STOCHASTIC PROCESSES WITH BOUNDED AND VANISHING MEAN OSCILLATION

KHOA LÊ

ABSTRACT. While functions and martingales of bounded mean oscillation (BMO) are well-studied due to the prominent works of John–Nirenberg (1961), Fefferman (1971) and Getoor–Sharpe (1972), BMO processes had not been introduced with one exception of Garisa (1973) who considered BMO sequences. Stochastic processes of bounded mean oscillation have appeared previously in various publications in probability theory however without a proper identity. Our main purpose is to identify the defining characters of such processes and establish their exponential integrability. In addition, we introduce a subclass of stochastic processes of vanishing mean oscillation (VMO), which comes with quantitative moment estimates. Some applications in rough stochastic differential equations, stochastic numerics and regularization by noise are discussed.

MATHEMATICS SUBJECT CLASSIFICATION (2020): Primary 60G07, Secondary 60H50, 60H35, 60H10.

KEYWORDS: BMO processes; VMO processes; BMO martingales; John–Nirenberg inequality; regularization by noise

1. INTRODUCTION

A real-valued locally integrable function \( f \) defined on \( \mathbb{R}^d \) is of bounded mean oscillation (BMO) if

\[
\sup_Q \frac{1}{|Q|} \int_Q |f(x) - f_Q|dx < \infty
\]

where the supremum is taken over all cubes \( Q \) in \( \mathbb{R}^d \), \( |Q| \) denotes the Lebesgue measure of \( Q \) and \( f_Q = \frac{1}{|Q|} \int_Q f(x)dx \). This definition is due to John and Nirenberg in [JN61], in which they obtained an exponential estimate for BMO functions known nowadays as John–Nirenberg inequality. Fefferman later in [Fef71] identified the space of BMO functions as the dual of the Hardy space \( H_1 \). Getoor and Sharpe introduced among other things in [GS72] continuous time BMO martingales and established the duality of \( H_1 \) and BMO martingales, which is a probabilistic analogue of Fefferman’s earlier result. This duality for discrete time martingales is due to Fefferman and Stein [FS72], Garsia and Herz [Gar73a, Her74]. The stochastic exponential of a BMO martingale is a uniformly integrable martingale. This is one of many reasons which make BMO martingales directly related to many problems in probability, see [Kaz94] and [BN08, DMS+97, Gei05] for further applications in financial mathematics.
Recent progress in rough stochastic differential equations and regularization by noise leads naturally to stochastic processes of bounded mean oscillation. To give the precise definition, let \((\Omega, \mathcal{G}, \mathbb{P})\) be a probability space equipped with a filtration \(\{\mathcal{F}_t\}_{t \geq 0}\) satisfying the usual conditions. Let \(\tau > 0\) be a fixed number. For each stopping time \(S\), \(\mathbb{E}_S\) denotes the conditional expectation with respect to \(\mathcal{F}_S\) and for each \(G \in \mathcal{G}\), \(\mathbb{P}_S(G) := \mathbb{E}_S(1_G)\). For each \(m \in [1, \infty]\), \(\| \cdot \|_m\) denotes the norm in \(L^m(\Omega, \mathcal{G}, \mathbb{P})\).

**Definition 1.1.** Let \((V_t)_{t \in [0, \tau]}\) be a real valued adapted right continuous process with left limits (RCLL). \(V\) is of bounded mean oscillation (BMO) if

\[
[V]_{\text{BMO}} := \sup_{0 \leq S \leq T \leq \tau} \| \mathbb{E}_S[V_T - V_S] \|_\infty < \infty
\]

where the supremum is taken over all stopping times \(S, T\) and we set \(V_{0-} = V_0\) by convention.

It is straightforward to see that the class of BMO processes starting from 0 equipped with the (semi-)norm \([\cdot]\)_{\text{BMO}} is a Banach space. If one formally takes \(\tau = \infty\) and assumes additionally that \(V\) satisfies the martingale property, the above definition defines precisely continuous time BMO martingales over infinite time horizon ([Kaz94, GS72]). **Definition 1.1** is also motivated from Garsia’s definition of BMO sequences in [Gar73b]. When \(V\) is continuous, the supremum in (1.1) can be taken over deterministic times \(S, T\) without changing the BMO property, see **Proposition 2.2** below. Examples of BMO processes are given.

**Example 1.2.** (a) Let \((X_t^x)_{t \geq 0, x \in \mathbb{R}^d}\) be a Markov process and let \(f : [0, \tau] \times \mathbb{R}^d \to \mathbb{R}\) be a measurable function. Suppose that one has the uniform Krylov estimate

\[
\sup_{0 \leq s \leq t \leq \tau} \sup_{x \in \mathbb{R}^d} \left| \int_s^t f(r, X_{r-}^x) dr \right| \leq C
\]

for some finite constant \(C\). If furthermore the process \(K_t := \int_0^t f(r, X_r^0) dr\) is a.s. continuous then it is BMO. (Indeed, by Markov property, \(\mathbb{E}_s[K_t - K_s] = \| \int_s^t f(r, X_{r-}^x) dr \|_{x=X_0^x} \leq C\) so that \(K\) is BMO by **Proposition 2.2** herein.) Krylov estimate is an important tool in the study of stochastic differential equations (SDEs), see for instance [RZ21, Zha16].

(b) Let \(f : [0, 1] \times \mathbb{R}^d \to \mathbb{R}\) be a Borel function which is uniformly bounded by 1 and \(B\) be a standard Brownian motion in \(\mathbb{R}^d\). Define for each integer \(n \geq 1\), the process \(V^n_t = \int_0^t [f(r, B_r) - f(r, B_{[nt]/n})] dr, t \in [0, 1]\), which corresponds to the error of the quadrature rule approximating the functional \(\int_0^t f(r, B_r) dr\). Then, \((V^n_t)_{t \in [0, 1]}\) is BMO with \([V^n]_{\text{BMO}} \leq (N \log(n+1)/n)^{1/2}\) for some finite constant \(N\). Indeed, it is shown by Dareiotis and Gerencsér in [DG20] (see Lemma 2.1 therein) that

\[
\text{ess sup}_{\omega} (\mathbb{E}_\omega[V^n_t - V^n_s]^2)^{1/2} \leq (N \log(n+1)/n)^{1/2} \text{ for every } s \leq t \leq 1.
\]

Without the conditional expectation, (1.3) is also obtained by Altmeyer in [Alt21]. The authors of both works rely on explicit moment computations and therefore are restricted to the second moment. Quadrature error estimates such as the above are directly related to the strong
convergence rate of the Euler–Maruyama scheme for SDEs. It is of important interests to have quadrature error estimates for all moments. We will revisit these processes in Example 2.4.

BMO processes appeared in [KR05, RZ21, Zha16] on singular diffusions, in [Dav07, ABLM20, Lê22, GG22] on regularization by noise phenomenon, in [DG20, LL21, DGL22] on strong convergence rate of numerical methods for stochastic differential equations, in [FHL21] on rough stochastic differential equations. It is surprising that in all of these occurrences, BMO property was not identified and the connection with John–Nirenberg inequality was not established. We will revisit some of these problems later in Section 4. The purpose is to demonstrate that knowledge of BMO processes effectively shortens proofs and improves results.

Stochastic processes with vanishing mean oscillation form an important subclass of BMO processes. In most of the aforementioned references, whenever a process is realized as a BMO process, it is actually VMO.

Our main focus in the current article is the John–Nirenberg inequality for BMO and VMO stochastic processes. Our work differs from the previous ones on the topic in the following ways:

(i) we work with general RCLL stochastic processes instead of martingales;
(ii) we consider processes over a finite time interval $[0, \tau]$ instead of the whole non-negative axis $[0, \infty)$.

This setup is well motivated by the problems considered in the aforementioned references. The restriction over finite time interval does bring out some new properties of BMO processes even when restricted to BMO martingales. For instance, we establish a connection between John–Nirenberg inequality and Khasminskii’s lemma and consequently, obtain better exponential estimates, see Theorem 2.3 and Remark 2.7 below. When the modulus of mean oscillation can be quantified, we obtain quantitative exponential estimates, see Theorem 3.2 and Corollary 3.3.

