u, w_r|w|^{q-2}\rangle dt - \frac{1}{2}\sum_{r=1}^d \int_s^\tau \langle (\nabla a)\cdot w, \phi_r\rangle dt,\end{aligned}$$

where we have used $\sum_{r=1}^d \langle A\nabla_r u, w_r|w|^{q-2}\rangle dt = \int_s^\tau \langle Aw, w|w|^{q-2}\rangle dt$, with $A$ applied to vector field $w$ componentwise. We evaluate:

$$\int_s^\tau \langle Aw, w|w|^{q-2}\rangle dt = I_q^a + (q-2)J_q^a$$

Since $a \geq I$ (i.e. the ellipticity condition with $\kappa = 1$), and so $I_q^a \geq I_q$ and $J_q^a \geq J_q$, we obtain inequality

$$\frac{1}{q}\langle |w(\tau)|^q\rangle - \frac{1}{q}\langle |\nabla f|^q\rangle + \frac{1}{2}I_q + \frac{(q-2)}{2}J_q \leq -\frac{1}{2}\sum_{r=1}^d \int_s^\tau \langle [\nabla_r, A]u, w_r|w|^{q-2}\rangle dt - \frac{1}{2}\sum_{r=1}^d \int_s^\tau \langle (\nabla a)\cdot w, \phi_r\rangle dt. \tag{6.2}$$

Thus, our goal is to estimate the two terms in the right-hand side of (6.2) in terms of $I_q$ and $J_q$. This will be done in the next two claims.

*Claim* 3.

$$\left|\sum_{r=1}^d \int_s^\tau \langle [\nabla_r, A]u, w_r|w|^{q-2}\rangle dt\right| \leq M_1 I_q + N_1 J_q + C_1 \sum_{r,l=1}^d \int_s^\tau g_{\gamma_{rl}}(t)\langle |w|^q\rangle dt,$$

*where*

$$M_1 := \frac{1}{4\alpha_1}, \qquad N_1 := \alpha_1\gamma\frac{q^2}{4} + (q-2)\Big(\alpha_2\gamma\frac{q^2}{4} + \frac{1}{4\alpha_2}\Big), \quad C_1 := \big(\alpha_1 + (q-2)\alpha_2\big),$$

*where $\alpha_1, \alpha_2 > 0$ are arbitrary, and*

$$\gamma := \sum_{r,l=1}^d \gamma_{rl}.$$

Note that $M_1$ and $N_1$ can be made as small as needed by assuming that the form-bounds $\delta_a$ and $\gamma_{rl}$'s are sufficiently small (provided $\alpha_1, \alpha_2 > 0$ are first chosen to be sufficiently large).

*Proof of Claim 3.* We have $[\nabla_r, A] = \nabla_r \sum_{i,l=1}^d \nabla_i(a_{il}\nabla_l) - \sum_{i,l=1}^d \nabla_i(a_{il}\nabla_l\nabla_r)$, so, using integration by parts, we evaluate:

$$\begin{aligned}\sum_{r=1}^d \int_s^\tau \langle [\nabla_r, A]u, w_r|w|^{q-2}\rangle dt &= \sum_{r,i,l=1}^d \int_s^\tau \langle (\nabla_r a_{il})w_l, (\nabla_i w_r)|w|^{q-2}\rangle dt \\ &\quad + (q-2)\sum_{r,i,l=1}^d \int_s^\tau \langle (\nabla_r a_{il})w_l, w_r|w|^{q-2}\nabla_i|w|\rangle dt \\ &\quad \text{(use Cauchy-Schwarz)} \\ &\leq \alpha_1 \sum_{r,l=1}^d \int_s^\tau \langle |\nabla_r a_{\cdot l}|^2|w|^q\rangle dt + \frac{1}{4\alpha_1}I_q \\ &\quad + (q-2)\Big[\alpha_2 \sum_{r,l=1}^d \int_s^\tau \langle |\nabla_r a_{\cdot l}|^2|w|^q\rangle dt + \frac{1}{4\alpha_2}J_q\Big].\end{aligned}$$

Since $\nabla_r a_{\cdot l} \in \mathbf{F}_{\gamma_{rl}}$,

$$\begin{aligned}\sum_{r,l=1}^d \int_s^\tau \langle |\nabla_r a_{\cdot l}|^2|w|^q\rangle dt &\leq \sum_{r,l=1}^d \gamma_{rl}\frac{q^2}{4}\int_s^\tau \langle (\nabla|w|)^2|w|^{q-2}\rangle\, dt + \sum_{r,l=1}^d c_{\gamma_{rl}}\int_s^\tau \langle |w|^q\rangle dt \\ &= \gamma\frac{q^2}{4}J_q + \sum_{r,l=1}^d c_{\gamma_{rl}}\int_s^\tau \langle |w|^q\rangle dt.\end{aligned}$$

Thus,

$$\sum_{r,l=1}^{d}\int_s^\tau \langle [\nabla_r, A]u, w_r|w|^{q-2}\rangle dt \leq \frac{1}{4\alpha_1} I_q + \Big[\alpha_1\gamma\frac{q^2}{4} + (q-2)\big(\alpha_2\gamma\frac{q^2}{4} + \frac{1}{4\alpha_2}\big)\Big] J_q$$
$$+ \big(\alpha_1 + (q-2)\alpha_2\big)\sum_{r,l=1}^{d} c_{\gamma_{rl}}\int_s^\tau \langle |w|^q\rangle dt,$$

which is the claimed inequality. □

*Claim* 4.

$$\left|\sum_{r=1}^{d}\int_s^\tau \langle (\nabla a)\cdot w, \phi_r\rangle dt\right| \leq M_2 I_q + N_2 J_q + C_2 c_{\delta_a}\int_s^\tau \langle |w|^q\rangle dt,$$

*where*

$$M_2 = \frac{\nu_1}{4}\frac{4\varkappa}{4\varkappa - q + 2}\frac{q}{2}, \qquad N_2 = \frac{\nu_1}{4}\frac{4\varkappa}{4\varkappa - q + 2}(q-2)\Big(\varkappa + \frac{1}{2}\Big) + \frac{1}{\nu_1}\frac{q^2}{4}\delta_a + (q-2)\Big[\nu_2 + \frac{\delta_a}{4\nu_2}\frac{q^2}{4}\Big],$$
$$C_2 = \frac{1}{\nu_1} + \frac{q-2}{4\nu_2},$$

*where* $\varkappa > \frac{q-2}{4}$ *and* $\nu_1$, $\nu_2 > 0$ *are arbitrary.*

Again, we can make $M_2$, $N_2$ arbitrarily small by assuming that the form-bound $\delta_a$ of the vector field $\nabla a$ is sufficiently small (provided that $\nu_1$, $\nu_2$ are first chosen to be sufficiently small).

*Proof of Claim 4.* Write $\sum_{r=1}^d \int_s^\tau \langle (\nabla a)\cdot w, \phi_r\rangle dt = \int_s^\tau \langle (\nabla a)\cdot w, \phi\rangle dt$, where

$$\phi := \sum_{r=1}^{d}\phi_r \equiv -\nabla\cdot(w|w|^{q-2}).$$

Now, instead of integrating by parts right away as we did in Claim 3, we first evaluate the test function $\phi$:

$$\phi = (-\Delta u)|w|^{q-2} - w\cdot\nabla|w|^{q-2}, \tag{6.3}$$

so

$$\int_s^\tau \langle (\nabla a)\cdot w, \phi\rangle dt = -\int_s^\tau \langle (\nabla a)\cdot w, \Delta u\,|w|^{q-2}\rangle dt - \int_s^\tau \langle (\nabla a)\cdot w, w\cdot\nabla|w|^{q-2}\rangle dt$$
$$=: W_1 + W_2. \tag{6.4}$$

The rest goes in two steps:

1. We estimate $W_1$ using Cauchy-Schwarz and applying $\nabla a \in \mathbf{F}_{\delta_a}$:

$$|W_1| \leq \frac{\nu_1}{4}\int_s^\tau \langle |w|^{q-2}|\Delta u|^2\rangle dt + \frac{1}{\nu_1}\int_s^\tau \big\langle |\nabla a|^2|w|^q\big\rangle\, dt$$
$$\leq \frac{\nu_1}{4}\int_s^\tau \langle |w|^{q-2}|\Delta u|^2\rangle dt + \frac{1}{\nu_1}\Big[\delta_a\frac{q^2}{4}J_q + c_{\delta_a}\int_s^\tau \langle |w|^q\rangle\Big]$$

In turn, representing $|\Delta u|^2 = (\nabla\cdot w)^2$ and integrating by parts twice, we obtain:

$$\int_s^\tau \langle |w|^{q-2}|\Delta u|^2\rangle dt = -\int_s^\tau \langle w\cdot\nabla|w|^{q-2}, \Delta u\rangle dt + \sum_{r=1}^{d}\int_s^\tau \big\langle w\cdot\nabla w_r, \nabla_r|w|^{q-2}\big\rangle\, dt + I_q$$
$$=: -F + H + I_q,$$

where we estimate

$$|F| \leq (q-2)\left(\frac{1}{4\varkappa}\int_s^\tau \langle |w|^{q-2}|\Delta u|^2\rangle dt + \varkappa J_q\right), \quad |H| \leq (q-2)\left(\frac{1}{2}I_q + \frac{1}{2}J_q\right).$$

Hence

$$\left(1-\frac{q-2}{4\varkappa}\right)\int_s^\tau \langle |w|^{q-2}|\Delta u|^2\rangle dt \leq I_q + (q-2)\left(\varkappa J_q + \frac{1}{2}I_q + \frac{1}{2}J_q\right), \quad \varkappa > \frac{q-2}{4}, \tag{6.5}$$

so

$$|W_1| \leq \frac{\nu_1}{4}\frac{4\varkappa}{4\varkappa - q+2}\left(I_q + (q-2)\left(\varkappa J_q + \frac{1}{2}I_q + \frac{1}{2}J_q\right)\right) + \frac{1}{\nu_1}\left[\delta_a\frac{q^2}{4}J_q + c_{\delta_a}\int_s^\tau \langle |w|^q\rangle dt\right].$$

2. We estimate $W_2$:

$$\begin{aligned}
|W_2| &\leq (q-2)\int_s^\tau \langle |w|^{\frac{q-2}{2}}|\nabla|w|||\nabla a||w|^{\frac{q}{2}}\rangle dt \\
&\leq (q-2)\left[\nu_2\int_s^\tau \langle |w|^{q-2}|\nabla|w||^2\rangle dt + \frac{1}{4\nu_2}\int_s^\tau \langle |\nabla a|^2|w|^q\rangle\, dt\right] \\
&\leq (q-2)\left[\nu_2 J_q + \frac{\delta_a}{4\nu_2}\frac{q^2}{4}J_q + \frac{1}{4\nu_2}c_{\delta_a}\int_s^\tau \langle |w|^q\rangle dt\right].
\end{aligned}$$

It remains to apply these estimates on $|W_1|$, $|W_2|$ in (6.4) to end the proof. □

We now apply Claims 3 and 4 in (6.2):

$$\frac{2}{q}\langle |w(\tau)|^q\rangle + (1-\sum_{i=1}^2 M_i)I_q + (q-2-\sum_{i=1}^2 N_i)J_q \leq \frac{2}{q}\langle |\nabla f|^q\rangle + K\int_s^\tau \langle |w|^q\rangle dt,$$

$$K := C_1\sum_{r,l=1}^d c_{\gamma_{rl}} + C_2 c_{\delta_a},$$

where, in view of the remarks after the claims, $1-\sum_{i=1}^2 M_i > 0$ and $q-2-\sum_{i=1}^2 N_i > 0$, provided that the form-bounds $\gamma_{rl}$ and $\delta_a$ are sufficiently small.

