Bernoulli **24**(4B), 2018, 3222–3245
https://doi.org/10.3150/17-BEJ958

Applications of pathwise Burkholder–Davis–Gundy inequalities

PIETRO SIORPAES[1]

Imperial College London, Huxley Building, Office 6M.19 180 Queen's Gate, South Kensington, London, SW72AZ, United Kingdom. E-mail: p.siorpaes@imperial.ac.uk

In this paper, after generalizing the *pathwise* Burkholder–Davis–Gundy (BDG) inequalities from discrete time to cadlag semimartingales, we present several applications of the pathwise inequalities. In particular we show that they allow to extend the classical BDG inequalities

1. to the Bessel process of order $\alpha \geq 1$
2. to the case of a *random* exponent p
3. to martingales stopped at a time τ which belongs to a well studied class of *random* times

Keywords: Bessel process; Burkholder–Davis–Gundy; pathwise martingale inequalities; pseudo stopping time; semimartingale; variable exponent

1. Introduction

In recent years, a new method of systematically proving martingale inequalities through pathwise counterparts has emerged. This approach, which historically arose from considerations in robust mathematical finance in the seminal paper by Hobson [14], has in particular been applied to derive the pathwise Burkholder–Davis–Gundy (BDG) inequalities: see [5], where in Section 2 one can also find more information on the history of the subject. Notice that one can find early instances of this approach in the proofs of [31], Chapter 4, Propositions 4.3 and 4.4.

The first goal of this paper is to generalize the pathwise BDG inequalities of [5] from discrete to continuous time; specifically we to show that, if X is a cadlag semimartingale and Φ a very general function, one can explicitly construct integrands $K, \tilde{K}$ such that for some constant C_Φ the following pathwise BDG inequalities hold

$$\Phi\big(\sqrt{[X]_t}\big) \leq C_\Phi \Phi\big(X_t^*\big) + (K \cdot X)_t, \qquad \Phi\big(X_t^*\big) \leq C_\Phi \Phi\big(\sqrt{[X]_t}\big) + (\tilde{K} \cdot X)_t; \tag{1}$$

this turns out to be easy if $\Phi(t) = t$ but hard in general, even in the case where $\Phi(t) = t^p/p$ for some $p > 1$.

The second goal of this paper is to show that the pathwise inequalities are strictly more powerful than the classical ones, in useful ways. Indeed, we present several applications of (1), in which we are able to extend the classical BDG inequalities beyond their traditional domain of

[1]This paper was written and submitted for publication while the author was a postdoctoral researcher at the University of Oxford and was supported by the European Research Council under (FP7/2007-2013)/ERC Grant agreement no. 335421.

1350-7265

validity. Although we concentrate our attention exclusively on the BDG inequalities, it is clear that also for other martingale inequalities the pathwise version is going to be analogously 'better' than the classical one; in this regard, it is interesting to keep in mind that every martingale inequality in finite discrete time admits a pathwise equivalent: see [3,8].

How can these additional applications arise? Trivially the pathwise martingale inequalities imply their classical equivalent by (localizing and) taking expectations and using the fact that $\mathbb{E}[(H \cdot X)_\tau] = 0$ when X is a local martingale, τ is a stopping time (one small enough to make $\sup_{t \le \tau} |(H \cdot X)_t|$ integrable) and $H = K, \tilde{K}$. The whole idea is then that there are interesting examples where X is a not a local martingale or τ is a not stopping time and yet one can somehow control the size of $\mathbb{E}[(H \cdot X)_\tau]$ for $H = K, \tilde{K}$. Let us now review our applications of (1) one by one.