Remark 1.3. Herein, we focus on estimations for moments of BMO processes and for this problem, there is no loss of generality when restricting to real valued processes. Indeed, if $Z$ is an adapted RCLL process taking values in some metric space $(E, d)$ such that $\sup_{0 \leq t \leq \tau} \|\mathbb{E} d(V_{s}, V_{t})\|_{\infty} < \infty$, then the processes $Z_{t} = d(V_{0}, V_{t})$ is a real valued BMO process as of Definition 1.1. This is an immediate consequence of the triangle inequality. In fact, the maximal process $Z_{t}^{*} = \sup_{s \leq t} |Z_{s}|$ is also of BMO, see Proposition 2.6 below.

Let us now detail our plan. In Section 2, we establish John–Nirenberg inequality for BMO processes and its connection with the Khasminskii’s lemma. These results are applied to VMO processes in Section 3. We observe here an interesting connection with Lyons’ estimates for multiplicative functionals ([Lyo98]). In Section 4, we discuss three applications of John–Nirenberg inequality for BMO and VMO stochastic processes. The first one is about exponential moments of a class of stochastic controlled rough paths considered in [FHL21]. In the second application, we explain that Davie’s exponential estimate ([Dav07]) can be simplified and deduced from a single estimate for the second moment. The last application is about strong convergence rate of the tamed Euler–Maruyama scheme for SDEs with integrable drifts. Such rate has been obtained previously in [LL21] however only for small moments. Here, we are able to derive the
same rate for all moments using John–Nirenberg inequality in conjunction with the recent
stability estimate of Galeati and Ling in [GL22]. The appendix contains two auxiliary results
which are well-known but are adjusted to our setting.

2. BMO processes

We recall Definition 1.1 of BMO processes. For each BMO process \( V \), we define its modulus
of mean oscillation \( \rho(V) : \{ (s, t) \in [0, \tau]^2 : s \leq t \} \to [0, \infty) \) by

\[
\rho_{s,t}(V) = \sup_{s \leq S \leq T \leq t} \| \mathbb{E}_S[V_T - V_{S-}] \|_{\infty}, \quad 0 \leq s \leq t \leq \tau,
\]

where the supremum is taken over all stopping times \( S, T \) satisfying \( s \leq S \leq T \leq t \). The
function \( \rho(V) \) is monotone in the sense that

\[
\rho_{u,v}(V) \leq \rho_{s,t}(V) \text{ whenever } [u,v] \subset [s,t]. \tag{2.1}
\]

We define the function \( \kappa(V) : \{ (s, t) \in [0, \tau]^2 : s \leq t \} \to [0, \infty) \) by the relation

\[
\kappa_{s,t}(V) = \lim_{h \downarrow 0} \sup_{s \leq u \leq S \leq t, v-h \leq u \leq v} \rho_{u,v}(V). \tag{2.2}
\]

It is easy to see that the above limit always exists, that \( \kappa(V) \) is monotone and \( \kappa_{s,t}(V) \in [0, \rho_{s,t}(V)] \) for each \( s \leq t \).

An immediate consequence of (1.1) is that all jumps of a BMO process are uniformly bounded
by \( \kappa \).

**Proposition 2.1.** If \( V \) is BMO then \( \sup_{t \in [0, \tau]} \| V_t - V_{t-} \|_{\infty} \leq \kappa_{0,\tau}(V) \).

**Proof.** By definitions, \( \| \mathbb{E}_S[V_T - V_{S-}] \|_{\infty} \leq \rho_{0,\tau}(V) \) for all stopping times in \([0, \tau]\). Taking \( T = S \) yields that \( \sup_S | V_S - V_{S-} |(\omega) \leq \rho_{0,\tau}(V) \) for all \( \omega \) in a set of full probability. For such \( \omega \) and for
each \( r \in [0, \tau] \) such that \( V_r(\omega) \neq V_{r-}(\omega) \), there is a stopping time \( S \) such that \( S(\omega) = r \) (Prop.
2.26 [KS91]). With this stopping time, one deduces that \( | V_r - V_{r-} |(\omega) \leq \rho_{0,\tau}(V) \). The argument
can be applied over arbitrary sub-intervals of \([0, \tau]\). Hence, for all \( \omega \) in a set of full probability,
for every rationals \( s \leq t \) in \([0, \tau]\) and every \( r \in [s, t] \), we have \( | V_r - V_{r-} |(\omega) \leq \rho_{s,t}(V) \). This
implies the result. \( \square \)

If \( V \) is RCLL and has uniformly bounded jumps, one can replace the stopping times in (1.1)
by deterministic times.

**Proposition 2.2.** Let \( V \) be an adapted RCLL process and assume that

\[
\sup_{s \leq t \leq \tau} \| \mathbb{E}_S[V_t - V_{s-}] \|_{\infty} \leq B \quad \text{and} \quad \sup_{r \in [0, \tau]} | V_r - V_{r-} | \leq C
\]

for some finite constants \( B, C \). Then for every stopping times \( S \leq T \leq \tau \), one has

\[
\mathbb{E}_S[V_T - V_{S-}] \leq 2B + 3C. \tag{2.3}
\]

Consequently, \( V \) is BMO.
**Theorem 2.3** (John–Nirenberg inequality). Let $V$ be a BMO process and let $r$ be a fixed number in $[0, \tau]$. Then

\[ \| \mathbb{E}_r \sup_{t \leq \tau} |V_t - V_r|^p \|_\infty \leq p!(11\rho_{r,\tau}(V))^p \text{ for every integer } p \geq 1, \quad (2.5) \]

and

\[ \mathbb{E}_r e^{\lambda \sup_{t \leq \tau} |V_t - V_r|} < \infty \text{ for every } \lambda < (11\kappa_{r,\tau}(V))^{-1}. \quad (2.6) \]

We revisit the example from the introduction and discuss the implication of the previous theorem.

**Example 2.4.** We recall **Example 1.2**.

(a) From **Theorem 2.3**, we have

\[ \text{ess sup}_\omega \mathbb{E}_s \left( \sup_{t \in [s, \tau]} \left| \int_s^t f(r, X_r^0)dr \right|^p \right) \leq p!(11C)^p \quad (2.7) \]

for every integer $p \geq 1$ and every $s \in [0, \tau]$. The John–Nirenberg inequality for BMO processes can thus be considered as a passage from uniform Krylov estimate to moment estimates of all orders. Such passage has been known previously only when $f$ is non-negative through Khasminskii’s lemma or when the modulus of mean oscillation is sufficiently regular through the stochastic sewing lemma from [Lê20] (see also Remark 5.3 in [ABLM20]).
(b) Theorem 2.3 implies the following new estimate without any book-keeping calculations of high moments

\[ \text{ess sup}_{\omega} \left( \mathbb{E}_s \sup_{t \leq s \leq 1} \left| \int_s^t [f(r, B_r) - f(r, B_{[nr]/n})] \, dr \right|^p \right) \leq p!(121N \log(n + 1)/n)^\beta \]  

(2.8)

for every integer \( p \geq 1 \) and every \( s \leq 1 \).

Later in Theorem 3.2, we will see that when the modulus of mean oscillation can be quantified, one can improve the growth constant \( p! \) in (2.7) and (2.8).

The proof of Theorem 2.3 relies on the following two intermediate results.

**Proposition 2.5.** Let \((A_t)_{t \in [0, \tau]}\) be a BMO process which is non-decreasing. Then for every \( r \in [0, \tau] \) and every \( \lambda < (\kappa_{r, \tau}(A))^{-1} \), there is a finite constant \( c = c(\lambda, \rho(A)|_{[r, \tau]}^2) \) such that \( \mathbb{E}_r e^{\lambda(A_r - A_{A_r})} \leq c \).