The previous inequality holds for all $s < \tau \leq T$. In fact, since $I_q(\tau)$, $J_q(\tau)$ and the last term are monotonically increasing in $\tau$, we have

$$\begin{aligned}
\frac{2}{q}\sup_{\tau\in[s,T]}\langle |w(\tau)|^q\rangle + (1-\sum_{i=1}^2 M_i)I_q(T) + (q-2-\sum_{i=1}^2 N_i)J_q(T) &\leq \frac{2}{q}\langle |\nabla f|^q\rangle \\
&+ K\int_s^T \langle |w|^q\rangle.
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{2}{q}\sup_{\tau\in[s,T]}\langle |w(\tau)|^q\rangle + (1-\sum_{i=1}^2 M_i)I_q(T) + (q-2-\sum_{i=1}^2 N_i)J_q(T) &\leq \frac{2}{q}\langle |\nabla f|^q\rangle \\
&+ K(T-s)\sup_{\tau\in[s,T]}\langle |w(\tau)|^q\rangle.
\end{aligned}$$

Assuming that the interval $[s,T]$ is sufficiently small, we obtain

$$C_0\sup_{\tau\in[s,T]}\langle |w(\tau)|^q\rangle + (1-\sum_{i=1}^2 M_i)I_q(T) + (q-2-\sum_{i=1}^2 N_i)J_q(T) \leq \frac{2}{q}\langle |\nabla f|^q\rangle,$$

where

$$C_0 := \frac{2}{q} - K(T-s) > 0.$$

Further, applying elementary inequality $I_q \geq J_q$, we arrive at

$$C_0 \sup_{\tau \in [s,T]} \langle |w(\tau)|^q \rangle + C J_q \leq \frac{2}{q} \langle |\nabla f|^q \rangle.$$

for $C>0$, which yields the sought inequality with $K_1 = \frac{C}{C_0}$ and $K_2 = \frac{2}{qC_0}$ – but so far only on a small interval $[s,T]$. We can, however, extend the previous inequality to the interval of arbitrary length using the semigroup property at the expense of increasing the constants $K_1$, $K_2$. This ends the proof in the case $F=0$ and $\rho=1$.

3. Let us now handle non-zero right-hand side $F \neq 0$. Since $\rho \equiv 1$ for now, we assume that $F \in L^2([0,T], L^2(\mathbb{R}^d))$. In this case, after multiplying the parabolic equation by the test function $\phi_r$, summing in $1 \leq r \leq d$ and integrating, we obtain the following term

$$\sum_{r=1}^d \int_s^\tau \langle |F|, \phi_r \rangle =: S.$$

We need to estimate it from above in terms of $I_q$ and $J_q$. To this end, we repeat the argument from [KM2, Proof of Theorem 2.2], namely, using (6.3), we write

$$\begin{aligned} S = \int_s^\tau \langle |w|^{q-2} \Delta v, |F| \rangle + \int_s^\tau \langle w \cdot \nabla |w|^{q-2}, |F| \rangle \\ =: Z_1 + Z_2. \end{aligned}$$

Applying Cauchy-Schwarz, we have

$$\begin{aligned} Z_1 &\leq \frac{\alpha_1}{4} \int_s^\tau \langle |w|^{q-2} |\Delta v|^2 \rangle dt + \frac{1}{\alpha_1} \int_s^\tau \langle |w|^{q-2} |F|^2 \rangle dt \qquad (\alpha_1 > 0) \\ &\text{(we are using (6.5))} \\ &\leq \frac{\varkappa \alpha_1}{4\varkappa - q + 2} \big[ I_q + (q-2)(\varkappa J_q + \frac{1}{2} I_q + \frac{1}{2} J_q) \big] + \frac{1}{\alpha_1} L_q \end{aligned}$$

where

$$L_q := \int_s^\tau \langle |w|^{q-2} |F|^2 \rangle dt.$$

Next,

$$\begin{aligned} Z_2 &\leq (q-2) \int_s^\tau \langle |w|^{\frac{q}{2}-1} \nabla |w|, |w|^{(q-2)/2} |F| \rangle \quad (\alpha_2 > 0) \\ &\leq (q-2) \Big[ \alpha_2 J_q + \frac{1}{4\alpha_2} L_q \Big]. \end{aligned}$$

It remains to estimate $L_q$:

$$
\begin{aligned}
L_q &= \int_s^\tau \langle |F|^{2-\frac{4}{q}} |w|^{q-2}, |F|^{\frac{4}{q}} \rangle \\
&\text{(apply Young's inequality)} \\
&\leq \frac{q-2}{q} \varepsilon^{\frac{q}{q-2}} \int_s^\tau \langle |F|^2 |w|^q \rangle + \frac{2}{q} \varepsilon^{-\frac{q}{2}} \int_s^\tau \langle |F|^2 \rangle \\
&\text{(we are using } F \in \mathbf{F}_\mu) \\
&\leq \frac{q-2}{q} \varepsilon^{\frac{q}{q-2}} \left[ \mu \frac{q^2}{4} J_q + \mu \int_s^\tau \langle |w|^q \rangle \right] + \frac{2}{q} \varepsilon^{-\frac{q}{2}} \int_s^\tau \langle |F|^2 \rangle.
\end{aligned}
$$

This completes the estimate of the right-hand side term $S = \sum_{r=1}^d \int_s^\tau \langle |F|, \phi_r \rangle$.

4. Next, in the weighted setting, i.e. when the weight $\rho$ is non-constant, after we multiply the equation by the test function $\phi_r$, we integrate with respect to the measure $\rho(x)dx$. All occurrences of $\nabla \rho$ resulting from the integration by parts are removed using (1.7). In the previous estimate we will have $\int_s^\tau \langle |F|^2 \rho \rangle$, as claimed, instead of $\int_s^\tau \langle |F|^2 \rangle$. For details, if needed, see [KM2], or refer to the end of the proof of Proposition 3 below.

## 7. Proof of Theorem 1(*ii*): Feller propagator

The assumptions of Theorem 1(*ii*) imply that the symmetric matrix fields $a = \sigma\sigma^\top$, $a_n = \sigma_n \sigma_n^\top$ satisfy

$$
\begin{aligned}
a, a_n \in [L^\infty(\mathbb{R}^{1+d})]^{d\times d}, &\quad a, a_n \geq \kappa I \text{ for some fixed } \kappa > 0, \\
a_n \to a &\quad \text{in } [L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^{d\times d}, \\
\nabla a_n \to \nabla a &\quad \text{in } [L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^d
\end{aligned}
$$

and

$$
(\nabla_r a_{n,il})_{i=1}^d \in \mathbf{F}_{\gamma_{rl}} \quad \text{for all } r,l = 1,\dots,d, \qquad \nabla a_n \in \mathbf{F}_{\delta_a}
$$

with the same constants $c_{\delta_a}$, $c_\nu$. The form-bounds $\delta_a$ and $\gamma_{rl}$ are bounded from above by $C\nu$ for appropriate $C > 0$, where $\nu$ is the form-bound of $|\nabla\sigma|$ – this follows from the product rule and $\|\sigma_n\|_\infty \leq \|\sigma\|_\infty < \infty$. So, it suffices for us to show that if $\delta_a$ and $\gamma_{rl}$ are sufficiently small, then assertion (*ii*) of Theorem 1 holds.

Without loss of generality, the ellipticity constant $\kappa = 1$. We can reverse the direction of time and carry out the proof for solutions of the corresponding initial-value problem written in the divergence form:

$$
\begin{cases}
\partial_t u - \frac{1}{2}\nabla\cdot a_n \cdot \nabla + \frac{1}{2}\nabla a_n \cdot \nabla u_n = 0 & \text{in } ]s,\infty[\times\mathbb{R}^d, \\
u_n|_{t=s} = f.
\end{cases}
\tag{7.1}
$$

It suffices to construct the forward Feller propagator for $f \in C_c^\infty$ and then use the density argument and the $L^\infty$ contractivity.

We subtract the approximating equations in (7.1):

$$
(\partial_t + \frac{1}{2}A_n + \frac{1}{2}\nabla a_n \cdot \nabla)g = \frac{1}{2}\nabla\cdot(a_n - a_m)\cdot\nabla u_m - \frac{1}{2}(\nabla a_n - \nabla a_m)\cdot\nabla u_m \tag{7.2}
$$

where, recall, $A_n = -\nabla\cdot a_n\cdot\nabla$, and we set $g := u_n - u_m$.

**Lemma 7.1** (Iteration inequality). *Fix some $0<\alpha<1$ and $\varsigma$, $\lambda>1$ such that*

$$\varsigma<\frac{d}{d-2+2\alpha},$$

*and*

$$\frac{1/(1-\alpha)}{\lambda}=\frac{d/(d-2+2\alpha)}{\varsigma}.$$

*Then there exist $\theta>0$, $k>1$ and $m_0$ such that for all $m,n\geq m_0$, for all $r\geq r_0>\frac{2}{2-\sqrt{\delta_a}}$, we have*

$$\|u_m-u_n\|_{L^{\frac{r}{1-\alpha}}([s,s+\theta],L^{\frac{rd}{d-2+2\alpha}})}\leq \Big(C_0\|\nabla u_m\|^2_{L^{2\lambda'}([s,s+\theta],L^{2\varsigma'})}\Big)^{\frac{1}{r}}(r^{2k})^{\frac{1}{r}}\|u_m-u_n\|^{1-\frac{2}{r}}_{L^{(r-2)\lambda}([s,s+\theta],L^{(r-2)\varsigma})},\quad (7.3)$$

*for a constant $C_0$ that does not depend on $m$ or $s$. Here, as usual, $\frac{1}{\varsigma}+\frac{1}{\varsigma'}=1$, $\frac{1}{\lambda}+\frac{1}{\lambda'}=1$.*

The proof repeats the proof of [K1, Lemma 2], see also [KS2], and consists of applying Cauchy-Schwarz inequality and using the form-boundedness of $\nabla a_n$. The only new term that we get, after multiplying (7.2) by $g=u_m-u_n$, is

$$\int_s^\tau\langle\nabla\cdot(a_n-a_m)\cdot\nabla u_m,\eta|\eta|^{1-\frac{2}{r}}\rangle$$

where $\eta:=g|g|^{\frac{r-2}{2}}$. It is estimated, after integrating by parts and applying Cauchy-Schwarz, using the uniform boundedness of $a_n$ (in $(t,x)$ and $n$).