If B is a Brownian motion in $\mathbb{R}^n$ started at B_0, $X := \|B\|_{\mathbb{R}^n}$ and $\Phi(t) = t^p/p$ with[2] $p > 0$, the BDG inequalities applied to B imply that for some c, C

$$c\mathbb{E}\Phi\big(\sqrt{[X]_\tau}\big) \le \mathbb{E}\Phi\big(X^*_\tau\big) \le C\mathbb{E}\Phi\big(\sqrt{[X]_\tau}\big) \qquad \text{for all stopping times } \tau. \tag{2}$$

In other words, the BDG inequalities hold for such a process X, even if X is not a local martingale; X is called an n-dimensional Bessel process. More generally (but without making a connection with Brownian motion) one can define the α-dimensional Bessel process X for all $\alpha \in \mathbb{R}$. This is a positive Feller process with continuous paths, and it is a semimartingale if $\alpha \notin (0, 1)$, so it is natural to ask for which values of $\alpha \notin (0, 1)$ the BDG inequalities hold. This questions was answered by [12], Theorem 4.1, where one can find a proof (only provided for the case $p = 1$) of the fact that (2) holds if $\alpha \ge 1$. We will show how this is just a corollary of the pathwise BDG inequalities, and that in the case $p = 1$ this is really easy to show. Even better (and curiously enough) we do not recover the BDG inequalities for Bessel processes for $p \ge 1$ from the corresponding $p \ge 1$ pathwise BDG inequalities (which are laborious to prove), but rather applying a strengthened version of the easy-to-prove pathwise Davis inequalities (i.e., the case $p = 1$), obtained by a delicate modification of the arguments in [11], Chapter 7, Lemma 91. So, while the ideas of [12] yield constants with the 'correct' scaling in α and ours do not, our approach has the advantage of simplicity, and allows to derive these inequalities relying on a systematic way to generalize martingale inequalities, instead of applying ad-hoc methods.

Then, we develop a variant of the BDG inequalities, connecting the expectations of the p-th power of $\sqrt{[X]_t}$ and X^*_t, in the case in which the exponent $p \ge 1$ is a random variable. In particular when p is $\mathcal{F}_0$-measurable our inequalities imply that

$$\mathbb{E}\sqrt{[X]^p_\infty} \le \mathbb{E}\big(6pX^*_\infty\big)^p, \qquad \mathbb{E}\big(X^*_\infty\big)^p \le \mathbb{E}\big(6p\sqrt{[X]_\infty}\big)^p \tag{3}$$

for every cadlag local martingale X; this implies[3] that (3) holds if $p \ge 1$ is any random variable independent of X (of course this also easily follows taking conditional expectations with respect

[2] We recall that, while in general the BDG inequalities only hold for $p \ge 1$, they hold for any $p > 0$ for *continuous* local martingales. The inequalities also hold for very general functions Φ: for the cadlag (resp. continuous) case one can take Φ as in Theorem 3 (resp. Theorem 5).

[3] Indeed in this case X is also a local martingale with respect to the filtration $\mathcal{F}_t := \sigma(p, (X_s)_{s \le t})$.

to p and then applying the classical BDG inequalities). In the case where the probability space has no atoms and p is bounded above, (3) has essentially[4] appeared in [23], Theorem 3.1 (which is still unpublished), which claims that the result also holds without the assumption that p is $\mathcal{F}_0$-measurable. Under additional hypotheses, Doob's inequality has appeared in [16], Theorem 3.5, and in [2], Corollary 1, for a $\mathcal{F}_0$-measurable p. In light of Doob's pathwise inequality [1], Proposition 2.1(i) (see also [13], Eqn. (5)), our pathwise method easily allows to recover also Doob's inequality for an arbitrary $\mathcal{F}_0$-measurable variable exponent $p > 1$, without requiring any of the additional hypotheses made in[5] [2] or in[6] [16].

It turns out that (2) hold not only when τ is a stopping time, but also for many random times. Indeed, if τ is a finite random time such that $(K \cdot X)_\tau$ and $(\tilde{K} \cdot X)_\tau$ are in L^1 and have zero expectation, trivially (1) implies (2). This can be useful: if

$$\mathbb{H}^1 := \{N \text{ is a martingale and } N^*_\infty \in L^1(\mathbb{P})\}, \tag{4}$$

given any random time τ the set of $M \in \mathbb{H}^1$ for which $\mathbb{E}[M_\tau] = 0$ is 'large' (it has co-dimension at most 1 in $\mathbb{H}^1$) and can be quite explicitly characterized: see [26], Section 3. Moreover, perhaps surprisingly there are quite a number of interesting examples of random times τ (called pseudo stopping times) which are not stopping times and for which $\mathbb{E}M_\tau = 0$ holds for *any* $M \in \mathbb{H}^1$; these times have been studied in [28], where one can find several equivalent characterizations and examples.