**Proof.** It suffices to show the result for \( r = 0 \). By assumption, \( \|\mathbb{E}_S(A_t - A_{S-})\|_\infty \leq \rho_{s,t}(A) \) for every times \( s \leq t \leq \tau \) and every stopping time \( s \leq S \leq t \). We apply the energy inequality, Lemma A.2, to obtain that \( \mathbb{E}_s[(A_r - A_t)^p] \leq p!(\rho_{s,t}(A))^p \) for every \( s \leq t \leq \tau \) and every integer \( p \geq 1 \). For each \( \lambda < (\kappa_{0,\tau}(A))^{-1} \), there is an \( h_0 > 0 \) such that \( \lambda \rho_{s,t}(A) < 1 \) whenever \( t - s \leq h_0 \). For such \( s, t \), we have by Taylor’s expansion that

\[ \|\mathbb{E}_s e^{\lambda(A_r - A_{A_r})}\|_\infty \leq (1 - \lambda \rho_{s,t}(A))^{-1} < \infty. \]

Now partition \([0, \tau]\) by points \( 0 = t_0 < t_1 < \ldots < t_n = \tau \) so that \( \max_{1 \leq k \leq n}(t_k - t_{k-1}) \leq h_0 \). Then

\[ \mathbb{E}_r e^{\lambda(A_{A_r} - A_r)} = \mathbb{E}_r e^{\lambda(A_{t_0} - A_{A_r})} e^{\lambda(A_{t_0} - A_{A_{t_0}})} = \mathbb{E}_r e^{\lambda(A_{t_0} - A_{A_{t_0}})} \mathbb{E}_{t_0} e^{\lambda(A_{t_0} - A_{A_{t_0}})} \]

\[ \leq \mathbb{E}_r e^{\lambda(A_{t_0} - A_{A_r})} \|\mathbb{E}_{t_0} e^{\lambda(A_{t_0} - A_{A_{t_0}})}\|_\infty. \]

Iterating the previous inequality yields

\[ \|\mathbb{E}_r e^{\lambda(A_{A_r} - A_r)}\|_\infty \leq \|\mathbb{E}_r e^{\lambda(A_{j} - A_{A_j})}\|_\infty \prod_{k=j+1}^n \|\mathbb{E}_{t_{k-1}} e^{\lambda(A_{t_{k-1}} - A_{A_{t_{k-1}}})}\|_\infty, \]

where \( j \) is such that \( t_{j-1} \leq r < t_j \). This implies the bound

\[ \|\mathbb{E}_r e^{\lambda(A_{A_r} - A_r)}\|_\infty \leq \prod_{k=1}^n (1 - \lambda \rho_{t_{k-1}, t_k}(A))^{-1}, \]

which yields the result. \( \square \)

**Proposition 2.6.** Let \( V \) be a BMO process. Define \( V_t^* = \sup_{s \leq t} |V_s - V_0| \). Then \( V^* \) is BMO with \( \rho(V^*) \leq 11\rho(V) \).

**Proof.** Fix \( s \leq t \). For stopping times \( s \leq S \leq T \leq t \), we have

\[ \mathbb{E}_S |V_T - V_{S-}| \leq c \text{ with } c = \rho_{s,t}(V). \]
We define \( D_{s,t} = \sup_{s \leq r \leq t} |V_r - V_s| \) and apply Lemma A.1 to obtain that
\[
\beta \mathbb{P}_s((D_{s,t} - \beta)^+ \geq \alpha) \leq \beta \mathbb{P}_s(D_{s,t} \geq \alpha + \beta) \leq \rho_{s,t}(V) \mathbb{P}_s(D_{s,t} \geq \alpha)
\]
for every \( \alpha, \beta > 0 \). It follows that \( \beta \mathbb{E}_s((D_{s,t} - \beta)^+) \leq c \mathbb{E}_s D_{s,t} \). Choosing \( \beta = 2c \), we have
\[
2c \mathbb{E}_s(D_{s,t} - 2c) \leq c \mathbb{E}_s D_{s,t},
\]
that is \( \mathbb{E}_s \sup_{s \leq r \leq t} |V_r - V_s| \leq 4 \rho_{s,t}(V) \). Combining with the elementary estimate \( V_t^* - V_s^- \leq \sup_{s \leq r \leq t} |V_r - V_s^-| \), we obtain that
\[
\mathbb{E}_s |V_t^* - V_s^-| \leq 4 \rho_{s,t}(V).
\]
We also have \( V_t^* - V_t^- \leq V_t - V_t^- \) for every \( t \leq \tau \) and hence an application of Proposition 2.2 yields the result. \( \square \)

**Proof of Theorem 2.3.** It suffices to show the result for \( r = 0 \). Define \( A_r = \sup_{s \in [0,t]} |V_s - V_0| \). By Proposition 2.6, \( A \) is BMO and \( \rho(A) \leq 11 \rho(V) \). We obtain (2.5) and (2.6) by applying Lemma A.2 and Proposition 2.5 respectively. \( \square \)

When \( A_t = \int_0^t \beta(r) dr \) for a non-negative process \( \beta \), Proposition 2.5 deduces to a general form of Khasminskii’s lemma (Chapter I.1 [Por90]). We note that in earlier versions of Khasminskii’s lemma ([Kha59, AS82]), a smallness condition was imposed. The relation between the exponential constant \( \lambda \) and the function \( \kappa \) goes back at least to Portenko [Por90]. The John–Nirenberg inequality for discrete martingales originated from [JN61] for a specific probability space. Most of the literatures focuses on BMO martingales ([GS72, Kaz94]) with an exception of Garsia and Stroock who considered BMO sequences in [Gar73b, Str73] from which Definition 1.1 is inspired. We note that in the case of discrete time, stopping times are not needed.

**Remark 2.7.** In all the aforementioned references, BMO processes are defined over the whole positive axis while the relation between the exponential exponent \( \lambda \) with the function \( \kappa \) is a unique feature for BMO processes on finite time domains. To be more precise, we recall that a continuous martingale \((X_t)_{t \geq 0}\) is BMO if
\[
[X]_{BMO}^2 := \sup_S \|\mathbb{E}_S|X_\infty - X_S|^2\|_{\infty} < \infty,
\]
where the supremum is taken over all stopping times \( S \). Osekowski shows in [Ose15] that the inequality
\[
\mathbb{E}e^{\lambda \sup_{t \geq 0} |X_t - X_0|} \leq \int_0^\infty e^{\lambda |X|_{BMO}^2} e^{-\lambda t} dt \quad (\lambda > 0)
\]  
(2.9)
is true and sharp, i.e. there is a martingale \( X \) with \( 0 < |X|_{BMO} < \infty \) for which both sides are equal. One sees that \( \kappa(X) \) plays no role whatsoever for BMO processes over infinite time horizon. In the other direction, let \((V_t)_{t \in [0,T]}\) be a continuous BMO martingale and define \( X_t = V_{t \wedge T} \). Then \((X_t)_{t \geq 0}\) is a continuous BMO martingale on the whole positive axis and is subjected to (2.9). In particular, one sees that \( \mathbb{E}e^{\lambda \sup_{t \in [0,T]} |V_t - V_0|} \) is finite if \( \lambda |V_{T \wedge T}|_{BMO} < 1 \). Since \( |V_{T \wedge T}|_{BMO} \) is comparable to \( \rho_{[0,T]}(V) \) (by John–Nirenberg inequality), we deduce that \( \mathbb{E}e^{\lambda \sup_{t \in [0,T]} |V_t - V_0|} \) is finite if \( \lambda \rho_{[0,T]}(V) < c \) for some universal constant \( c \). This is much more restrictive than the
condition \( \lambda k_{[0,\tau]}(V) < 11 \) provided by Theorem 2.3, especially for processes of vanishing mean oscillation (cf. Theorem 3.2).

3. VMO processes

Definition 3.1. A BMO process \((V_t)_{t \in [0,\tau]}\) is VMO if \( \kappa_{0,\tau}(V) = 0 \).

An immediate consequence of Proposition 2.1 is that every VMO process has continuous sample paths. It is straightforward to see that the class of VMO processes starting from 0 forms a closed subspace of the space of BMO processes starting from 0.

Let \( \alpha \) be a number in \((0,1]\) and let \( w : \{(s,t) \in [0,\tau]^2 : s \leq t\} \rightarrow [0,\infty) \) be a control, i.e. \( w(s,u) + w(u,t) \leq w(s,t) \) whenever \( s \leq u \leq t \). We assume throughout the article that \( w \) is continuous. An adapted process \( V \) belongs to \( \text{VMO}^\alpha_w \) if there is a finite constant \( C \) such that \( \rho_V(s,t) \leq Cw(s,t)^\alpha \) for every \( s \leq t \). For such process, we define

\[
[V]_{\text{VMO}^\alpha_w} := \sup_{s < t \leq \tau} \frac{\rho_{s,t}(V)}{w_{s,t}^\alpha}.
\]

When \( \alpha > 1 \), the space \( \text{VMO}^\alpha_w \) contains only constant processes. In the case of trivial control \( w(s,t) = t-s \), we simply write \( \text{VMO}^\alpha \).