Now, one can iterate the inequality in Lemma 7.1; these are Moser's iterations. The crucial point here is that we can appeal to the gradient bound of Proposition 2, after applying the Sobolev embedding theorem and using the interpolation inequality, to establish that in (7.3)

$$\sup_m\|\nabla u_m\|^2_{L^{2\lambda'}([s,s+\theta],L^{2\varsigma'})}<\infty.$$

Set

$$D_T=\{(s,t)\mid 0\leq s\leq t\leq T\},\quad D_{T,\theta}:=D_T\cap\{(s,t)\mid 0\leq t-s\leq\theta\},\quad \theta<T.$$

Let $U_n=\{U_n^{t,s}\}_{0\leq s\leq t\leq T}$ denote the (forward) Feller propagator corresponding to the initial-value problem (7.1).

**Lemma 7.2.** *For any $r_0>\frac{2}{2-\sqrt{\delta_a}}$ there exist $\theta>0$, constants $B<\infty$ and*

$$\gamma:=\big(1-\frac{\varsigma d}{d+2}\big)\Big(1-\frac{\varsigma d}{d+2}+\frac{2\varsigma}{r_0}\Big)^{-1}>0$$

*($1<\varsigma<\frac{d+2}{d}$) independent of $m,n$ such that*

$$\|U_mf-U_nf\|_{L^\infty(D_{T,\theta}\times\mathbb{R}^d)}\leq B\sup_{0\leq s\leq T-\theta}\|U_mf-U_nf\|^\gamma_{L^{r_0}([s,s+\theta],L^{r_0})}\quad \textit{for all } n,m.\quad (7.4)$$

See [K1, proof of Lemma 3], see also [KS2].

That $\{U_mf\}$ indeed converges in $L^{r_0}([s,s+\theta],L^{r_0})$ is proved in the same way as in [K1, Lemma 4]. That is, it suffices to prove that $\{U_mf\}$ is a Cauchy sequence in $L^2([s,s+\theta]\times\mathbb{R}^d)$ and then interpolate with the $L^\infty$ contractivity bound for $u_n$, $u_m$. In turn, the proof of convergence in $L^2([s,s+\theta],L^2)$ consists of two steps:

1) Subtracting again the equations for $u_m$, $u_n$ and applying Cauchy-Schwarz and the uniform in $n$ form-boundedness of $\nabla a_n$, but this time working in the weighted space $L^2_\rho$. The only new term that we get is

$$\int_s^\tau \langle \nabla\cdot(a_n-a_m)\cdot\nabla u_m, \rho(u_m-u_n)\rangle.$$

We need to show that it tends to zero as $n$, $m\to\infty$. This is done using the integration by parts where the property (1.7) plays a crucial role, invoking the gradient bound of Proposition 2 and the convergence $a_n-a_m\to 0$ in $L^2_{\rm loc}(\mathbb{R},L^2_\rho(\mathbb{R}^d))$.

2) "cutting tails" of $u_n$, i.e. working with weight $1-\rho$ and using the compact support assumption on the initial function $f$ (so $u_n$ must be small, uniformly in $n$, when considered away from the support of $f$).

This yields the existence of the limit

$$\exists\ s\text{-}C_\infty\text{-}\lim_{m\to\infty} U^{t,s}_m =: U^{t,s}f \quad (\text{uniformly in } (s,t)\in D_{T,\theta}),$$

which can then be extended to $D_T$ by showing the consistency of the limits. The sought backward Feller propagator is now obtained as $P^{r,t}:=U^{T-r,T-t}$. □

## 8. Proof of Theorem 2

Following the discussion after Theorem 2, we only need to comment on the proof of the strong existence part $(i)$.

In the next estimate, assume that the functions $f_i\in L^2_{\rm loc}(\mathbb{R}^{d+1})$ $(i\geq 1)$ are form-bounded

$$\|f_i(t,\cdot)\varphi\|_2^2\leq \nu\|\nabla\varphi\|_2^2+g_\nu(t)\|\varphi\|_2^2 \tag{8.1}$$

for some $\nu>0$, with $0\leq g_\nu\in L^{1+\varepsilon}(\mathbb{R})$.

By the assumptions of the theorem,

$$b\in\mathbf{F}_{\delta,g_\delta}\qquad\text{with } 0\leq g_\delta\in L^{1+\varepsilon}(\mathbb{R}). \tag{8.2}$$

Everywhere below, drift $b$ and functions $f_i$ are additionally assumed to be smooth. However, the constants in the estimates will not depend on smoothness or boundedness of $b$ and $f_i$.

By classical theory, there exists a unique strong solution $X^x_t$ to

$$X^x_t=x+\int_0^t b(\tau,X^x_\tau)d\tau+W_t.$$

Let $0\leq T_0\leq T_1\leq T$ where $T$ is fixed.

**Proposition 3′.** *Assume that $\theta>0$ in the definition of $\rho$ is fixed sufficiently small. There exist positive constants $C_0$ and $K$, independent of the smoothness and boundedness of $b$ and $f_i$, such that, for every $n\geq 1$, every $0\leq T_0<T_1\leq T$, and every $1\leq\alpha_i\leq d$,*

$$\bigg\|\mathbf{E}\int_{\Delta_n(T_0,T_1)}\prod_{i=1}^n\nabla_{\alpha_i}f_i(t_i,X^{\cdot}_{t_i})\,dt_1\cdots dt_n\bigg\|^2_{L^2_\rho(\mathbb{R}^d)}\leq C_0K^n(T_1-T_0)^{\frac{\varepsilon}{1+\varepsilon}}. \tag{8.3}$$

*Moreover, $K$ can be made arbitrarily small by requiring the form-bounds $\delta$ and $\nu$ in (8.2) and (8.1) to be sufficiently small.*

*Proof.* 1. We first assume

$$\operatorname{sprt} f_i \subset \mathbb{R} \times B_R(0), \quad i \geq 1, \tag{A$_b$}$$

for some $R > 0$ independent of $i$, and work with $\rho \equiv 1$. Fix $n$, put $u_{n+1} = 1$, and define consecutively

$$h_k = (\nabla_{\alpha_k} f_k) u_{k+1}, \qquad k = 1, \dots, n,$$

where $u_k$ solves the terminal-value problem

$$\partial_t u_k + \frac{1}{2}\Delta u_k + b \cdot \nabla u_k + h_k = 0, \qquad u_k(T_1, \cdot) = 0 \quad \text{on } [T_0, T_1]. \tag{8.4}$$

As in [RZ, Proof of Lemma 4.2],

$$\mathbf{E}_{\mathcal{F}^W_{T_0}} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n \nabla_{\alpha_i} f_i(t_i, X^x_{t_i})\, dt_1 \cdots dt_n = u_1(T_0, X^x_{T_0}). \tag{8.5}$$

Set

$$v(t,x) := \mathbf{E} u_1(T_0, X^x_{T_0}), \quad 0 \leq t \leq T_0,$$

and define

$$\tilde{u}(t,\cdot) := \begin{cases} v(t,\cdot), & 0 \leq t \leq T_0, \\ u_1(t,\cdot), & T_0 \leq t \leq T_1, \end{cases} \qquad \tilde{h}(t,\cdot) := \begin{cases} 0, & 0 \leq t < T_0, \\ h_1(t,\cdot), & T_0 \leq t \leq T_1. \end{cases}$$

Then

$$\partial_t \tilde{u} + \frac{1}{2}\Delta \tilde{u} + b \cdot \nabla \tilde{u} + \tilde{h} = 0, \quad \tilde{u}(T_1, \cdot) = 0,$$

and

$$\tilde{u}(0,x) = \mathbf{E} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n \nabla_{\alpha_i} f_i(t_i, X^x_{t_i})\, dt_1 \cdots dt_n.$$

Put, as in the proof of Proposition 3,

$$F(t) := \lambda \int_0^t \big[g_\delta(s) + g_\nu(s)\big]\, ds, \quad G(t) := \lambda \int_{T_0}^t \big[g_\delta(s) + g_\nu(s)\big]\, ds. \tag{8.6}$$

In particular,

$$e^{F(t)} = e^{F(T_0)} e^{G(t)}, \qquad T_0 \leq t \leq T_1.$$

Fix $M, \beta > 0$. Multiplying the equation for $\tilde{u}$ by $e^F \tilde{u}$ and integrating over $[0, T_1] \times \mathbb{R}^d$, we obtain

$$\begin{aligned} \frac{1}{2}\langle \tilde{u}^2(0)\rangle + \frac{1}{2}\int_0^{T_1} e^F \langle |\nabla \tilde{u}|^2\rangle\, dt + \frac{\lambda}{2}\int_0^{T_1} (g_\delta + g_\nu) e^F \langle \tilde{u}^2\rangle\, dt \\ = \int_0^{T_1} e^F \langle b \cdot \nabla \tilde{u}, \tilde{u}\rangle\, dt + \int_{T_0}^{T_1} e^F \langle (\nabla_{\alpha_1} f_1) u_2, u_1\rangle\, dt. \end{aligned}$$

The drift term satisfies

$$\int_0^{T_1} e^F \langle b \cdot \nabla \tilde{u}, \tilde{u}\rangle\, dt \leq \sqrt{\delta} \int_0^{T_1} e^F \langle |\nabla \tilde{u}|^2\rangle\, dt + \frac{1}{2\sqrt{\delta}} \int_0^{T_1} g_\delta e^F \langle \tilde{u}^2\rangle\, dt.$$

Integrating by parts in the last term of the preceding energy identity and applying the Cauchy-Schwarz inequality twice, we get

$$\begin{aligned}
&\int_{T_0}^{T_1} e^F\langle(\nabla_{\alpha_1} f_1)u_2, u_1\rangle\,dt\\
&= -\int_{T_0}^{T_1} e^F\langle f_1(\nabla_{\alpha_1} u_2), u_1\rangle\,dt - \int_{T_0}^{T_1} e^F\langle f_1 u_2, \nabla_{\alpha_1} u_1\rangle\,dt\\
&\le M\int_{T_0}^{T_1} e^F\langle f_1^2, u_1^2\rangle\,dt + \frac{1}{4M}\int_{T_0}^{T_1} e^F\langle|\nabla u_2|^2\rangle\,dt\\
&\quad + \beta\int_{T_0}^{T_1} e^F\langle f_1^2, u_2^2\rangle\,dt + \frac{1}{4\beta}\int_{T_0}^{T_1} e^F\langle|\nabla u_1|^2\rangle\,dt\\
&\le \left(M\nu + \frac{1}{4\beta}\right)\int_{T_0}^{T_1} e^F\langle|\nabla u_1|^2\rangle\,dt\\
&\quad + \left(\beta\nu + \frac{1}{4M}\right)\int_{T_0}^{T_1} e^F\langle|\nabla u_2|^2\rangle\,dt\\
&\quad + M\int_{T_0}^{T_1} g_\nu e^F\langle u_1^2\rangle\,dt + \beta\int_{T_0}^{T_1} g_\nu e^F\langle u_2^2\rangle\,dt.
\end{aligned}$$

Set

$$C_1 := \frac{1}{2} - \sqrt{\delta} - M\nu - \frac{1}{4\beta}, \qquad C_2 := \beta\nu + \frac{1}{4M}.$$

We first choose $M$ and $\beta$ sufficiently large, then choose a constant $c > 0$ sufficiently large, and finally require $\delta$ and $\nu$ to be sufficiently small, so that $C_1 > 0$ and $K := \frac{C_2}{C_1} \vee \frac{\beta}{cC_1} < 1$. We can furthermore choose the aforementioned constants to make $K$ as small as needed. Fix $\lambda$ sufficiently large that

$$\frac{\lambda}{2} - \frac{1}{2\sqrt{\delta}} \ge 0, \qquad \frac{\lambda}{2} - M \ge cC_1.$$

Then

$$\langle\widetilde{u}^2(0)\rangle \le CK\left(\int_{T_0}^{T_1} e^G\langle|\nabla u_2|^2\rangle\,dt + c\int_{T_0}^{T_1} g_\nu e^G\langle u_2^2\rangle\,dt\right), \tag{8.7}$$

where the factor $e^{F(T_0)}$ has been absorbed into $C$.