If τ is a finite[7] stopping time and $A_t := 1_{[\tau,\infty)}(t)$, then $f(\tau) = \int_0^\infty f(s)\,dA_s$; it is then natural to ask if one can generalize (2) and obtain that

$$c\mathbb{E}\int_0^\infty \Phi\big(\sqrt{[X]_s}\big)\,dA_s \le \mathbb{E}\int_0^\infty \Phi\big(X^*_s\big)\,dA_s \le C\mathbb{E}\int_0^\infty \Phi\big(\sqrt{[X]_s}\big)\,dA_s \tag{5}$$

holds for any local martingale X, increasing adapted A with $A_\infty = 1$ and general Φ. It turns out that this is true and simple to prove, although (perhaps surprisingly) this follows not integrating the pathwise BDG inequalities but rather considering the classical ones on an enlarged space; this observation is probably not new, although we include it since were not able to locate a reference in the literature.

One can ask whether the BDG inequalities hold not only for local martingales, but also for many semimartingales which admit an equivalent local martingale measure; the latter processes being of particular importance in mathematical finance, due to the Fundamental Theorem of Asset Pricing (see [10]). As proved in [33,34], one can exactly characterize the equivalent measures under which the so called 'weighted BDG inequalities' hold. Indeed, given $\hat{\mathbb{P}} \sim \mathbb{P}$ let $Z_t := \mathbb{E}[d\hat{\mathbb{P}}/d\mathbb{P}|\mathcal{F}_t]$, so that[8] $Z_t = \exp(M_t - [M]_t/2)$ for a unique local martingale M with

[4] We use the modular [23], Eqn. (1.1), instead of the related Luxemburg norm [23], Eqn. (1.2),

[5] I.e. the probability space to be without atoms, p to be bounded away from 1 and bounded above, $\mathcal{F}_0$ to be countably generated.

[6] I.e. $\mathcal{F}_n$ to be countably generated for every n, p to be bounded away from 1 and bounded above and to satisfy the inequality [16], Eqn. (1.4).

[7] This is not really needed, as we will see.

[8] See [15], Chapter 2, Theorem 8.3, and [31], Chapter 8, Proposition 1.6; the fact that M is a local martingale follows from $dM = (Z_-)^{-1}\,dZ$, since Z is a martingale.

$M_0 = 0$. Then, as one can read in [19], Theorem 3.17 and 3.18, if the underlying filtration is such that every $\mathbb{P}$-martingale is continuous, $M \in \mathrm{BMO}(\mathbb{P})$ iff there exist c, C such that

$$c\hat{\mathbb{E}}\sqrt{[X]_\infty} \le \hat{\mathbb{E}}X^*_\infty \le C\hat{\mathbb{E}}\sqrt{[X]_\infty} \tag{6}$$

holds for every local $\mathbb{P}$-martingale X. While we cannot use the pathwise BDG inequalities to obtain with a simple proof the above extremely satisfying result in complete generality, we can easily prove a weaker statement which does not require any knowledge about BMO-martingales.

Then, we briefly discuss what happens to the pathwise and the standard BDG inequalities (and their optimal constants) in higher dimension (finite and infinite).

The outline of the rest of the paper is then as follows. In Section 2, we introduce most of the notations and we derive the pathwise Davis inequalities for cadlag semimartingales from their discrete time version. In Section 3, we state the pathwise BDG inequalities for cadlag semimartingales; we relegate the corresponding proof to the appendix, as it is computationally demanding. In Section 4, we prove the BDG inequalities for the Bessel processes. In Section 5 we consider the case of a random exponent p. In Section 6, we show that the BDG inequalities hold for martingales stopped at many random times, and in Section 7, we discuss (5). In Section 8, we consider what happens after a change of measure. In Section 9, we discuss the multidimensional case and the optimal constants. In Section 10, we show that Davis pathwise inequalities for *continuous* semimartingales can be given an alternative proof, based only on Ito's formula.