**Theorem 3.2.** Let \((V_t)_{t \in [0,\tau]}\) be VMO. Then

\[ E_r e^{\lambda \sup_{t \in [r,\tau]} |V_t - V_r|} < \infty \text{ for every } r \in [0,\tau] \text{ and every } \lambda > 0. \] (3.1)

If \( V \) belongs to \( \text{VMO}^\alpha_w \), then

\[ \sup_{r \in [0,\tau]} E_r e^{\lambda \sup_{t \in [r,\tau]} |V_t - V_r|} \leq 2^{1+(22[V]_{\text{VMO}^\alpha_w,\lambda})^{1/\alpha} w_{0,\tau}} \text{ for every } \lambda > 0. \] (3.2)

**Proof:** The estimate (3.1) is a direct consequence of Theorem 2.3.

Suppose now that \( V \) is \( \text{VMO}^\alpha_w \) and put \( \rho = 11[V]_{\text{VMO}^\alpha_w} w^\alpha \). We define \( t_0 = 0 \) and for each integer \( k \geq 1 \),

\[ t_k = \sup \{ t \in [t_{k-1},\tau] : \lambda \rho_{t_{k-1},t} \leq 1/2 \}. \]

By continuity of \( w \), we have \( \lambda \rho_{t_{k-1},t_k} = 1/2 \) for \( k = 1, \ldots, n \) and \( \lambda \rho_{t_{n-1},t_n} \leq 1/2 \). By definition of controls, we have

\[ \frac{n-1}{(22[V]_{\text{VMO}^\alpha_w,\lambda})^{1/\alpha}} \leq \sum_{k=1}^{n} w_{t_{k-1},t_k} \leq w_{0,\tau}, \]

which yields \( n \leq 1+(22[V]_{\text{VMO}^\alpha_w,\lambda})^{1/\alpha} w_{0,\tau} \). Fix \( r \in [0,\tau] \) and define \( A_r = \sup_{s \in [r,\tau]} |V_s - V_r| \). Proposition 2.6 shows that \( A \) is BMO with \( \rho(A) \leq \rho \). Let \( j \) be such that \( t_{j-1} \leq r < t_j \). Following the proof of Proposition 2.5, we have

\[ E_r e^{\lambda A_r} \leq \prod_{k=j}^{n} (1-\lambda \rho_{t_{k-1},t_k})^{-1} \leq 2^n. \]

These estimates imply (3.2). \( \square \)
Corollary 3.3. Let $(V_t)_{t \in [0,\tau]}$ be $VMO^\alpha_w$. If $\alpha \in (0, 1)$, then there are constants $c_\alpha, C_\alpha$ such that for all $\lambda$ satisfying $\lambda \left( [V]_{VMO^\alpha w^\alpha_{0,\tau}} \right)^{1-\alpha} < c_\alpha$,

$$\mathbb{E} \exp \left( \lambda \sup_{t \leq \tau} |V_t - V_0|^{1-\alpha} \right) < \infty,$$  \hspace{1cm} (3.3)

and for all $p \geq 1$,

$$\| \mathbb{E} s \left[ \sup_{t \in [s,\tau]} |V_t - V_s|^p \right] \|_\infty \leq c_\alpha \Gamma(\alpha) \left( [V]_{VMO^\alpha w^\alpha_{s,t}} \right)^p,$$  \hspace{1cm} (3.4)

where $\Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du$ is the Gamma function.

If $\alpha = 1$ then

$$\mathbb{P} \left( |V_t - V_s| \leq 22 [V]_{VMO^1_w w_{s,t}} \text{ for all } s \leq t \leq \tau \right) = 1.$$  \hspace{1cm} (3.5)

Proof. Consider first the case $\alpha < 1$. Define $Z = \sup_{t \leq \tau} |V_t - V_0|$ and $a = 1/\alpha$. By Chebyshev inequality and (3.2), we have

$$\mathbb{P}(Z > x) = \mathbb{P}(e^{\lambda Z} > e^{\lambda x}) \leq e^{-\lambda x} \mathbb{E} e^{\lambda Z} \leq C e^{-\lambda x + c \beta^a},$$

where $\beta = ([V]_{VMO^\alpha w^\alpha_{0,\tau}})^a$ and $c = c(\alpha), C = C(\alpha)$ are some universal positive constants. One can optimize in $\lambda$ to obtain that for every $x$ bounded away from 0,

$$\mathbb{P}(Z > x) \leq C e^{-c_\alpha \beta^a \frac{x^a}{a} - \frac{1}{a} \left( \frac{1}{a} \right)}, \quad \frac{1}{a} + \frac{1}{a'} = 1,$$

where $c_\alpha, C$ are some other positive constants. In view of the layer cake representation

$$\mathbb{E} e^{\lambda Z} = \lambda a' \int_0^\infty e^{\lambda x^a' - \frac{1}{a}} x^a' \mathbb{P}(Z > x) dx,$$

we see that $\mathbb{E} e^{\lambda Z}$ is finite if $\lambda \beta^{1/(a-1)} < c_\alpha$. We obtain (3.3) by observing that $a' = 1/(1 - \alpha)$.

The estimate (3.4) is obtained in an analogous way. Define $Y = \sup_{t \in [s,\tau]} |V_t - V_s|$. Reasoning as previously,

$$\mathbb{P}_s(Y > x) \leq C e^{-c_\alpha \beta^a \frac{x^a}{a} - \frac{1}{a} \left( \frac{1}{a} \right)}.$$

By the layer cake representation,

$$\mathbb{E}_s Y^p = p \int_0^\infty x^{p-1} \mathbb{P}_s(Y > x) dx \leq C_p \int_0^\infty x^{p-1} e^{-c_\alpha \beta^{1/(a-1)} x^a} dx.$$

After the change of variable $y = c_\alpha \beta^{-1/(a-1)} x^a$, using the identity $\Gamma(z+1) = z \Gamma(z)$, we arrive at (3.4).

In the case $\alpha = 1$, a similar argument with Chebyshev inequality and (3.2) leads to

$$\mathbb{P}(|V_t - V_s| > x) \leq e^{-\lambda x + 22 [V]_{VMO^1_w w_{s,t}} \lambda}$$

for every $x > 0$ and $\lambda > 0$. When $x > 22 [V]_{VMO^1_w w_{s,t}}$, we can send $\lambda \to \infty$ to obtain that $\mathbb{P}(|V_t - V_s| > x) = 0$. This implies that $\mathbb{P}(|V_t - V_s| \leq 22 [V]_{VMO^1_w w_{s,t}}) = 1$. Since $V$ has continuous sample paths, this implies (3.5). \hfill \Box
Estimate (3.4) is inspired by the precise estimate from Lyons’ first extension theorem (inequality (2.21) in [Lyo98]), which is obtained through the so-called neo-classical inequality. Note that there is a smallness condition in the aforementioned article, which is not present in Corollary 3.3.

4. Applications

4.1. Rough stochastic differential equations. In [FHL21], the authors consider a hybrid rough stochastic differential equation of the type

\[ dY_t = b_t(Y_t)dt + \sigma_t(Y_t)dB_t + (f_t, f'_t)(Y_t)dX_t \]

where \( B \) is a standard Brownian motion and \( X = (X, X) \) is a Hölder rough path. The coefficients \( b, \sigma, f, f' \) are progressively measurable and regular in the \( Y \)-component. Under some natural regularity conditions, [FHL21] shows that (4.1) has a unique continuous solution in a certain class of stochastic controlled rough paths denoted by \( \mathbf{D}_X^aL_{m,\infty} \) for some \( \alpha \in (1/3, 1/2] \) and \( m \geq 2 \). Such processes are adapted and satisfy

\[ \sup_{s < t \leq r} \frac{(\|E_t[Y_t - Y_s]^m\|_\infty)^{1/m}}{(t - s)^\alpha} < \infty, \]  

(4.2)

together with some controllness conditions. As is shown in [FHL21], this class of stochastic controlled rough paths are stable under composition with smooth vector fields and rough stochastic integration. Because most of these properties are irrelevant for our purpose, we refer to the cited reference for further details.