We next apply the same calculation to $u_k$, $2 \le k \le n-1$. Multiplying (8.4) by $e^G u_k$ and integrating over $[T_0, T_1] \times \mathbb{R}^d$ gives

$$\begin{aligned}
&C_1\int_{T_0}^{T_1} e^G\langle|\nabla u_k|^2\rangle\,dt + \left(\frac{\lambda}{2} - M\right)\int_{T_0}^{T_1} g_\nu e^G\langle u_k^2\rangle\,dt\\
&\quad + \left(\frac{\lambda}{2} - \frac{1}{2\sqrt{\delta}}\right)\int_{T_0}^{T_1} g_\delta e^G\langle u_k^2\rangle\,dt\\
&\le C_2\int_{T_0}^{T_1} e^G\langle|\nabla u_{k+1}|^2\rangle\,dt + \beta\int_{T_0}^{T_1} g_\nu e^G\langle u_{k+1}^2\rangle\,dt.
\end{aligned}$$

Consequently,

$$\int_{T_0}^{T_1} e^G\left(\langle|\nabla u_k|^2\rangle + c\,g_\nu\langle u_k^2\rangle\right)\,dt \le K\int_{T_0}^{T_1} e^G\left(\langle|\nabla u_{k+1}|^2\rangle + c\,g_\nu\langle u_{k+1}^2\rangle\right)\,dt. \tag{8.8}$$

For $n \geq 2$, iteration gives

$$\int_{T_0}^{T_1} e^G \left(\langle|\nabla u_2|^2\rangle + c\, g_\nu \langle u_2^2\rangle\right)\, dt \leq K^{n-2} \int_{T_0}^{T_1} e^G \left(\langle|\nabla u_n|^2\rangle + c\, g_\nu \langle u_n^2\rangle\right)\, dt.$$

It remains to estimate the last integral. Since $u_{n+1} = 1$, we have $h_n = \nabla_{\alpha_n} f_n$. The energy inequality for $u_n$ and the integration by parts give

$$\begin{aligned}
&\int_{T_0}^{T_1} e^G \left(\langle|\nabla u_n|^2\rangle + c\, g_\nu \langle u_n^2\rangle\right)\, dt \\
&\quad \leq C \left| \int_{T_0}^{T_1} e^G \langle \nabla_{\alpha_n} f_n, u_n\rangle\, dt \right| \\
&\quad = C \left| \int_{T_0}^{T_1} e^G \langle f_n, \nabla_{\alpha_n} u_n\rangle\, dt \right| \\
&\quad \leq \frac{1}{2} \int_{T_0}^{T_1} e^G \left(\langle|\nabla u_n|^2\rangle + c\, g_\nu \langle u_n^2\rangle\right)\, dt + C \int_{T_0}^{T_1} e^G \langle f_n^2\rangle\, dt.
\end{aligned}$$

Hence,

$$\int_{T_0}^{T_1} e^G \left(\langle|\nabla u_n|^2\rangle + c\, g_\nu \langle u_n^2\rangle\right)\, dt \leq C \int_{T_0}^{T_1} e^G \langle f_n^2\rangle\, dt. \tag{8.9}$$

Choose $\varphi \in C_c^\infty(\mathbb{R}^d)$ such that $\varphi \geq 1$ on $B_R(0)$. By ($\mathrm{A}_b$) and (8.1),

$$\langle f_n^2(t)\rangle \leq \langle f_n^2(t), \varphi^2\rangle \leq \nu\langle|\nabla\varphi|^2\rangle + g_\nu(t)\langle\varphi^2\rangle.$$

Since $G$ is bounded on $[T_0, T_1]$, Hölder's inequality yields

$$\int_{T_0}^{T_1} e^G \langle f_n^2\rangle\, dt \leq C \left[T_1 - T_0 + \int_{T_0}^{T_1} g_\nu(t)\, dt\right] \leq C(T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}}, \tag{8.10}$$

where we used that $T$ is fixed.

Combining (8.7)-(8.10), we obtain, for $n \geq 2$,

$$\left\| \mathbf{E} \int_{\Delta_n(T_0,T_1)} \prod_{i=1}^n \nabla_{\alpha_i} f_i(t_i, X_{t_i}^{\cdot})\, dt_1 \cdots dt_n \right\|_2^2 \leq C K^{n-1} (T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}}.$$

After changing the constant $C$, this has the asserted form $C_0 K^n (T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}}$.

When $n = 1$, the same conclusion follows directly by applying the energy inequality to $\tilde{u}$ and using

$$\int_{T_0}^{T_1} e^F \langle \nabla_{\alpha_1} f_1, \tilde{u}\rangle\, dt = -\int_{T_0}^{T_1} e^F \langle f_1, \nabla_{\alpha_1}\tilde{u}\rangle\, dt,$$

followed by the preceding cutoff estimate for $f_1$. Taking $C_0$ large, if necessary, gives the sought estimate for $n = 1$.

2. Let us remove the assumption ($\mathrm{A}_b$). We repeat the preceding calculation in $L^2_\rho(\mathbb{R}^d)$. Applying the form-boundedness condition on $f_i$ to "test function" $\sqrt{\rho}u$ gives, for every fixed $\eta > 0$,

$$\langle f_i^2, u^2\rangle_\rho \leq (1+\eta)\nu\langle|\nabla u|^2\rangle_\rho + \big(g_\nu + C_\eta \nu\theta\big)\langle u^2\rangle_\rho,$$

and the analogous estimate holds for $b$ (with $\delta$ and $g_\delta$). Here we have used (1.7). The integration by parts in the diffusion term produces additional terms bounded by

$$\eta\langle|\nabla u|^2\rangle_\rho + C_\eta\theta\langle u^2\rangle_\rho.$$

In addition, integration by parts in $\langle(\nabla_{\alpha_k} f_k)u_{k+1}, u_k\rangle_\rho$ produces

$$-\int_{\mathbb{R}^d} f_k u_{k+1} u_k \, \nabla_{\alpha_k} \rho \, dx.$$

This term is estimated by

$$\begin{aligned}\left|\int_{\mathbb{R}^d} f_k u_{k+1} u_k \, \nabla_{\alpha_k} \rho \, dx\right| &\leq \eta \langle f_k^2, u_{k+1}^2\rangle_\rho + C_\eta \left\langle \frac{|\nabla \rho|^2}{\rho^2}, u_k^2 \right\rangle_\rho \\ &\leq \eta \langle f_k^2, u_{k+1}^2\rangle_\rho + C_\eta \theta \langle u_k^2\rangle_\rho.\end{aligned}$$

Thus, after first fixing $\eta$ and $\theta$ sufficiently small, all the additional terms are absorbed by the preceding argument, with $g_\delta + g_\nu$ in (8.6) replaced by $g_\delta + g_\nu + C\theta$.

Finally, using (8.1) with $\sqrt{\rho}$ as a test function,

$$\langle f_n^2(t)\rangle_\rho \leq \nu \langle |\nabla \sqrt{\rho}|^2\rangle + g_\nu(t)\langle \rho \rangle \leq C\nu\theta + C g_\nu(t).$$

It follows that

$$\int_{T_0}^{T_1} e^G \langle f_n^2(t)\rangle_\rho \, dt \leq C(T_1 - T_0)^{\frac{\varepsilon}{1+\varepsilon}}.$$

The weighted versions of (8.7)-(8.10) now give the asserted $L^2_\rho(\mathbb{R}^d)$ estimate. □

The next proposition, which is a consequence of Proposition 3′, verifies compactness of the Wiener-Sobolev space following [RZ]; the reader can compare it with Proposition 4 in the proof of Theorem 1.

**Proposition 4′.** *For every $r \geq 1$, there exist constants $K_1$, $K_2$ (independent of smoothness or boundedness of $b$) such that*

*(i)* $\sup_{y\in\mathbb{R}^d} \|\nabla X_t^x - I\|_{L^{2r}(B_1(y), L^r(\Omega))} \leq K_1 t^{\frac{1}{4r}\frac{\varepsilon}{1+\varepsilon}}$ *for all* $0 \leq t \leq T$;

*(ii)* $\sup_{y\in\mathbb{R}^d} \|D_s X_t^x - I\|_{L^{2r}(B_1(y), L^r(\Omega))} \leq K_1 (t-s)^{\frac{1}{4r}\frac{\varepsilon}{1+\varepsilon}}$ *for a.e.* $s \in [0,T]$ *and* $0 \leq s \leq t \leq T$;

*(iii)* $\sup_{y\in\mathbb{R}^d} \|D_s X_t^x - D_{s'} X_t^x\|_{L^{2r}(B_1(y), L^r(\Omega))} \leq K_2 |s-s'|^{\frac{1}{8r}\frac{\varepsilon}{1+\varepsilon}}$ *for a.e.* $s, s' \in [0,T]$ *and* $0 \leq s, s' \leq t \leq T$.

In detail, since $b$ is bounded and smooth, one has

$$\nabla X_t^x - I = \int_0^t \nabla b(s, X_s^x) \cdot \nabla X_s^x ds. \tag{8.11}$$

We iterate this identity,

$$\nabla X_t^x - I = \sum_{n=1}^\infty \int_{\Delta_n(t)} \prod_{i=1}^n \nabla b(t_i, X_{t_i}^x) dt_1 \ldots dt_n,$$

so

$$\|\nabla X_t^x - I\|_{L^{2r}(\mathbb{R}^d, L^r(\Omega))} \leq \sum_{n=1}^\infty \left\| \int_{\Delta_n(t)} \prod_{i=1}^n \nabla b(t_i, X_{t_i}^x) dt_1 \ldots dt_n \right\|_{L^{2r}(\mathbb{R}^d, L^r(\Omega))}. \tag{8.12}$$

It is at this step that we apply Proposition 3′ with $f_i$ taken to be the derivatives of the components of $b$. The Malliavin derivatives satisfy an SDE of the same type as (8.11) and are controlled in the same way using Proposition 3′. The remainder of the proof is identical to [KM1, Proof of Theorem 1] and follows closely the proof of Theorem 1.

## Appendix A. Example of $\{b_n\}$

In Theorem 2, we consider a bounded smooth approximation $\{b_m\}$ of form-bounded $b \in \mathbf{F}_{\delta,g_\delta}$ that satisfies

1) $\{b_m\}$ are form-bounded with the same $\delta$ and $\sup_m \|g_{\delta,m}\|_{L^{1+\varepsilon}(\mathbb{R})} < \infty$;

2)
$$b_m \to b \quad \text{in } [L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^d \quad \text{as } m \to \infty. \tag{A.1}$$

One straightforward way to construct $\{b_m\}$ is as follows.