2. Pathwise Davis inequalities for cadlag semimartingales

In this section, we introduce most of the notations used throughout the paper, and then we easily obtain a version of the pathwise Davis inequalities for cadlag semimartingales by passing to the limit their discrete time version.

We will work on an underlying filtered probability space $(\Omega, \mathbb{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ whose filtration $(\mathcal{F}_t)_t$ satisfies the usual conditions.[9] Given cadlag adapted processes S, X, A, and assuming that X is a semimartingale, A is of finite variation (on compact sets) and the following integrals exist, we will use the following notations. The cag predictable process S_- has value $S_{t-} := \lim_{u\uparrow t} S_u$ at time t, the jump of S at t is $\Delta S_t = S_t - S_{t-}$, the running maximum S^* of S is given by $S^*_t := \sup_{u\le t} |S_u|$, $[X]$ is the quadratic variation of X, A_∞ is the (possibly infinite) limit $\lim_{t\to\infty} A_t$ (which always exists if A is increasing), $(H\cdot X)_t$ is the stochastic integral $\int_{(0,t]} H_u\, dX_u$, and $\int_0^t H_u\, dA_u$ (resp. $\int_{0-}^{t-} H_u\, dA_u$) is the Lebesgue-Stieltjes integral $\int_{(0,t]} H_u\, dA_u$ (resp. $\int_{[0,t)} H_u\, dA_u$). Given arbitrary processes $K, \tilde{K}$ we will write that $K \le \tilde{K}$ if $K_t \le \tilde{K}_t$ holds $\mathbb{P}$ a.s. and for all $t \ge 0$, and we define by convention $K_{0-} := 0$ and $0/0 := 0$, so that in particular $X_{0-} = X^*_{0-} = [X]_{0-} = 0$, the integrands H in Theorem 1 and $H, G, F^{(s)}_t$ in Theorem 3 are well defined and the measure $dX^*_\cdot$ (resp. $d\sqrt{[X]_\cdot}$) has mass $X^*_0 = |X_0|$ (resp. $\sqrt{[X]_0} = |X_0|$) at 0. We denote with $a \vee b$ the maximum of $a, b \in \mathbb{R}$, and with $\mathbb{R}_+$ the interval $[0, \infty)$.

[9] Meaning it is right continuous and $\mathcal{F}_0$ contains all the negligible sets of $\mathcal{F}_\infty$.

Theorem 1. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a cadlag semimartingale, and set*

$$H_t := \frac{X_{t-}}{\sqrt{[X]_{t-} + (X^*_{t-})^2}}. \tag{7}$$

Then H is cag, predictable, has values in $[-1, 1]$, and satisfies

$$\sqrt{[X]} \le 3X^* - (H \cdot X), \qquad X^* \le 6\sqrt{[X]} + 2(H \cdot X); \tag{8}$$

Proof. Applying [5], Theorem 1.2, to $x_i := X^t_{i/2^n}(\omega)$, $0 \le i \le N$ with $2^n t \le N$ gives

$$Q^n(X)_t := \sqrt{\sum_{i\in\mathbb{N}} \big(X^t_{(i+1)/2^n} - X^t_{i/2^n}\big)^2} \le 3M^n(X)_t - \big(H^n \cdot X\big)_t, \tag{9}$$

where $M^n(X)_t := \max_{i\in\mathbb{N}} |X^t_{i/2^n}|$ and

$$H^n_s := \sum_{i\in\mathbb{N}} 1_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(s) \frac{X_{\frac{i}{2^n}}}{\sqrt{Q^n(X)_{\frac{i}{2^n}} + (M^n(X)_{\frac{i}{2^n}})^2}}. \tag{10}$$

Since $Q^n(X)_s \to [X]_s$ uniformly on compacts in probability, passing to a subsequence (without relabeling) we get that $Q^n(X)_s \to [X]_s$ a.s. for all $s \ge 0$; since X is a.s. cadlag, $M^n(X)_s \to X^*_s$ a.s. and so[10] $H^n_s \to H_s$ a.s. for all $s \ge 0$. Since $|H^n| \le 1$, using the stochastic dominated convergence theorem we get that $(H^n \cdot X)_. \to (H \cdot X)_.$ uniformly on compacts in probability, so we can take limits in (9) and obtain that $\sqrt{[X]_t} \le 3X^*_t - (H \cdot X)_t$ a.s. for all $t \ge 0$. The proof of the second inequality is identical. □

The traditional Davis inequalities (12) are a simple corollary of their pathwise equivalent (8).