Based upon earlier sections, any continuous adapted process satisfying the property (4.2) is VMO\(^\alpha\). Hence, such stochastic controlled rough paths are subjected to the John–Nirenberg inequality. Although [FHL21] also discusses exponential estimates for the solution of (4.1) by means of Lyons’ multiplicative functionals, their result comes with some additional restrictions on \( m, \alpha \) and the connection with VMO processes was not present there. On the other hand, our results actually accommodate minimal conditions that \( \alpha \in (0, 1] \) and \( m = 1 \).

**Theorem 4.1.** Let \( (Y_t)_{t \in [0, r]} \) be a continuous process in \( \mathbf{D}_X^aL_{1,\infty} \) (see Section 3 of [FHL21] for the precise definition) for some \( \alpha \in (0, 1] \). Then \( Y \) is VMO\(^\alpha\) and

\[ \mathbb{E}e^{\lambda \sup_{t \in [0, r]} |Y_t - Y_s|} \leq 2^{1 + (22Y_{VMO^\alpha})^1/\alpha} \text{ for every } \lambda > 0. \]

**Proof.** As discussed earlier, \( Y \) satisfies (4.2) with \( m = 1 \). By Proposition 2.2 and sample path continuity, \( Y \) is necessarily VMO\(^\alpha\). The exponential estimate is a direct consequence of (3.2). \[\Box\]

Another class of processes introduced in [FHL21] is \( C^\alpha L_{m,\infty} \) (with \( \alpha \in (0, 1] \) and \( m \geq 1 \)) consisting of adapted processes \( Y \) such that

\[ \sup_{t \in [0, r]} \|Y_t\|_m + \sup_{s < t \leq r} \frac{(\|E_t[Y_t - Y_s]^m\|_\infty)^{1/m}}{(t - s)^\alpha} < \infty. \]

We can classify these classes as VMO processes in the following way.
Proposition 4.2. Let $\alpha \in (0, 1]$ and $m \in [1, \infty]$ be some fixed numbers and $(Y_t)_{t \in [0, \tau]}$ be a continuous adapted process. $Y$ belongs to $C^\alpha L_{m, \infty}$ if and only if $Y_0$ is $L_m$-integrable and $Y$ is VMO$^\alpha$.

Proof. Straightforward from Proposition 2.2. □

4.2. Davie’s estimates. Let $(B_t)_{t \geq 0}$ be a standard Brownian motion and $g : [0, \tau] \times \mathbb{R}^d \to \mathbb{R}$ be a bounded Borel measurable function such that $|g(r, y)| \leq 1$ for all $(r, y) \in [0, \tau] \times \mathbb{R}^d$. Davie shows in [Dav07] that for any even integer positive integer $p$ and $x \in \mathbb{R}^d$,

$$
\mathbb{E} \left( \int_0^1 [g(t, B_t + x) - g(t, B_t)] \, dt \right)^p \leq C^p \Gamma(p/2 + 1)|x|^p, \tag{4.3}
$$

where $C$ is an absolute constant. This inequality exhibits the regularization effect of Brownian motion through temporal integration. Such regularization effect is developed further into the framework of nonlinear Young integration by Catellier and Gubinelli in [CG16]. Davie’s estimate is an important one and has been reproduced in different forms under other conditions and setups [Sha16, RZ21, Rez14, Lê22, Lê20, ABLM20].

Davie shows (4.3) by first expanding the moment into an iterated multiple integral. The lack of regularity in $g$ is compensated by the smooth density of the Brownian motion through integration by parts. This procedure produces a sum of iterated multiple integrals involving derivatives of the Gaussian density. He then estimates each of these multiple integrals carefully to obtain (4.3). Davie’s proof is beautiful yet intricate because of its analysis of high moments. We now explain how John–Nirenberg inequality in Theorems 2.3 and 3.2 could be utilized in this context. Indeed, following Davie’s proof in [Dav07], one has

$$
\mathbb{E} \left( \int_s^t [g(r, B_r + x) - g(r, B_r)] \, dr \right)^2 \leq C^2 |x|^2 (t - s).
$$

Since Brownian motion has independent increments, we can upgrade the above inequality to the following estimate

$$
\mathbb{E}_s \left( \int_s^t [g(r, B_r + x) - g(r, B_r)] \, dr \right)^2 \leq C^2 |x|^2 (t - s).
$$

This shows that the process $V_t = \int_0^t [g(r, B_r + x) - g(r, B_r)] \, dr$ is VMO$^{1/2}$ with $[V]_{\text{VMO}^{1/2}} \leq C|x|$. The estimate (3.4) gives

$$
\mathbb{E} \left( \sup_{t \in [0, 1]} \int_0^t [g(r, B_r + x) - g(r, B_r)] \, dr \right)^p \leq C^p \Gamma(p/2 + 1)|x|^p. \tag{4.4}
$$

This estimate is comparable to or perhaps stronger than (4.3) because of the supremum in its left-hand side.
4.3. Quadrature error estimates and strong convergence rate of Euler method. Consider the stochastic differential equation
\[ dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x_0, \quad t \in [0, 1], \]  
where \( d \geq 1, \ b : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d, \ \sigma : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d \) are Borel measurable functions, \((B_t)_{t \geq 0}\) is a \( d \)-dimensional standard Brownian motion defined on some complete probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) and \( x_0 \) is an \( \mathcal{F}_0 \)-random variable. The tamed Euler–Maruyama scheme associated to \((4.5)\) is
\[ dX^n_t = b^n(t, X^n_{k_n(t)})dt + \sigma(t, X^n_{k_n(t)})dB_t, \quad X^n_0 = x^n_0, \quad t \in [0, 1], \]  
where \( x^n_0 \) is a \( \mathcal{F}_0 \)-random variable and \( b^n \) is an approximation of the vector field \( b \) and \( k_n(t) = \frac{j}{n} \) whenever \( \frac{j}{n} \leq t < \frac{j+1}{n} \) for some integer \( j \geq 0 \).

We note that \((4.6)\) with the choice \( b^n = b \) is the usual Euler–Maruyama scheme, which, however, is not well-defined for a merely integrable function \( b \) even when \( b \) is replaced by \( b 1_{\{b < \infty\}} \). This is because the simulation for the usual Euler–Maruyama scheme may enter a neighborhood of a singularity of \( b \), making the scheme unstable and uncontrollable.

The recent article [LL21] establishes strong rate of convergence of the tamed Euler–Maruyama scheme \((4.6)\) to \((4.5)\) under some integrability condition of the drift \( b \). To state their result, we first recall some notation from [LL21]. Let \( p, q \in [1, \infty] \) be some fixed parameters. \( L_p(\mathbb{R}^d) \) and \( L_p(\Omega) \) denote the Lebesgue spaces respectively on \( \mathbb{R}^d \) and \( \Omega \). For each \( \nu \in \mathbb{R}, L_{\nu, p}(\mathbb{R}^d) := (1 - \Delta)^{-\nu/2} (L_p(\mathbb{R}^d)) \) is the usual Bessel potential space on \( \mathbb{R}^d \) equipped with the norm \( \| f \|_{L_{\nu, p}(\mathbb{R}^d)} := \| (1 - \Delta)^{\nu/2} f \|_{L_p(\mathbb{R}^d)}, \) where \( (1 - \Delta)^{\nu/2} f \) is defined through Fourier’s transform. \( L_{\nu, p}^q([0, 1]) \) denotes the space of measurable function \( f : [0, 1] \to L_{\nu, p}(\mathbb{R}^d) \) such that \( \| f \|_{L_{\nu, p}^q([0, 1])} \) is finite.

Here, for each \( s, t \in [0, 1] \) satisfying \( s \leq t \), we denote
\[ \| f \|_{L_{\nu, p}^q([s, t])} := \left( \int_{s}^{t} \| f(r, \cdot) \|_{L_{\nu, p}(\mathbb{R}^d)}^q dr \right)^{\frac{1}{q}} \]
with obvious modification when \( q = \infty \). When \( \nu = 0 \), we simply write \( L_p^q([0, 1]) \) instead of \( L_{0, p}^q([0, 1]) \). In particular, \( L_p^q([0, 1]) \) contains Borel measurable functions \( f : [0, 1] \times \mathbb{R}^d \to \mathbb{R} \) such that \( \int_0^1 \left[ \int_{\mathbb{R}^d} |f(t, x)|^p dx \right]^{q/p} dt \) is finite. As in [LL21], we assume the following conditions.