Let $\{\gamma_\varepsilon\}$ denote the standard mollifier on $\mathbb{R}^{1+d}$, i.e. a family of positive smooth functions in $C_c^\infty(\mathbb{R}^{1+d})$ defined by
$$\gamma_\varepsilon(s,y) := \frac{1}{\varepsilon^{1+d}}\gamma(s/\varepsilon, y/\varepsilon), \quad (s,y) \in \mathbb{R}^{1+d}$$
with $\int_{\mathbb{R}^{1+d}} \gamma = 1$.

**Lemma A.1.** *The vector fields*
$$b_m := \gamma_{\varepsilon_m} \star b,$$
*where $\star$ is the convolution on $\mathbb{R}^{1+d}$ and $\varepsilon_m \downarrow 0$, satisfy 1), 2).*

**Remark 7.** It is easy to see that $\{b_m\}$ defined slightly differently,
$$b_m := \gamma_{\varepsilon_m} \star (\mathbf{1}_m b),$$
where $\mathbf{1}_m$ is the indicator function of the set $\{(t,x) \in \mathbb{R}^{1+d} \mid |t| \leq m, |x| \leq m, |b(t,x)| \leq m\}$, are form-bounded with form-bound slightly larger – but still independent of $m$ – than $\delta$ (not a problem since all our conditions on all form-bounds are strict inequalities), provided that we choose $\varepsilon_m \downarrow 0$ sufficiently rapidly. Indeed, we can write
$$b_m := \mathbf{1}_m b + \big(\gamma_{\varepsilon_m} \star \mathbf{1}_m b - \mathbf{1}_m b\big).$$
The first term in the RHS, i.e. vector field $\mathbf{1}_m b$, is obviously form-bounded with the same $\delta$ and even with the same function $g_\delta$. The second term is a difference between a bounded vector field with compact support and its mollification; it is not difficult to modify the first example in Section 2 dealing with the Lebesgue class drifts to show that, provided that $\varepsilon_m \downarrow 0$ sufficiently rapidly, the form-bound of $\gamma_{\varepsilon_m} \star \mathbf{1}_m b - \mathbf{1}_m b$ goes to zero as $m \to \infty$. Finally, we can use the fact that the sum of two form-bounded vector fields is a form-bounded vector field.

*Proof of Lemma A.1.* The boundedness and smoothness of $b_m$ follow from the standard properties of mollifiers.

1. Let us prove 1). For every $t \in \mathbb{R}$ and $\varphi \in C_c^\infty(\mathbb{R}^d)$,
$$\|b_m(t,\cdot)\varphi\|_2^2 \leq \int_{\mathbb{R}} \int_{\mathbb{R}^d} \gamma_{\varepsilon_m}(s,y) \int_{\mathbb{R}^d} |b(t-s,x-y)|^2 |\varphi(x)|^2 dx dy ds.$$
Hence
$$\|b_m(t,\cdot)\varphi\|_2^2 \leq \int_{\mathbb{R}} \int_{\mathbb{R}^d} \gamma_{\varepsilon_m}(s,y) \int_{\mathbb{R}^d} |b(t-s,z)|^2 |\varphi(z+y)|^2 dz dy ds.$$
Applying the form-boundedness of $b(t-s,\cdot)$ to $\varphi(\cdot+y)$ and using the translation-invariance of the $L^2$ norm, we obtain
$$\|b_m(t,\cdot)\varphi\|_2^2 \leq \delta\|\nabla\varphi\|_2^2 + g_{\delta,m}(t)\|\varphi\|_2^2, \tag{A.2}$$

where $g_{\delta,m}$ is the convolution of $t \mapsto \int_{\mathbb{R}^d} \gamma_{\varepsilon_m}(\cdot, y)dy$ and $t \mapsto g_\delta(t)$. Since the former is itself a mollifier (on $\mathbb{R}$), and the mollification preserves the $L^{1+\varepsilon}$ norm, we have

$$\|g_{\delta,m}\|_{L^{1+\varepsilon}(\mathbb{R})} \leq \|g_\delta\|_{L^{1+\varepsilon}(\mathbb{R})},$$

i.e. we have verified 1).

2. To verify 2), we first note that, by the standard properties of mollifiers,

$$b_m \to b \quad \text{in } [L^2_{\text{loc}}(\mathbb{R}^{1+d})]^d. \tag{A.3}$$

Our goal now is to upgrade this convergence to (A.1).

To this end, for each $R > 0$, define functions

$$\eta_R(y) := \xi_R(|y|), \qquad \xi_R(r) := \begin{cases} 0, & 0 \leq r < R, \\ r - R, & R \leq r \leq R+1, \\ 1, & r > R+1. \end{cases}$$

Then

$$\mathbf{1}_{\mathbb{R}^d \setminus B_{R+1}} \leq \eta_R^2 \qquad \text{and} \qquad |\nabla \eta_R| \leq \mathbf{1}_{B_{R+1} \setminus B_R}.$$

Next, it is easily seen that $b_m - b$ is form-bounded with parameters $4\delta$ and $2(g_{\delta,m} + g_\delta)$, where, by the argument in the previous step, $\sup_m \|2(g_{\delta,m} + g_\delta)\|_{L^1(\mathbb{R})} < \infty$ (note that we have $L^1$ here, not $L^{1+\varepsilon}$, although this is insignificant at this step). Therefore, using the form-boundedness of $b_m - b$ with test function $\eta_R\sqrt{\rho}$, we have

$$\int_0^T \int_{\mathbb{R}^d \setminus B_{R+1}} |b_m(t,y) - b(t,y)|^2 \rho(y)\, dydt \leq 4\delta T \big\|\nabla(\eta_R\sqrt{\rho})\big\|_2^2 + 2\int_0^T (g_{\delta,m}(t) + g_\delta(t))dt \big\|\eta_R\sqrt{\rho}\big\|_2^2.$$

Since $\rho$ is in $L^1(\mathbb{R}^d)$, vanishes at infinity and satisfies (1.7), we have $\big\|\eta_R\sqrt{\rho}\big\|_2 \to 0$ and $\big\|\nabla(\eta_R\sqrt{\rho})\big\|_2 \to 0$ as $R \to \infty$. Therefore,

$$\lim_{R\to\infty} \sup_m \int_0^T \int_{\mathbb{R}^d \setminus B_{R+1}} |b_m(t,y) - b(t,y)|^2 \rho(y)\, dy\, dt = 0. \tag{A.4}$$

On the other hand, for each $R > 0$, we have by (A.3)

$$\begin{aligned} &\int_0^T \int_{B_{R+1}} |b_m(t,y) - b(t,y)|^2 \rho(y)\, dy\, dt \\ &\leq \|\rho\|_{L^\infty(B_{R+1})} \int_0^T \int_{B_{R+1}} |b_m(t,y) - b(t,y)|^2\, dy\, dt \to 0 \end{aligned}$$

as $m \to \infty$. This, and the previous convergence, yield (A.1). $\square$

## Appendix B. Example of $\{\sigma_n\}$

Let $\sigma$ be as in Theorem 1, i.e. for every $t \in \mathbb{R}$,

1) $\sigma(t,\cdot) \in [L^\infty(\mathbb{R}^d)]^{d\times d}$,
2) $\sigma(t,\cdot)\sigma(t,\cdot)^\top \geq \kappa I$ a.e. on $\mathbb{R}^d$,
3) the form-boundedness condition (1.6) on the derivatives of $\sigma$ holds, i.e. for a.e. $t \in \mathbb{R}$

$$\||\nabla\sigma(t,\cdot)|\varphi\|_2^2 \leq \nu\|\nabla\varphi\|_2^2 + c_\nu\|\varphi\|_2^2 \quad \forall\, \varphi \in C_c^\infty(\mathbb{R}^d),$$

4) (**H**) holds for $\sigma$, i.e. $\sigma$ is locally uniformly Hölder continuous as a $[L^2_\rho(\mathbb{R}^d))]^{d\times d}$-valued function.

Let $\{\eta_\varepsilon\}$ denote the standard (spatial) mollifier on $\mathbb{R}^d$, i.e. for a fixed $0 \leq \eta \in C_c^\infty(B_1)$, $\int_{\mathbb{R}^d}\eta(x)dx = 1$, put $\eta_\varepsilon(x) := \varepsilon^{-d}\eta(x/\varepsilon)$.

**Lemma B.1.** *Assuming that the form-bound $\nu$ is sufficiently small, the sequence $\sigma^\varepsilon(t,\cdot) := \eta_\varepsilon * \sigma(t,\cdot)$, satisfies, for all $\varepsilon$ sufficiently small, uniformly in a.e. $t \in \mathbb{R}$,*

*1′) $\|\sigma^\varepsilon(t,\cdot)\|_{L^\infty(\mathbb{R}^d)} \leq \|\sigma(t,\cdot)\|_{L^\infty(\mathbb{R}^d)}$*

*2′) $\sigma^\varepsilon(t,\cdot)(\sigma^\varepsilon(t,\cdot))^\top \geq \frac{\kappa}{4}I$;*

*3′) (1.6) holds for $\sigma^\varepsilon$ with the same form-bound $\nu$ and constant $c_\nu$, i.e. independent of $\varepsilon$;*

*4′) (**H**) holds for $\sigma^\varepsilon$ with the same Hölder continuity exponent and constant.*

*5′)*

$$\|\sigma(t,\cdot) - \sigma^\varepsilon(t,\cdot)\|_{L^2_\rho(\mathbb{R}^d)} \leq C_1\varepsilon, \quad t \in \mathbb{R}.$$

*6′) Set $a = \sigma\sigma^\top$, $a^\varepsilon = \sigma^\varepsilon(\sigma^\varepsilon)^\top$. Then*

$$\nabla_k a^\varepsilon \to \nabla_k a \quad \text{ in } [L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^{d\times d} \text{ as } \varepsilon \downarrow 0$$

*for all $1 \leq k \leq d$. That is, we have componentwise convergence of the derivatives of $a^\varepsilon$, which in particular implies the convergence of the row-divergence operators $\nabla a^\varepsilon \to \nabla a$ in $[L^2_{\rm loc}(\mathbb{R}, L^2_\rho(\mathbb{R}^d))]^d$.*

*Proof.* 1′) is straightforward. Let us prove 3′). We have $\nabla\sigma^\varepsilon(t,\cdot) = \eta_\varepsilon * \nabla\sigma(t,\cdot)$. Therefore, by Cauchy-Schwarz,

$$|\nabla\sigma^\varepsilon(t,\cdot)|^2 \leq \eta_\varepsilon * |\nabla\sigma(t,\cdot)|^2.$$

Now, we verify form-boundedness of $|\nabla\sigma^\varepsilon(t,\cdot)|$ uniformly in $t$:

$$\begin{aligned}
\langle|\nabla\sigma^\varepsilon(t,\cdot)|^2, \varphi^2\rangle &\leq \langle\eta_\varepsilon * |\nabla\sigma(t,\cdot)|, \varphi^2\rangle \\
&= \langle|\nabla\sigma(t,\cdot)|^2, \eta_\varepsilon * \varphi^2\rangle \\
&\quad \text{(use form-boundedness (1.6) of } |\nabla\sigma(t,\cdot)|) \\
&\leq \nu\langle|\nabla\sqrt{\eta_\varepsilon * \varphi^2}|^2\rangle + c_\nu\langle\eta_\varepsilon * \varphi^2\rangle \\
&\quad \text{(Cauchy-Schwarz yields } |\nabla\sqrt{\eta_\varepsilon * \varphi^2}|^2 \leq \eta_\varepsilon * |\nabla\varphi|^2) \\
&\leq \nu\langle|\nabla\varphi|^2\rangle + c_\nu\langle\varphi^2\rangle,
\end{aligned}$$

as claimed (see e.g. [KS1, Sect. 4.4] for details, if needed).