Corollary 2. *Under the assumptions of Theorem 1*

$$(H \cdot X)^* \le 3\big(X^* + \sqrt{[X]}\big) \quad \textit{and} \quad [H \cdot X] \le [X]. \tag{11}$$

If X is a local martingale, then so is $(H \cdot X)$, and

$$\mathbb{E}\sqrt{[X]_\infty} \le 3\mathbb{E}X^*_\infty, \qquad \mathbb{E}X^*_\infty \le 6\mathbb{E}\sqrt{[X]_\infty}; \tag{12}$$

if moreover $\mathbb{E}\sqrt{[X]_\infty} < \infty$ then $(H \cdot X)$ is a martingale *and $(H \cdot X)^*_\infty \in L^1(\mathbb{P})$.*

Proof. It trivially follows from (8) that $-3\sqrt{[X]} \le (H \cdot X) \le 3X^*$; since $|H| \le 1$, $[H \cdot X] = H^2 \cdot [X] \le [X]$, (11) hold. Assume now that X is a local-martingale, in which case also $H \cdot X$ is a local martingales (because H is cag). Let τ_n be sequence of stopping times which localizes

[10] If $[X]_{s-} + (X^*_{s-})^2 = 0$ then $H^n_s = H_s = 0$.

$H \cdot X$; applying (8) to X^{τ_n} (instead of X), taking expectations and then taking limits for $t, n \to \infty$ we get (12) by monotone convergence. If $\mathbb{E}\sqrt{[X]_\infty}$ is finite (11) gives that $\mathbb{E}(H \cdot X)^*_\infty < \infty$, so the dominated convergence theorem ensures that the local martingale $H \cdot X$ is a martingale. □

3. Pathwise BDG inequalities for cadlag semimartingales

In this section, we modify ideas of Garsia to obtain that the general (pathwise) BDG inequalities are a consequence of Davis' ones; this approach was already taken in [5] in the (technically much simpler) discrete-time case when $\Phi(t) = t^p$ for some $p > 1$. We actually relegate to the Appendix the computationally-intense proof. For expositions of Garsia's ideas, we refer to [25], pages 101 to 106, or [9]; a slightly modified version[11] is given in [11], Chapter 7, Lemma 91.

Notice that we *do not* derive the general pathwise BDG inequalities passing to the limit the discrete time statement [5], Theorem 6.3, as done for Davis inequalities. The problem with this approach is that for $p > 1$ it is not easy to show that the discretized integrands H^n, G^n corresponding to (10) should converge to their continuous time equivalent, since if we try to express H^n_s as a stochastic integral we obtain that[12] $H^n_s = \int_0^s K^n_u \, dY^n_u$, where also the integrator Y^n depends on n (and similarly for G^n). In principle however our approach to proving pathwise inequalities through discretization can be applied also to the general BDG inequalities. Indeed, the results of [3] strongly suggest the existence of predictable integrands H_t, G_t, each of the form $f(X_{t-}, [X]_{t-}, X^*_{t-})$ for some *continuous* function f (which depends on Φ), for which the pathwise BDG inequalities hold in discrete time. If that is indeed the case, they can then be obtained in continuous time simply by passing to the limit (as done in the proof of Theorem 1).