**Condition \( \mathcal{A} \).** The diffusion coefficient \( \sigma \) is a \( d \times d \)-matrix-valued measurable function on \([0, 1] \times \mathbb{R}^d \). There exists a constant \( K_1 \in [1, \infty) \) such that for every \( s \in [0, 1] \) and \( x \in \mathbb{R}^d \)
\[ K_1^{-1} I \leq (\sigma \sigma^*)(s, x) \leq K_1 I, \]  
where \( I \) denotes the identity matrix. Furthermore, the following conditions hold.

1. There are constants \( \alpha \in (0, 1) \) and \( K_2 \in (0, \infty) \) such that for every \( s \in [0, 1] \) and \( x, y \in \mathbb{R}^d \)
\[ |(\sigma \sigma^*)(s, x) - (\sigma \sigma^*)(s, y)| \leq K_2 |x - y|^{\alpha}. \]
2. \( \sigma(s, \cdot) \) is weakly differentiable for a.e. \( s \in [0, 1] \) and there are constants \( p_0 \in [2, \infty), \)
\( q_0 \in (2, \infty) \) and \( K_3 \in (0, \infty) \) such that
\[
\frac{d}{p_0} + \frac{2}{q_0} < 1 \quad \text{and} \quad \| \nabla \sigma \|_{L^{q_0}([0,1])} \leq K_3.
\]

**Condition \( \mathfrak{G} \).** \( x_0 \) belongs to \( L_p(\Omega, \mathcal{F}_0) \) and \( b \) belongs to \( L_p^q([0,1]) \) for some \( p, q \in [2, \infty) \) satisfying \( \frac{1}{p} + \frac{2}{q} < 1 \). For each \( n, x^n_0 \) belongs to \( L_p(\Omega, \mathcal{F}_0) \) and \( b^n \) belongs to \( L_p^q([0,1]) \cap L_p^\infty([0,1]) \) with \( p, q \) as above. Furthermore, there exist finite positive constants \( K_4, \theta \) and continuous controls \( \{\mu^n\}_n \) such that \( \sup_{n \geq 1}(\| b^n \|_{L_p^q([0,1])} + \mu^n(0,1)) \leq K_4 \) and
\[
(1/n)^{\frac{1}{2} - \frac{1}{q}} \| b^n \|_{L_{20}^q([1,t])} \leq \mu^n(s,t)^\theta \quad \forall t - s \leq 1/n.
\]

**Definition 4.3.** Let \( \lambda > 0 \) be a fixed number which is sufficiently large. Let \( U = (U^1, \ldots, U^d) \) where for each \( h = 1, \ldots, d, U^h \) is the solution to the following equation
\[
\partial_t U^h + \sum_{i,j=1}^d \frac{1}{2} (\sigma \sigma^*)^{ij} \partial_{ij} U^h + b^h \cdot \nabla U^h = \lambda U^h - b^h, \quad U^h(1, \cdot) = 0.
\]

Let \( X \) be the solution to (4.5). For each \( \bar{p} \in [1, \infty) \), we put
\[
\omega_n(\bar{p}) = \left\| \sup_{t \in [0,1]} \right\| \int_0^t (1 + \nabla U) \left[ b - b^n \right](r, X_r) dr \left\|_{L_p(\Omega)} \right. .
\]

Note that we have changed the definition of \( \omega_n \) from [LL21] by replacing \( b^{n,h} \) with \( b^h \) in (4.9).

The main result of [LL21] (Theorems 2.2 and 2.3 therein) asserts that for any \( \bar{p} \in (1, p) \cap (1, \frac{1}{\theta}(p \wedge p_0)) \) and any \( \gamma \in (0, 1) \), there exists a finite constant \( N \) such that
\[
\| \sup_{t \in [0,1]}|X^n_t - X_t|_{L_{\bar{p}}(\Omega)} \|_{L_p(\Omega)} \leq N \left[ \| x^n_0 - x_0 \|_{L_p(\Omega)} + (1/n)^{\frac{1}{2}} + (1/n)^{\frac{1}{2}} \log(n) + \omega_n(\bar{p}) \right].
\]

To obtain estimate, [LL21] first utilize stability results for (4.5) to show that the strong convergence rate is bounded by \( \omega_n(\bar{p}) \) and the quadrature error of the type
\[
\| \sup_{t \in [0,1]} \int_0^t g(r, X^n_r) (f(r, X^n_r) - f(r, X^n_{k_n(r)})) dr \|_{L_p(\Omega)}
\]
where \( f \in L_p^q \cap L_p^\infty \) and \( g \in L_p^q \cap L_p^\infty \). [LL21] then applies stochastic sewing techniques (introduced in [Lê20]) to obtain the rate \( (1/n)^{1/2} \log(n) \) for the quadrature error and to bound \( \omega_n \) by a suitable distance between \( b \) and \( b^n \).

While (4.10) produces the best available rate in the literature, it comes with an unnatural constraint on \( \bar{p} \). This restriction is purely technical and is necessary for both stability analysis and stochastic sewing arguments described previously. In the more recent article [GL22], a stability estimate which is valid for all moments has been obtained. In the present article, we
utilize the John–Nirenberg inequality (Theorem 2.3) to overcome the moment restriction in the stochastic sewing arguments used to estimate $\omega_n$ and the quadrature error. Our main contribution are the following two results, which remove the moment restrictions from Theorems 2.2 and 2.3 of [LL21].

**Theorem 4.4.** Assume that Conditions $\mathfrak{A}$-$\mathfrak{B}$ hold. Let $(X^n_t)_{t\in[0,1]}$ be the solution to (4.6) and $(X_t)_{t\in[0,1]}$ be the solution to (4.5). Then for any $\bar{p} \in (1, \infty)$ and any $\gamma > 1$, there exists a finite constant $N(K_1, K_2, K_3, K_4, \alpha, p_0, q_0, p, q, d, \bar{p}, \gamma)$ such that

$$
\big\| \sup_{t\in[0,1]} |X^n_t - X_t| \big\|_{L^p(\Omega)} \leq N \left[ \|X^n_0 - x_0\|_{L^p(\Omega)} + (1/n)^{\frac{q}{2}} + (1/n)^\frac{1}{2} \log(n) + \omega_n(\gamma \bar{p}) \right].
$$

(4.11)

**Theorem 4.5.** Assume that Conditions $\mathfrak{A}$-$\mathfrak{B}$ hold with $q_0 = \infty$ and $\frac{1}{p} + \frac{1}{p_0} < 1$. Let $\nu \in [0, 1)$ be such that

$$
\nu < \frac{3}{2} - \frac{d}{2p} - \frac{2}{q}.
$$

(4.12)

Then for every $\bar{p} \in (1, \infty)$, there exists a constant $N$ depending on $K_1, K_2, K_3, K_4, \alpha, p_0, q_0, p, q, d, \bar{p}, \nu$ such that

$$
\omega_n(\bar{p}) \leq N\|b - b^n\|_{L_{-\nu p}(\Omega)}.
$$

(4.13)

[LL21] also considers the case $\nu = 1$ in Theorem 4.5. Our argument also works in this case without much effort. We therefore leave it for interested readers.