Proof of 2′). We will use 3′), i.e. uniform in $\varepsilon$ form-boundedness of $|\nabla\sigma^\varepsilon|$. Let $t \in \mathbb{R}$. Our goal is to show that the minimal eigenvalue $\lambda_{\min}$ of $\sigma^\varepsilon(t,x)(\sigma^\varepsilon(t,x))^\top$ is greater or equal to $\frac{\kappa}{4}$. To this end, we first note that there exists $y \in B_\varepsilon(x)$ such that

$$|\sigma^\varepsilon(t,x) - \sigma(t,y)|_\circ \leq C\sqrt{\nu + c_\nu\varepsilon^2} \tag{B.1}$$

for a universal constant $C$, where $|A|_{\rm o} := \max_{v\in\mathbb{R}^d,|v|=1}|Av|$ (o stands for "operator norm"); the operator norm and the Euclidean norms of $A$ are equivalent, but it will be convenient to work with the operator norm. We prove (B.1) below. With (B.1) at hand, we can estimate

$$\sqrt{\lambda_{\min}(\sigma^\varepsilon(t,x)(\sigma^\varepsilon(t,x))^\top)} \geq \sqrt{\lambda_{\min}(\sigma(t,y)\sigma(t,y)^\top)} - |\sigma^\varepsilon(t,x)-\sigma(t,y)|_{\rm o},$$

which is, taking into account the variational characterization of eigenvalues of $\sigma\sigma^\top$, simply the inequality

$$\min_{|v|=1}|\sigma^\varepsilon(t,x)v| \geq \min_{|v|=1}|\sigma(t,y)v| - \max_{|v|=1}|(\sigma^\varepsilon(t,x)-\sigma(t,y))v|.$$

Therefore,

$$\begin{aligned}\sqrt{\lambda_{\min}(\sigma^\varepsilon(t,x)(\sigma^\varepsilon(t,x))^\top)} &\geq \sqrt{\kappa} - C\sqrt{\nu + c_\nu\varepsilon^2}\\ &\geq \frac{\sqrt{\kappa}}{2}\end{aligned}$$

provided $\nu$ is sufficiently small, starting with some small $\varepsilon$, as claimed.

It remains to prove (B.1). To this end, we compare $\sigma^\varepsilon(t,x)$ and $\sigma(t,y)$ to a common average: $\sigma_{B_\varepsilon(x)}(t) := \frac{1}{|B_\varepsilon(x)|}\int_{B_\varepsilon(x)}\sigma(t,z)dz$.

Step 1: We have

$$\begin{aligned}|\sigma_\varepsilon(t,x) - \sigma_{B_\varepsilon}(t)|_{\rm o}^2 &= \left|\int_{B_\varepsilon(x)}\eta_\varepsilon(x-z)(\sigma(t,z)-\sigma_{B_\varepsilon}(t))dz\right|_{\rm o}^2\\ &\leq \int_{B_\varepsilon(x)}\eta_\varepsilon(x-z)|\sigma(t,z)-\sigma_{B_\varepsilon}(t)|_{\rm o}^2dz\\ &\quad \text{(use } \|\eta_\varepsilon\|_\infty \leq c\varepsilon^{-d}\text{)}\\ &\leq c\varepsilon^{-d}\int_{B_\varepsilon(x)}|\sigma(t,z)-\sigma_{B_\varepsilon}(t)|^2dz\\ &\quad \text{(apply the Poincaré inequality)}\\ &\leq C\varepsilon^{2-d}\int_{B_\varepsilon(x)}|\nabla\sigma(t,z)|^2dz = C\varepsilon^{2-d}\int_{\mathbb{R}^d}\mathbf{1}_{B_\varepsilon(x)}|\nabla\sigma(t,z)|^2dz\\ &\quad \text{(apply form-boundedness of } |\nabla\sigma(t,\cdot)| \text{ after replacing } \mathbf{1}_{B_\varepsilon(x)}\\ &\quad \text{by non-negative smooth function } \varphi_\varepsilon = 1 \text{ on } B_\varepsilon(x),\ \varphi_\varepsilon = 0 \text{ on } \mathbb{R}^d - B_{2\varepsilon}(x))\\ &\leq C_1\varepsilon^{2-d}\big(\nu\varepsilon^{d-2} + c_\nu\varepsilon^d\big) = C_1\big(\nu + c_\nu\varepsilon^2\big).\end{aligned}$$

Step 2: Let us now compare $\sigma(t,y)$, for some choice of $y \in B_\varepsilon(x)$, and $\sigma_{B_\varepsilon}(t)$. By the estimates above,

$$\varepsilon^{-d}\int_{B_\varepsilon(x)}|\sigma(t,z)-\sigma_{B_\varepsilon}(t)|_{\rm o}^2dz \leq C\big(\nu + c_\nu\varepsilon^2\big).$$

Therefore,

$$\begin{aligned}|\{z \in B_\varepsilon(x) \mid |\sigma(t,z)-\sigma_{B_\varepsilon}(t)|_{\rm o}^2 > 2C\big(\nu+c_\nu\varepsilon^2\big)\}| &\leq \frac{1}{2C\big(\nu+c_\nu\varepsilon^2\big)}\int_{B_\varepsilon(x)}|\sigma(t,z)-\sigma_{B_\varepsilon}(t)|_{\rm o}^2dz\\ &\leq \frac{1}{2C\big(\nu+c_\nu\varepsilon^2\big)}\varepsilon^dC\big(\nu+c_\nu\varepsilon^2\big) = \frac{\varepsilon^d}{2}.\end{aligned}$$

Since the set of $x$ where the ellipticity condition $\sigma(t,x)\sigma(t,x)^\top \geq \kappa I$ is violated, if non-empty, has measure zero, by the previous estimate there exist (plenty of) $y \in B_\varepsilon(x)$ such that

$$|\sigma(t,y)-\sigma_{B_\varepsilon}(t)|_{\rm o}^2 \leq 2C\big(\nu + c_\nu \varepsilon^2\big).$$

Combining the estimates of Step 1 and Step 2, we arrive at (B.1), up to re-denoting $C$.

Proof of 5′). Fix $t \in \mathbb{R}$. We have

$$\begin{aligned}|\sigma(t,x)-\sigma^\varepsilon(t,x)|^2 &= \left|\int_{B_1} \eta(z)\big(\sigma(t,x)-\sigma(t,x-\varepsilon z)\big)\,dz\right|^2 \\ &\leq \int_{B_1} \eta(z)|\sigma(t,x)-\sigma(t,x-\varepsilon z)|^2\,dz, \end{aligned} \tag{B.2}$$

where, in turn, $\sigma(t,x)-\sigma(t,x-\varepsilon z) = \varepsilon \int_0^1 (\nabla \sigma)(t,x-s\varepsilon z)\, z ds$. Then

$$|\sigma(t,x)-\sigma(t,x-\varepsilon z)|^2 \leq \varepsilon^2 |z|^2 \int_0^1 |\nabla\sigma(t,x-s\varepsilon z)|^2 ds, \tag{B.3}$$

so, multiplying (B.2) by $\rho$, integrating over $\mathbb{R}^d$ and using (B.3), we obtain

$$\|\sigma(t,\cdot)-\sigma^\varepsilon(t,\cdot)\|^2_{L^2_\rho(\mathbb{R}^d)} \leq \varepsilon^2 \int_{B_1} \eta(z)|z|^2 \int_0^1 \int_{\mathbb{R}^d} \rho(x)|\nabla\sigma(t,x-s\varepsilon z)|^2 dxdsdz.$$

Making the change of variables $y = x - s\varepsilon z$ for fixed $z$, and using $|z| \leq 1$ and $0 \leq s\varepsilon \leq 1$, so that $|s\varepsilon z| \leq 1$, we have $\rho(y+s\varepsilon z) \leq C_d \rho(y)$. Hence we arrive at

$$\begin{aligned}\|\sigma(t,\cdot)-\sigma^\varepsilon(t,\cdot)\|^2_{L^2_\rho(\mathbb{R}^d)} &\leq C_d\varepsilon^2 \left(\int_{B_1} |z|^2\eta(z)\,dz\right) \int_{\mathbb{R}^d} \rho(y)|\nabla\sigma(t,y)|^2 dy \\ &\leq C\varepsilon^2 \|\nabla\sigma(t,\cdot)\|^2_{L^2_\rho(\mathbb{R}^d)}.\end{aligned}$$

It remains to apply the form-boundedness (1.6) of $|\nabla\sigma(t,\cdot)|$ and use the fact that both $\rho$ and $|\nabla\rho|$ are square integrable, to arrive at the claimed inequality $\|\sigma(t,\cdot)-\sigma^\varepsilon(t,\cdot)\|_{L^2_\rho(\mathbb{R}^d)} \leq C_1\varepsilon$.

Proof of 4′). Note first that, given a function (or matrix field) $h$ on $\mathbb{R}^d$, we have

$$\begin{aligned}\|\eta_\varepsilon * h\|^2_{L^2_\rho(\mathbb{R}^d)} &\leq C\int_{\mathbb{R}^d} \eta_\varepsilon(z)\int_{\mathbb{R}^d} \rho(x)|h(x-z)|^2 dxdz \\ &= C\int_{\mathbb{R}^d} \eta_\varepsilon(z)\int_{\mathbb{R}^d} \rho(y+z)|h(y)|^2 dydz \\ &\quad (\text{use that if } |z|\leq 1, \text{ then } \rho(y+z)\leq C_d\rho(y),\ y\in\mathbb{R}^d) \\ &\leq C\|h\|^2_{L^2_\rho(\mathbb{R}^d)}.\end{aligned}$$

Applying this inequality to $h(\cdot) = \sigma(t,\cdot)-\sigma(s,\cdot)$, we see that the locally uniform Hölder continuity of $\sigma$ in time as a $L^2_\rho$-valued function is preserved by the mollification in the spatial variables.