We will henceforth consider a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ which is cadlag, increasing, unbounded and such that $\phi(0) = 0$; as usual we define $\phi(0-) := 0$, so that in particular $d\phi$ has no atom at zero. The integral $\Phi(t) := \int_0^t \phi(s) \, ds$ is a convex increasing function such that $\Phi(t)/t \to \infty$ as $t \to \infty$. We will also assume that $\Phi(t)$ is 'tame', that is, that there exists some constant C_Φ such that $\Phi(2t) \le C_\Phi \Phi(t)$ for all t; equivalently, there exists some constant c_ϕ such that $\phi(2t) \le c_\phi \phi(t)$ for all t. Such functions Φ are well studied in connections to Orlicz spaces, and are often called 'Young functions', although they are also referred to by various other names. In particular, the interested reader should consult [20], where Φ would be called an 'N-function satisfying the Δ_2-condition'. One can then show that the 'exponent[13]'

$$p := \sup_{u>0} \frac{u\phi(u)}{\Phi(u)}, \tag{13}$$

is in $(1, \infty)$. Moreover, if $\psi(t) := \inf\{s : \phi(s) > t\}$ is the cad (so $\psi_-(t) := \psi(t-)$ is the cag) inverse of ϕ, and $\Psi(t) := \int_0^t \psi(s) \, ds$ is the convex conjugate of Φ, the following inequalities

[11] The quantity $\mathbb{E}[\Phi(|X|)]$ is replaced by the seminorm $\|X\|_\Phi := \inf\{\lambda > 0 : \mathbb{E}[\Phi(|X|/\lambda)] \le 1\}$.

[12] Indeed $Y^n_s = \Phi(\sqrt{Q^n(X)_s})$.

[13] If $\Phi(t) = t^c/c$ then p in (13) equals c.

(which we borrow from [9]) hold:

$$
\begin{aligned}
uv \le \Phi(u) + \Psi(v), \qquad \Phi(au) \le a^p \Phi(u) \qquad \text{if } a \ge 1, \\
\Psi(au) \le a\Psi(u) \qquad \text{if } a \le 1, \\
\Psi(s) \le (p-1)\Phi\big(\psi(s-)\big) \quad \text{and} \quad \psi_-\big(\phi(s)\big) \le s \quad \text{and so} \quad \Psi\big(\phi(s)\big) \le (p-1)\Phi(s).
\end{aligned}
\tag{14}
$$

Theorem 3. *Assume that $(X_t)_{t\in\mathbb{R}_+}$ is a cadlag semimartingale, Φ is as above and p is given by (13). Let C (resp. D) be the cad inverse of $\sqrt{[X]}$ (resp. X^*) and define $c_p := p(6p)^p$,*

$$
H_t := \int_{[0,\sqrt{[X]}_{t-})} pF_t^{(C_s)}\, d\phi(s), \qquad G_t := \int_{[0,X^*_{t-})} pF_t^{(D_s)}\, d\phi(s),
$$

where $(F_t^{(s)})_{(s,t)\in\mathbb{R}^2_+ : s<t}$ is defined as

$$
F_t^{(s)} := \frac{X_{t-} - X_{s-}}{\sqrt{[X]_{t-} - [X]_{s-} + \sup_{s\le u<t}(X_u - X_{s-})^2}}.
$$

Then H, G are cag predictable and

$$
\Phi\big(\sqrt{[X]}\big) \le c_p \Phi\big(X^*\big) - (H\cdot X), \qquad \Phi\big(X^*\big) \le c_p\Phi\big(\sqrt{[X]}\big) + 2(G\cdot X). \tag{15}
$$

Moreover if ϕ is continuous, then H and G are lad and

$$
H_t = \int_{[0,t)} pF_t^{(s)}\, d\phi\big(\sqrt{[X]_s}\big), \qquad G_t = \int_{[0,t)} pF_t^{(s)}\, d\phi\big(X_s^*\big). \tag{16}
$$

If $\Phi(t) = t^p/p$, the inequalities (15) hold with the better[14] constant $c_p = (6p)^p$.

We will derive in (45) integral expressions for H_{t+} and G_{t+}; these show that if X and ϕ are continuous then also H and G are continuous. Notice that the function $\Phi(t) = t$ does not satisfy the assumptions made on Φ in Theorem 3; despite of this, Theorem 1 affords the equivalent of (15).