**Proposition 4.6.** Let $p \in (1, \infty)$, $q \in (2, \infty)$ and assume that Condition $\mathfrak{A}$ holds with $q_0 = \infty$ and $\frac{1}{p} + \frac{1}{p_0} < 1$. Let $X$ be a solution to (4.5). Let $g$ be a function in $L^q_p([0, 1])$ and let $\nu \in [0, 1)$ such that $\frac{d}{p} + \frac{2}{q} + \nu < 2$. Then for any $\bar{p} \in [1, \infty)$, there exists a constant $N = N(\nu, d, p, q, \bar{p})$ such that

$$
\big\| \sup_{t\in[0,1]} \int_0^t g(r, X_r)dr \big\|_{L^p(\Omega)} \leq N\|g\|_{L^q_{-\nu p}(\Omega)}.
$$

(4.14)

**Proof.** By Girsanov transformation, we can assume without loss of generality that $b = 0$ (see the argument in the proof of Theorem 5.1 in [LL21]). Put $V_t = \int_0^t g(r, X_r)dr$. We note that by Krylov estimate, $\|V_t - V_\bar{t}\|_{L^p(\Omega)} \leq \|g\|_{L^q_{-\nu p}(\Omega)}(t - s)^{1 - \frac{d}{p} - \frac{2}{q} - \nu}$ for all $m \geq 1$ (see inequality (5.0) and Lemma 3.4 from [LL21]). Consequently, $V$ is a.s. continuous. The proof of Proposition 6.6 from [LL21] shows that

$$
\sup_{s \leq t \leq 1} (\|E[V_t - V_{\bar{t}}]\|_{L^p(\Omega)})^{1/p} \leq \|g\|_{L^{\infty}_{-\nu p}(\Omega)}.
$$

This shows that $V$ is BMO and hence estimate (4.14) follows from Theorem 2.3. \qed

**Proposition 4.7.** Assume that Conditions $\mathfrak{A}1$ and $\mathfrak{B}$ hold. Let $X^n$ be the solution to (4.6) and let $f, g$ be measurable functions on $[0, 1] \times \mathbb{R}^d$. Assume that $\|f\|_{L^q_p([0, 1])} = \|g\|_{L^\infty([0, 1])} + \|g\|_{L^q_{-\nu p}(\Omega)} =$...
and \( \beta_n(f) = \sup_{0 \leq j \leq n-1} \| f \|_{L^p_{\omega}(\{j/n(j+1)/n]\))} \) is finite. Then for any \( \bar{p} \geq 1 \), there exists a constant 
\( N = N(d, p, q, \bar{p}) \) such that

\[
\left\| \sup_{t \in [0, 1]} \left| \int_0^t g(r, X^n_r) [f(r, X^n_r) - f(r, X^n_{k_n(r)})] dr \right|_{L^p(\Omega)} \right\| \leq N \left[ (1/n)^{1 - \frac{1}{\bar{p}}} \beta_n(f) + (1/n)^{\frac{1}{\bar{p}}} + (1/n)^{1/2} \log(n) \right]. \tag{4.15}
\]

**Proof.** By Girsanov transformation, we can assume without loss of generality that \( b = 0 \) (see the argument in the proof of Theorem 5.1 in [LL21]). Define

\[
V_t = \int_0^t g(r, X^n_r) [f(r, X^n_r) - f(r, X^n_{k_n(r)})] dr.
\]

Proposition 5.12 of [LL21] and its proof shows that for every \( v + 4/n \leq s \leq t \leq 1 \),

\[
\| \mathbb{E}_s |V_t - V_s|^p \|_{\infty}^{1/p} \leq \left[ (1/n)^{\frac{v}{2}} + (1/n)^{\frac{v}{2}} \log(n) \right].
\]

By assumption and Hölder inequality, we have for every \( s \leq t \)

\[
|V_t - V_s| \leq \| f \|_{L^p_{\omega}([s,t])} (t - s)^{1 - \frac{1}{v}}.
\]

Combining the previous two estimates yields that for every \( s \leq t \)

\[
\| \mathbb{E}_s |V_t - V_s|^p \|_{\infty}^{1/p} \leq \left[ (1/n)^{\frac{v}{2}} + (1/n)^{\frac{v}{2}} \log(n) + (1/n)^{1 - \frac{1}{2}} \beta_n(f) \right].
\]

Since \( V \) is continuous, from Proposition 2.2, it is BMO (in fact VMO) and hence by applying Theorem 2.3, we obtain (4.15). \( \square \)

**Proof of Theorem 4.4.** The stability argument in [LL21] comes with a restriction on the moment and we must replace it by the recent stability estimate from [GL22]. Indeed, from Section 3.4.1 of the aforementioned reference, we have

\[
\| \sup_{t \in [0, 1]} |X_t - X^n_t| \|_{L^p(\Omega)} \leq \| x_0 - x^n_0 \|_{L^p(\Omega)} + \| \sup_{t \in [0, 1]} |V_t| \|_{L^p(\Omega)}, \tag{4.16}
\]

where

\[
V_t = \int_0^t \left( \frac{1}{2} [(R^2 + \sigma)(R^2 + \sigma)^* - \sigma \sigma^*) : D^2 U + R^1 \cdot (I + \nabla U) \right) (r, X^n_r) dr
\]

\[
+ \int_0^t [R^2(I + \nabla u)](r, Y_t) dB_r,
\]

\[
R^1_t = b^n(t, X^n_{k_n(t)}) - b(t, X^n_t), \quad R^2_t = \sigma(t, X^n_{k_n(t)}) - \sigma(t, X^n_t).
\]

Since \( \sigma \) is Hölder continuous, the moments of terms with \( R^2 \) are bounded by a constant multiple of \( (1/n)^{\alpha/2} \) (see Section 7 of [LL21] for some analogous estimates). To treat the term with \( R^1 \), we note that
To treat the first term, we apply Proposition 4.7, regularity of $U$ (Lemma 7.1 of [LL21]) and Condition $\mathcal{B}$ to have

$$\left\| \sup_{t \in [0,1]} \left| \int_0^t (b^n(r, X^n_{k_t(r)}) - b^n(X^n_r)) \cdot (I + \nabla U)(r, X^n_r) dr \right| \right\|_{L_p(\Omega)} \leq (1/n)^{1-\frac{1}{2}} \beta_n(b^n) + (1/n)^{\frac{1}{2}} + (1/n)^{\frac{1}{2}} \log(n) \leq (1/n)^{\frac{1}{2}} + (1/n)^{\frac{1}{2}} \log(n).$$

Moment of the second term is directly related to $\omega_n(\bar{\rho})$ through Definition 4.3. This leads us to the following estimate

$$\left\| \sup_{t \in [0,1]} |V_t| \right\|_{L_p(\Omega)} \leq (1/n)^{\frac{1}{2}} + (1/n)^{\frac{1}{2}} \log(n) + \omega_n(\bar{\rho}).$$

Combining with (4.16), we obtain (4.11).

**Proof of Theorem 4.5.** The proof follows in exactly the same way as the proof of Theorem 2.3 in [LL21] (Section 7 therein). The restriction $\bar{\rho} < \rho$ there is now lifted thanks to Proposition 4.6. \qed

**Appendix A. Auxiliary results**

**Lemma A.1 (Garsia’s upcrossing lemma, [Gar73b, Str73]).** Let $(X_t)_{t \in [0,\tau]}$ be a right continuous adapted process with left limits and let $s$ be a fixed time in $[0,\tau]$. Suppose that there is a non-negative integrable random variable $U$ such that

$$\mathbb{E}_S[X_T - X_{S-}] \leq \mathbb{E}_S U \quad \text{(A.1)}$$

for any pair $S,T$ of stopping times with $s \leq S \leq T \leq \tau$. Let $Y$ be an $\mathcal{F}_s$-random variable and define $X^* = \sup_{s \leq r \leq T} |X_r - Y|$. Then for every $\alpha, \beta > 0$, one has

$$\beta \mathbb{P}_s(X^* \geq \alpha + \beta) \leq \mathbb{E}_s(U 1_{(X^* \geq \alpha)}). \quad \text{(A.2)}$$

**Proof.** We adopt the arguments from [Kaz94]. Let $\alpha, \beta > 0$ be given and $G \in \mathcal{F}_s$. We set $X_t = X_{t \wedge T}$ for $t > \tau$ and define

$$S = \inf\{t \geq s : |X_t - Y| \geq \alpha\}, \quad T = \inf\{t \geq s : |X_t - Y| \geq \alpha + \beta\},$$

with the standard convention that $\inf(\emptyset) = \infty$. Clearly $S$ and $T$ are stopping times and $s \leq S \leq T$. We also have from the above definitions,

$$\{X^* \geq \alpha + \beta\} \subset \{|X_T - X_{S-}| \geq \beta, |X_S - Y| \geq \alpha\}. \quad \text{(A.3)}$$

It follows that

$$\mathbb{E}(1_{(X^* \geq \alpha + \beta)} 1_G) \leq \mathbb{E}(1_{|X_T - X_{S-}| \geq \beta} 1_{(|X_S - Y| \geq \alpha)} 1_G) \leq \frac{1}{\beta} \mathbb{E}(1_{|X_T - X_{S-}|} 1_{(|X_S - Y| \geq \alpha)} 1_G)$$
which implies the result. □

We note that right-continuity of the filtration is necessary so that \( \alpha, \beta \) defined in the previous proof are stopping times. In addition, the inclusion (A.3) does not hold if one replaces \( X_{S-} \) by \( X_S \) in (A.1). These technical conditions become irrelevant when dealing with continuous processes.