Proof of 6′). Fix $1 \leq k \leq d$. We need to establish convergence of $\nabla_k a^\varepsilon$ to $\nabla_k a$ in $L^2([t_0,t_1],L^2_\rho(\mathbb{R}^d))$. We assume that $t_0$, $t_1$ are fixed from now on. We have

$$\nabla_k a = (\nabla_k\sigma)\sigma^\top + \sigma(\nabla_k\sigma)^\top,$$

Similarly,

$$\nabla_k a^\varepsilon = (\nabla_k\sigma^\varepsilon)(\sigma^\varepsilon)^\top + \sigma^\varepsilon(\nabla_k\sigma^\varepsilon)^\top.$$

Subtracting these two identities from each other, we obtain

$$\begin{aligned}\nabla_k a^\varepsilon - \nabla_k a =&(\nabla_k\sigma^\varepsilon - \nabla_k\sigma)(\sigma^\varepsilon)^\top + (\nabla_k\sigma)(\sigma^\varepsilon-\sigma)^\top\\ &+(\sigma^\varepsilon-\sigma)(\nabla_k\sigma)^\top + \sigma^\varepsilon(\nabla_k\sigma^\varepsilon-\nabla_k\sigma)^\top.\end{aligned}$$

Therefore, using $\|\sigma^\varepsilon\|_\infty \leq \|\sigma\|_\infty$, we obtain

$$\|\nabla_k a^\varepsilon - \nabla_k a\|_{L^2([t_0,t_1],L^2_\rho)} \leq 2\|\sigma\|_\infty\|\nabla_k\sigma^\varepsilon - \nabla_k\sigma\|_{L^2([t_0,t_1],L^2_\rho)} + 2\big\||\nabla_k\sigma|\,|\sigma^\varepsilon-\sigma|\big\|_{L^2([t_0,t_1],L^2_\rho)}.$$

We can show that the first term in the RHS goes to zero as $\varepsilon \downarrow 0$ by adapting the proof of 2) in Lemma A.1. (Namely, if a sequence of uniformly in $\varepsilon$ form-bounded vector fields converges locally in $L^2$ – which, in our case, in a standard property of mollifiers – then we can upgrade this convergence to the global convergence with respect to weight $\rho$.) The second term goes to zero by $\sigma^\varepsilon(t,\cdot)\to\sigma(t,\cdot)$ a.e. on $\mathbb{R}^d$ (this is a standard property of mollifiers) for a.e. $t\in\mathbb{R}$, upon applying the Dominated convergence theorem. □

## Appendix C. Proof that $J$ is invertible (in the proof of the stochastic Gronwall claim 1)

Step 1. Let us make a preliminary observation regarding evaluating $f(J_t)$, where $f$ is a $C^2$ function of $d^2$ variables. The matrix identity $dJ_t = \sum_{l=1}^d A_t^l J_t dW_t^l$ yields scalar identities

$$dJ_t^{ij} = \sum_{l=1}^d (A_t^l J_t)^{ij} dW_t^l \tag{C.1}$$

The Itô's formula therefore yields

$$df(J_t) = \sum_{i,j=1}^d \partial_{ij} f(J_t) dJ_t^{ij} + \frac{1}{2}\sum_{i,j,k,r=1}^d \partial_{ij}\partial_{kr} f(J_t) d[J^{ij},J^{kr}]_t,$$

where the square brackets denote the cross-variation. Now, invoking identity (C.1) and its immediate consequence $d[J^{ij},J^{kr}]_t = \sum_{l=1}^d (A_t^l J_t)^{ij}(A_t^l J_t)^{kr}dt$, we obtain

$$df(J_t) = \sum_{l=1}^d\sum_{i,j=1}^d \partial_{ij} f(J_t)(A_t^l J_t)^{ij} dW_t^l + \frac{1}{2}\sum_{l=1}^d\sum_{i,j,k,r=1}^d \partial_{ij}\partial_{kr} f(J_t)(A_t^l J_t)^{ij}(A_t^l J_t)^{kr}dt.$$

It is convenient to rewrite the right-hand side in compact form in terms of the directional derivatives of $f$ ($\nabla$ is the gradient on $\mathbb{R}^{d\times d}$, so the direction is a $d\times d$ matrix):

$$df(J_t) = \sum_{l=1}^d \nabla_{A_t^l J_t} f(J_t) dW_t^l + \frac{1}{2}\sum_{l=1}^d \nabla^2_{A_t^l J_t, A_t^l J_t} f(J_t)dt. \tag{C.2}$$

Step 2. Now, armed with identity (C.2), we take $f(X) := \det X$. Our goal is to obtain, using (C.2), the SDE satisfied by $\det(J_t)$. The use of (C.2) thus requires evaluating directional derivatives of the determinant. These are given by well-known formulas obtained using Taylor polynomial expansion of $\det X$ as a function of $X\in\mathbb{R}^{d\times d}$:

$$\nabla_{BX}\det(X) = \det(X)\mathrm{tr}B, \qquad \nabla^2_{BX,BX}\det(X) = \det X\big((\mathrm{tr}\,B)^2 - \mathrm{tr}\,B^2\big).$$

Importantly, there is no $X$ under the trace sign: we arrive at the following *linear* SDE for $D_t := \det J_t$:

$$dD_t = D_t \sum_{l=1}^{d} \mathrm{tr}(A_t^l) dW_t^l + D_t \frac{1}{2} \sum_{l=1}^{d} \big((\mathrm{tr}\,(A_t^l))^2 - \mathrm{tr}\,(A_t^l)^2\big) dt, \quad D_s = 1.$$

We can solve it:

$$\det J_t = \exp\left( \sum_{l=1}^{d} \int_s^t \mathrm{tr}(A_r^l) dW_r^l - \frac{1}{2} \sum_{l=1}^{d} \int_s^t \mathrm{tr}\big((A_r^l)^2\big) dr \right).$$

Hence $\det J_t$ never vanishes.

## Appendix D. Proof of Claim 2

*Proof of Claim 2.* Fix $y \in \mathbb{R}^d$ and $0 \leq r < t \leq T$, and set $U := B_1(y)$, $F_m(x) := X_{r,t}^{x,m}$. The following argument works componentwise, so Lemma 3.1 applies equally to these vector-valued random fields. Recall that

$$X_{r,t}^{x,m} = x + \int_r^t \sigma_m(\tau, X_{r,\tau}^{x,m})\, dW_\tau, \qquad \sup_m \|\sigma_m\|_\infty \leq M.$$

We apply Proposition 4 with $p = 2$ to $X_{r,\cdot}^{x,m}$ – crucially, the constants $K_1$, $K_2$ in this proposition are independent not only of $m$ and the centre $y$, but also of the initial time $r$.

*Verification of* (3.1). By Itô's isometry,

$$\mathbf{E}|F_m(x)|^2 \leq 2|x|^2 + 2\mathbf{E} \int_r^t |\sigma_m(\tau, X_{r,\tau}^{x,m})|^2\, d\tau \leq 2|x|^2 + 2M^2 T.$$

Since $y$ is fixed,

$$\sup_m \mathbf{E} \int_U |F_m(x)|^2\, dx \leq C_{y,T}. \tag{D.1}$$

Proposition 4($i$) and the Cauchy-Schwarz inequality yield

$$\mathbf{E} \int_U |\nabla_x X_{r,t}^{x,m} - I|^2\, dx \leq |U|^{1/2} \|\nabla_x X_{r,t}^{\cdot,m} - I\|_{L^4(U,L^2(\Omega))}^2 \leq C|t-r|^{1/4}.$$

Hence

$$\sup_m \mathbf{E} \int_U |\nabla_x F_m(x)|^2\, dx \leq C_T. \tag{D.2}$$

Now, (D.1) and (D.2) yield (3.1).

*Verification of* (3.2). Let us write in what follows

$$G_m(s,x) := D_s F_m(x) \ \big( = D_s X_{r,t}^{x,m} \big).$$

Since $X_{r,t}^{x,m}$ depends only on the increments of the Brownian motion on $[r,t]$,

$$G_m(s,x) = 0 \qquad \text{for a.e. } s \notin [r,t]. \tag{D.3}$$

For $s \in [r,t]$, Proposition 4($ii$) gives

$$\|G_m(s,\cdot) - \sigma_m(s, X_{r,s}^{\cdot,m})\|_{L^4(U,L^2(\Omega))} \leq C|t-s|^{1/8}.$$

Since $\|\sigma_m(s, X_{r,s}^{\cdot,m})\|_{L^4(U,L^2(\Omega))} \leq M|U|^{1/4}$, we thus obtain

$$\sup_m \sup_{s \in [r,t]} \|G_m(s,\cdot)\|_{L^4(U,L^2(\Omega))} \leq C_T. \tag{D.4}$$

Hence, (D.3) and Cauchy-Schwarz, yield

$$\sup_m \mathbf{E}\int_U\int_0^T |D_sF_m(x)|^2 dsdx = \sup_m \int_r^t \mathbf{E}\int_U |G_m(s,x)|^2 dxds \le C_T$$

$\Rightarrow$ (3.2).

*Verification of* (3.3). Case 1: Let $s, s' \in [r,t]$. Then, by Proposition 4(*iii*),

$$\|G_m(s,\cdot) - G_m(s',\cdot)\|_{L^4(U,L^2(\Omega))} \le K_2\left(|s-s'|^{1/16} + |s-s'|^{\gamma/4}\right).$$

Put $\alpha := \frac{1}{16} \wedge \frac{\gamma}{4}$ and select some $0 < \beta < \alpha$. Since we work on a finite time interval, the previous estimate yields $\|G_m(s,\cdot) - G_m(s',\cdot)\|_{L^4(U,L^2(\Omega))} \le C_T|s-s'|^{\alpha}$. Thus, using the Cauchy-Schwarz inequality, we have

$$\mathbf{E}\int_U |G_m(s,x) - G_m(s',x)|^2\, dx \le C_T|s-s'|^{2\alpha}.$$

It follows that

$$\begin{aligned}
&\mathbf{E}\int_U\int_r^t\int_r^t \frac{|G_m(s,x) - G_m(s',x)|^2}{|s-s'|^{1+2\beta}} dsds'dx \\
&\le C_T\int_r^t\int_r^t |s-s'|^{-1+2(\alpha-\beta)} dsds' < \infty
\end{aligned}$$

since $\beta < \alpha$.

Case 2: Let $s \in [r,t]$ and $s' \in [0,T]\setminus[r,t]$. Then $G_m(s',\cdot) = 0$. Therefore, using (D.4), we can estimate

$$\begin{aligned}
\mathbf{E}\int_U\int_{s\in[r,t],s'\in[0,T]\setminus[r,t]} &\frac{|G_m(s,x) - G_m(s',x)|^2}{|s-s'|^{1+2\beta}} dsds'dx \\
&\le C_T\int_r^t\left(\int_0^r \frac{ds'}{(s-s')^{1+2\beta}} + \int_t^T \frac{ds'}{(s'-s)^{1+2\beta}}\right)ds \\
&\quad \text{(discard two negative terms coming from the endpoints } s'=0 \text{ and } s'=T) \\
&\le C'_{T,\beta}\int_r^t \left[(s-r)^{-2\beta} + (t-s)^{-2\beta}\right] ds \le C_{T,\beta}(t-r)^{1-2\beta},
\end{aligned}$$

where in the last integration we have used $\beta < \frac{1}{2}$.

Case 3: The case $s' \in [r,t]$ and $s \in [0,T]\setminus[r,t]$ gives the same contribution in the right-hand side of (3.3) as the Case 2.

Case 4: If both $s, s' \notin [r,t]$, then $G_m(s',\cdot) = G_m(s,\cdot) = 0$, so the contribution to the right-hand side of (3.3) is zero.

Cases 1-4 yield (3.3). This ends the proof of Claim 2. $\square$

## Appendix E. Kolmogorov–Chentsov theorems

For the reader's convenience, we state below Kunita's Kolmogorov–Chentsov-type theorems [Ku] that we have used in the proof of Theorem 1($i$).