Of course, given a sequence of real numbers $(x_n)_{n\ge 0}$ and the probability space $\{\hat\omega\}$ made of one point, applying (15) (resp. (8)) to $X_t(\hat\omega) := \sum_n x_n 1_{[n,n+1)}(t)$ we obtain pathwise BDG inequalities for functions of a real variable, which if $\Phi(t) = t^p$ reduce to [5], Theorem 6.3, for $p>1$ (resp. to [5], Theorem 1.2, for $p=1$).

Moreover, if ϕ is continuous the integrands H and G in Theorem 3 are caglad, so (15) are really path-by-path inequalities. Indeed, for $\mathbb{P}$ a.e. ω one can compute $(H\cdot X)_\cdot(\omega)$ and $[X]_\cdot(\omega) := X^2(\omega) - (2X_-\cdot X)(\omega)$ only making use of $H_\cdot(\omega)$ and $X_\cdot(\omega)$ by taking limits of Riemann sums computed along appropriate sequences of hitting times (see[15] [6], Theorem 7.14,

[14] This fact, which may at first appear odd, corresponds to [9], Lemma 1, Eqn. (13').

[15] One can apply the cited theorem because $H^+ := (H_{t+})_t$ is adapted (since such is H, and the filtration is right continuous) and $H^+_{t-} = H_t$.

and [18]). Unfortunately, this remark does not apply to Davis inequalities (8), since H in Theorem 1 is not necessarily lad, not even if X is continuously differentiable.[16]

The traditional BDG inequalities are a simple corollary of the pathwise ones.

Corollary 4. *Under the assumptions of Theorem 3 for all* $y \geq 1$

$$\sqrt{[H \cdot X]} \leq p^2 \Phi\big(\sqrt{[X]}\big), \qquad \sqrt{[G \cdot X]} \leq (py)^p \Phi\big(\sqrt{[X]}\big) + \frac{p-1}{y} \Phi\big(X^*\big), \tag{17}$$

and if X *is a local-martingale then so are* $(H \cdot X)$ *and* $(G \cdot X)$, *and*

$$\mathbb{E}\Phi\big(\sqrt{[X]_\infty}\big) \leq c_p \mathbb{E}\Phi\big(X^*_\infty\big), \qquad \mathbb{E}\Phi\big(X^*_\infty\big) \leq c_p \mathbb{E}\Phi\big(\sqrt{[X]_\infty}\big). \tag{18}$$

In particular, if X *is a local-martingale and* $\mathbb{E}\Phi(\sqrt{[X]_\infty}) < \infty$ *then* $(H \cdot X)_t$ *and* $(G \cdot X)_t$ *are* martingales *and* $(H \cdot X)^*_\infty, (G \cdot X)^*_\infty \in L^1(\mathbb{P})$.

4. The Bessel process

As a corollary of the pathwise Davis inequalities, we will now prove BDG inequalities for the Bessel process for general Φ; notice how this turns out to be very easy in the case $\Phi(t) = t$ (i.e. Davis inequalities).

Theorem 5. *Let* X *be the Bessel process of dimension* $\alpha \in [1, \infty)$ *started at* $X_0 \geq 0$ *and* $\Phi(t) = t^p$ *for* $p > 0$, *then there exist constants* c, C *such that*

$$c\mathbb{E}\Phi\big(\sqrt{[X]_\tau}\big) \leq \mathbb{E}\Phi\big(X^*_\tau\big) \leq C\mathbb{E}\Phi\big(\sqrt{[X]_\tau}\big) \qquad \text{for all stopping times } \tau. \tag{19}$$

More generally, (19) *holds if* $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ *is cadlag, increasing, such that* $\Phi(x) = 0$ *iff* $x = 0$ *and for which* $\sup_{t>0} \Phi(\beta t)/\Phi(t) < \infty$ *for some (and thus all)* $\beta > 1$.

The only facts about X which we will need in the following proof is that $X_\cdot > 0$ $\mathbb{P} \otimes \mathcal{L}^1$ a.e. and X is a weak solution[17] of

$$dX = \frac{\alpha - 1}{2X} dt + dW, \tag{20}$$

where W is a standard Brownian motion w.r.t some underlying filtration $(\mathcal{F}_t)_t$. In fact, X is positive and it never hits 0 (after time zero) if $\alpha \geq 2$, whereas for $\alpha \in (0, 2)$ a.s. X hits zero but the set $\{s : X_s = 0\}$ has Lebesgue measure zero (for all these statements see [31], page 442). That X solves (20) is stated in [31], Chapter 11, Exercise 1.26, and if $\alpha \geq 2$ the proof is simple and can be found just before [31], Chapter 11, Proposition 1.10. Notice in particular that X has continuous paths and $[X]_t = X_0^2 + t$.