**Lemma A.2** (Energy inequality). Let \( c \) be a deterministic constant and \( (A_t)_{t \geq 0} \) be an adapted, right-continuous, non-decreasing process. Let \( \tau > 0 \) be fixed and suppose that

\[
\|\mathbb{E}(A_{\tau} - A_{S-})\|_{\infty} \leq c \text{ for every stopping time } S \leq \tau.
\]  

Then for every \( s \in [0, \tau] \) and every integer \( p \geq 1 \),

\[
\|\mathbb{E}(A_{\tau} - A_s)^p\|_{\infty} \leq p! c^p.
\]

**Proof.** When \( A_t \) takes the specific form \( \int_0^t \beta(r)dr \) for some \( \beta \geq 0 \), this result deduces to the Khasminskii’s lemma ([Kha59]). In the general form, it is known as energy inequality and can be found in [Mey66, Kik92]. Our statement is for processes over finite time intervals which differs from previous ones and needs justifications.

Let \( s \in [0, \tau] \) be fixed and \( G \) be an event in \( \mathcal{F}_s \). For each \( r \geq 0 \), define \( \tilde{A}_r = 1_G(A_{(r+s)\wedge \tau} - A_r) \). The process \( \tilde{A} \) is adapted with respect to the filtration \( \tilde{\mathcal{F}} := \{\mathcal{F}_{r+s}\}_{r \geq 0} \), right-continuous, satisfies \( \tilde{A}_0 = 0 \) and \( \|\mathbb{E}(\tilde{A}_{\tau} - \tilde{A}_{S-})\|_{\infty} \leq c \) for all \( \tilde{\mathcal{F}} \)-stopping times \( S \). Applying Theorem 4 of [Kik92] to the process \( \tilde{A} \), we get that \( \mathbb{E}(1_G(A_{\tau} - A_s)^p) \leq p! c^p \). Since \( G \) is arbitrary, this implies (A.5). □

**Acknowledgment**

The author thanks Peter Friz for motivating discussions on exponential integrability and his interest in this work.

**Funding**

The author was supported by Alexander von Humboldt Research Fellowship during the early stage of this work.

**References**

[ABLM20] Siva Athreya, Oleg Butkovsky, Khoa Lê, and Leonid Mytnik. Well-posedness of stochastic heat equation with distributional drift and skew stochastic heat equation. *arXiv preprint arXiv:2011.13498*, 2020.

[Alt21] Randolf Altmeyer. Approximation of occupation time functionals. *Bernoulli*, 27(4):2714–2739, 2021.

[AS82] M. Aizenman and B. Simon. Brownian motion and Harnack inequality for Schrödinger operators. *Comm. Pure Appl. Math.*, 35(2):209–273, 1982.

[BN08] Jocelyne Bion-Nadal. Dynamic risk measures: time consistency and risk measures from BMO martingales. *Finance Stoch.*, 12(2):219–244, 2008.
[CG16] R. Catellier and M. Gubinelli. Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.*, 126(8):2323–2366, 2016.

[Dav07] A. M. Davie. Uniqueness of solutions of stochastic differential equations. *Int. Math. Res. Not. IMRN*, (24):Art. ID rnm124, 26, 2007.

[DG20] Konstantinos Dareiotis and Máté Gerencsér. On the regularisation of the noise for the Euler-Maruyama scheme with irregular drift. *Electron. J. Probab.*, 25:Paper No. 82, 18, 2020.

[DGL22] Konstantinos Dareiotis, Máté Gerencsér, and Khoa Lê. Quantifying a convergence theorem of Gyöngy and Krylov. *Annals of Applied Probability*, 2022+.

[DMS97] Freddy Delbaen, Pascale Monat, Walter Schachermayer, Martin Schweizer, and Christophe Stricker. Weighted norm inequalities and hedging in incomplete markets. *Finance and Stochastics*, 1(3):181–227, 1997.

[Fef71] Charles Fefferman. Characterizations of bounded mean oscillation. *Bull. Amer. Math. Soc.*, 77:587–588, 1971.

[FHL21] Peter Friz, Antoine Hocquet, and Khoa Lê. Rough stochastic differential equations. *arXiv preprint arXiv:2106.10340*, 2021.

[FS72] C. Fefferman and E. M. Stein. $H^p$ spaces of several variables. *Acta Math.*, 129(3-4):137–193, 1972.

[FZ18] Peter K. Friz and Huilin Zhang. Differential equations driven by rough paths with jumps. *J. Differential Equations*, 264(10):6226–6301, 2018.

[Gar73a] Adriano M. Garsia. The Burgess Davis inequalities via Fefferman’s inequality. *Ark. Mat.*, 11:229–237, 1973.

[Gar75b] Adriano M. Garsia. *Martingale inequalities: Seminar notes on recent progress*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., Reading, Mass.-London-Amsterdam, 1973.

[Gei05] Stefan Geiss. Weighted BMO and discrete time hedging within the Black-Scholes model. *Probab. Theory Related Fields*, 132(1):13–38, 2005.

[GG22] Lucio Galeati and Máté Gerencsér. Solution theory of fractional SDEs in complete subcritical regimes. *arXiv preprint arXiv:2207.03475*, 2022.

[GL22] Lucio Galeati and Chengcheng Ling. Stability estimates for singular SDEs and applications. *arXiv preprint arXiv:2208.03670*, 2022.

[GS72] R. K. Getoor and M. J. Sharpe. Conformal martingales. *Invent. Math.*, 16:271–308, 1972.

[Her74] Carl Herz. Bounded mean oscillation and regulated martingales. *Trans. Amer. Math. Soc.*, 193:199–215, 1974.

[JN61] F. John and L. Nirenberg. On functions of bounded mean oscillation. *Comm. Pure Appl. Math.*, 14:415–426, 1961.

[Kaz94] Norihiko Kazamaki. *Continuous exponential martingales and BMO*, volume 1579 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1994.

[Kha59] R. Z. Khasminskii. On positive solutions of the equation $Uu + Vu = 0$. *Theor. Probability Appl.*, 4:309–318, 1959.

[Kik92] Masato Kikuchi. A note on the energy inequalities for increasing processes. In *Séminaire de Probabilités, XXVI*, volume 1526 of *Lecture Notes in Math.*, pages 533–539. Springer, Berlin, 1992.

[KR05] N. V. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, 131(2):154–196, 2005.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[Lé20] Khoa Lê. A stochastic sewing lemma and applications. *Electron. J. Probab.*, 25:Paper No. 38, 55, 2020.

[Lé22] Khoa Lê. Stochastic sewing in Banach space. *arXiv preprint arXiv:2105.09364*, 2022.

[LL21] Khoa Lê and Chengcheng Ling. Taming singular stochastic differential equations: A numerical method. *arXiv preprint arXiv:2110.01343*, 2021.
[Lyo98] Terry J. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[Mey66] Paul-A. Meyer. *Probability and potentials*. Blaisdell Publishing Co. [Ginn and Co.], Waltham, Mass.-Toronto, Ont.-London, 1966.

[Ose15] Adam Osekowski. Sharp maximal estimates for BMO martingales. *Osaka J. Math.*, 52(4):1125–1142, 2015.

[Por90] N. I. Portenko. *Generalized diffusion processes*, volume 83 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1990. Translated from the Russian by H. H. McFaden.

[Rez14] Fraydoun Rezakhanlou. Regular flows for diffusions with rough drifts. *arXiv preprint arXiv:1405.5856*, 2014.

[RZ21] Michael Röckner and Guohuan Zhao. Sdes with critical time dependent drifts: strong solutions. *arXiv preprint arXiv:2103.05803*, 2021.

[Sha16] A. V. Shaposhnikov. Some remarks on Davie’s uniqueness theorem. *Proc. Edinb. Math. Soc.* (2), 59(4):1019–1035, 2016.

[Str73] Daniel W. Stroock. Applications of Fefferman-Stein type interpolation to probability theory and analysis. *Comm. Pure Appl. Math.*, 26:477–495, 1973.

[Zha16] Xicheng Zhang. Stochastic differential equations with Sobolev diffusion and singular drift and applications. *Ann. Appl. Probab.*, 26(5):2697–2732, 2016.

School of Mathematics, University of Leeds, U.K.

*Email address*: k.1e@leeds.ac.uk