**Theorem 3** ([Ku, Theorem 1.4.1]). *Let $X(x)$, $x \in D$, be a random field with values in a Banach space $B$, where $D$ is a domain in $\mathbb{R}^d$. Assume that there exist positive constants $\gamma$, $C$, and $\alpha_i$, $i = 1, \dots, d$ with $\sum_{i=1}^d \alpha_i^{-1} < 1$ satisfying*

$$\mathbf{E}[\|X(x) - X(y)\|^\gamma] \leq C \left( \sum_{i=1}^d |x^i - y^i|^{\alpha_i} \right), \quad x, y \in D.$$

*where, $\|\cdot\|$ stands for the norm of the Banach space and $x = (x^1, \dots, x^d)$, $y = (y^1, \dots, y^d)$. Then $X$ has a continuous modification.*

**Theorem 4** ([Ku, Theorem 1.4.7]). *Let $\{X_n(x) \mid x \in I = [0,1]^d\}$ be a sequence of continuous random fields with values in a real separable semi-reflexive Fréchet space $S$ with seminorms $\|\cdot\|_N$, $N = 1, 2, \dots$. Assume that, for each $N$, there exist positive constants $\gamma$, $C$, and $\alpha_1, \dots, \alpha_d$, with $\sum_{i=1}^d \alpha_i^{-1} < 1$, such that*

$$\mathbf{E}\left[\|X_n(x) - X_n(y)\|_N^\gamma\right] \leq C \left( \sum_{i=1}^d |x^i - y^i|^{\alpha_i} \right), \quad x, y \in I,$$

*and*

$$\mathbf{E}\left[\|X_n(x)\|_N^\gamma\right] \leq C, \quad x \in I,$$

*for every $n$. Then $\{X_n\}$ is tight with respect to the semi-weak topology of $C(I, S)$, i.e. the topology of the uniform convergence on $I$ with respect to the weak topology of $S$.*

## AI usage disclosure

The authors made use of OpenAI ChatGPT 5.6 for mathematical discussions related to the proof of Proposition 4. We also used ChatGPT for editorial assistance. The text of the article – including all em dashes – is written in its entirety by the authors.

## References


[A] D.R. Adams, Weighted nonlinear potential theory, Trans. Amer. Math. Soc., 297 (1986), 73-94.

[BS] V. Bally and B. Saussereau, A relative compactness criterion in Wiener–Sobolev spaces and application to semi-linear stochastic PDEs, J. Funct. Anal., 210(2) (2004), 465-515.

[B] R. Bass, Diffusions and Elliptic Operators, Springer, 1998.

[BFGM] L. Beck, F. Flandoli, M. Gubinelli and M. Maurelli, Stochastic ODEs and stochastic linear PDEs with critical drift: regularity, duality and uniqueness, Electron. J. Probab., 24 (2019), Paper No. 136, 72 pp (arXiv:1401.1530).

[BKRS] V.I. Bogachev, N.V. Krylov, M. Röckner and S.V. Shaposhnikov, Fokker-Planck-Kolmogorov Equations, Amer. Math. Soc., 2015.

[CK] E. Carlen and P. Krée, $L^p$ Estimates on iterated stochastic integrals, Ann. Probab., 19(1) 1991, 354-368.

[CWW] S.Y.A. Chang, J.M. Wilson and T.H. Wolff, Some weighted norm inequalities concerning the Schrödinger operator, Comment. Math. Helv. 60 (1985), 217-246.

[C] A. S. Cherny, On the uniqueness in law and the pathwise uniqueness for stochastic differential equations, Theory Probab. Appl., 46(3) (2002), 406-419.

[DB] R. D. Deblassie, Scale-invariant diffusions: transience and non-polar points, Bernoulli, 5 (1999), 589-614.

[D] H. Dong, Recent progress in the $L^p$ theory for elliptic and parabolic equations with discontinuous coefficients, Analysis in Theory and Applications, 36 (2) (2020), 161-199.

[FJ] N. Fournier and B. Jourdain, Stochastic particle approximation of the Keller-Segel equation and two-dimensional generalization of Bessel process, Ann. Appl. Probab. 27 (2017), 2807-2861.

[GC] A. Gabrielli and F. Cecconi, Clustering and coalescence from multiplicative noise: the Kraichnan ensemble, J. Phys. A: Math. Theor., 41(23) (2008), 235003.

[GV] K. Gawędzki and M. Vergassola, Phase transition in the passive scalar advection, Physica D 138 (2000), 63-90.

[GvC] A. Gulisashvili and J. A. van Casteren, Non-autonomous Kato Classes and Feynman-Kac Propagators, World Scientific, 2006.

[H] M. Hairer, Advanced Stochastic Analysis, Lecture notes, 2026.

[HL] T. Hoffmann-Ostenhof and A. Laptev, Hardy inequalities with homogeneous weights, J. Funct. Anal. 268 (2015), 3278-3289.

[K1] D. Kinzebulatov, Feller evolution families and parabolic equations with form-bounded vector fields, Osaka J. Math., 54 (2017), 499-516.

[K2] D. Kinzebulatov, Form-boundedness and SDEs with singular drift, INdAM Series, 56 (2024), 147-262.

[K3] D. Kinzebulatov, Parabolic equations and SDEs with time-inhomogeneous Morrey drift, NoDEA Nonlinear Differential Equations Appl., 32 (2025), Paper No. 117, 30 pp.

[K4] D. Kinzebulatov, On particle systems and critical strengths of general singular interactions, Ann. Inst. Henri Poincaré (B) Probab. Stat., 62 (2026), no. 2, 1191-1251.

[KM1] D. Kinzebulatov and K. R. Madou, Strong solutions of SDEs with singular (form-bounded) drift via Roeckner-Zhao approach, Stoch. Dyn., 24(7) (2024), 2450050.

[KM2] D. Kinzebulatov and K. R. Madou, Stochastic equations with time-dependent singular drift, J. Differential Equations, 337 (2022), 255-293.

[KS1] D. Kinzebulatov and Yu. A. Semënov, On the theory of the Kolmogorov operator in the spaces $L^p$ and $C_\infty$, Ann. Sc. Norm. Sup. Pisa (5) **21** (2020), 1573-1647 (arXiv:1709.08598).

[KS2] D. Kinzebulatov and Yu.A. Semënov, Remarks on parabolic Kolmogorov operator, Theory Probab. Appl. 70(2) (2025), 282-300.

[KS3] D. Kinzebulatov and Yu.A. Semënov, Feller generators with singular drifts in the critical range, J. Differential Equations, 433 (2025), Paper No. 113262, 30 pp.

[KS4] D. Kinzebulatov and Yu.A. Semënov, Sharp solvability for singular SDEs, Electron. J. Probab. 28 (2023), article no. 69, 1-15.

[KS5] D. Kinzebulatov and Yu.A. Semënov, Feller generators and stochastic differential equations with singular (form-bounded) drift, Osaka J. Math. 58 (2021), 855-883 (arXiv:1904.01268).

[KS6] D. Kinzebulatov and Yu.A. Semënov, Brownian motion with general drift, Stoch. Proc. Appl., 130 (2020), 2737-2750 (arXiv:1710.06729).

[KSS] D. Kinzebulatov, Yu. A. Semënov and R. Song, Stochastic transport equation with singular drift, Ann. Inst. Henri Poincaré (B) Probab. Stat., 60(1) (2024), 731-752 (arXiv:2102.10610).

[KV] D. Kinzebulatov and R. Vafadar, SDEs with critical general distributional drifts: sharp solvability and blow ups, arXiv:2506.09244.

[KY] D. Kinzebulatov and C. Yue, Non-local SDEs, critical drifts and local blow-ups, arXiv:2606.02786.

[KPS] V. F. Kovalenko, M. A. Perelmuter and Yu. A. Semënov, Schrödinger operators with $L^{1/2}_W(R^l)$-potentials, J. Math. Phys., 22 (1981), 1033-1044.

[KS] V. F. Kovalenko and Yu. A. Semënov, $C_0$-semigroups in $L^p(\mathbb{R}^d)$ and $C^\infty(\mathbb{R}^d)$ spaces generated by differential expression $\Delta + b \cdot \nabla$, Theory Probab. Appl. 35 (1990), 443-453.

[Ku] H. Kunita, Stochastic Flows and Stochastic Differential Equations, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1990.

[Kr1] N. V. Krylov, On strong solutions of Itô's equations with $D\sigma$ and $b$ in Morrey classes containing $L_d$, Ann. Probab. 51 (2023), no. 5, 1729-1751.

[Kr2] N. V. Krylov, On strong solutions of time inhomogeneous Itô's equations with Morrey diffusion gradient and drift. A supercritical case, Stoch. Proc. Appl. 185 (2025), Article 104619, 20 pp.

[Kr3] N. V. Krylov, Stochastic Itô Equations and Parabolic Second-Order Equations with Singular Drift, 198 pp, arXiv:2605.03236.

[KrR] N. V. Krylov and M. Röckner, Strong solutions of stochastic equations with singular time dependent drift, Probab. Theory Related Fields 131 (2005), no. 2, 154-196.

[KrV] N.V. Krylov and A.Yu. Veretennikov, On explicit formulas for solutions of stochastic equations, Math. USSR-Sb. 29 (1976), no. 2, 239-256.

[KrZ] N.V. Krylov and A.K. Zvonkin, On strong solutions of stochastic differential equations (Russian), Proceedings of the School and Seminar on the Theory of Random Processes (Druskininkai, 1974), Part II, pp. 9-88 Inst. Fiz. i Mat. Akad. Nauk Litovsk. SSR, Vilnius, 1975. English translation in Sel. Math. Sov. 1(1) (1981), 19-61.

[MNS] G. Metafune, L. Negro and C. Spina, Sharp kernel estimates for elliptic operators with second-order discontinuous coefficients, J. Evol. Equ. 18 (2018), 467-514.

[MP] T. Meyer-Brandis and F. Proske, Construction of strong solutions of SDEs via Malliavin calculus, J. Funct. Anal., 258(11) (2010), 3922-3953.

[MNP] A. Mohammed, T. Nilssen and F. Proske, Sobolev differentiable stochastic flows for SDEs with singular coefficients: applications to the transport equation, Ann. Probab., 43(3) (2015), 1535-1576.

[N] D. Nualart, The Malliavin Calculus and Related Topics, Springer, Berlin, 2006.

[R] F. Rezakhanlou, Regular flows for diffusions with rough drifts, arXiv:1405.5856.

[RZ] M. Röckner and G. Zhao, SDEs with critical time dependent drifts: strong solutions. Probab. Theory Related Fields. 192 (2025), 1071-1111.

[T] Y. Tardy, Weak convergence of the empirical measure for the Keller-Segel model in both subcritical and critical cases, Electron. J. Probab., 29 (2024), Paper No. 142, 1-35.

[Z] X. Zhang, Stochastic differential equations with Sobolev coefficients and applications, Ann. Appl. Probab. 26 (2016), 2697-2732.

Université Laval, Département de mathématiques et de statistique, Québec, Canada and Laboratoire de Mathématiques Appliquées, Université Ferhat Abbas, Sétif 1, Campus El Bez, Sétif, Algeria

*Email address*: `sallah-eddine.boutiah.1@ulaval.ca`

Université Laval, Département de mathématiques et de statistique, Québec, Canada

*Email address*: `damir.kinzebulatov@mat.ulaval.ca`