[16] If $X_t := t^2 \sin(1/t)$ for $t > 0$ and $X_0 := 0$, X is C^1 and thus has finite variation, so it is a semimartingale with $[X] = 0$, and $H_t = X_{t-}/X^*_{t-}$ keeps oscillating between 1 and -1 as $t \downarrow 0$.

[17] The explosion time of (20) is ∞, i.e. the solution to (20) is defined for all $t \in \mathbb{R}_+$.

Lemma 6. *Let X be a semimartingale such that $X_\cdot + C > 0$ $\mathbb{P} \otimes \mathcal{L}^1$ a.e. and*

$$dX_t = \frac{\gamma}{2(X_t + C)}\,dt + dW_t, \qquad X_0 \geq 0, \tag{21}$$

where W is a standard Brownian motion w.r.t some underlying filtration $(\mathcal{F}_t)_t$, $C \geq 0$ is a $\mathcal{F}_0$-measurable random variable and $\gamma > 0$. Then

$$\mathbb{E}\big(X^*_\tau|\mathcal{F}_0\big) \leq (6+2\gamma)\mathbb{E}\big(\sqrt{[X]_\tau}|\mathcal{F}_0\big) \tag{22}$$

and if $C = 0$, then

$$\mathbb{E}\big(\sqrt{[X]_\tau}|\mathcal{F}_0\big) \leq 3\mathbb{E}\big(X^*_\tau|\mathcal{F}_0\big). \tag{23}$$

Proof. Taking $H_t := X_t/\sqrt{[X]_t + (X^*_t)^2}$, since $|H| \leq 1$ we get that $\mathbb{E}\int_0^t H_s^2\,ds \leq t < \infty$, so $H \cdot W$ is a martingale and (21) gives

$$\mathbb{E}\big((H \cdot X)_{\tau\wedge t}|\mathcal{F}_0\big) = \mathbb{E}\bigg(\int_0^{\tau\wedge t} \frac{\gamma}{2\sqrt{[X]_s + (X^*_s)^2}} \frac{X_s}{X_s + C}\,ds\Big|\mathcal{F}_0\bigg). \tag{24}$$

In particular, since $[X]_t = X_0^2 + t$ and $X_\cdot/(X_\cdot + C) \leq 1$ $\mathbb{P} \otimes \mathcal{L}^1$ a.e., deleting the positive X^* term from (24) we can bound $\mathbb{E}((H \cdot X)_{\tau\wedge t}|\mathcal{F}_0)$ from above with

$$\mathbb{E}\bigg(\int_0^{\tau\wedge t} \frac{\gamma}{2\sqrt{X_0^2 + s}}\,ds\Big|\mathcal{F}_0\bigg) = \gamma\Big(\mathbb{E}\Big(\sqrt{X_0^2 + \tau \wedge t}\Big|\mathcal{F}_0\Big) - |X_0|\Big) \leq \gamma\mathbb{E}\big(\sqrt{[X]_{\tau\wedge t}}|\mathcal{F}_0\big).$$

Now evaluate the second inequality (8) at time $\tau \wedge t$, take $\mathbb{E}(\cdots|\mathcal{F}_0)$, apply the bound we proved for $\mathbb{E}((H \cdot X)_{\tau\wedge t}|\mathcal{F}_0)$ and take limits[18] as $t \to \infty$ to get (22). If $C = 0$, then $X_\cdot = X_\cdot + C > 0$ $\mathbb{P} \otimes \mathcal{L}^1$ a.e., so trivially from (24) we get the bound $\mathbb{E}((H \cdot X)_{\tau\wedge t}|\mathcal{F}_0) \geq 0$, so (23) follows evaluating (8) at time $\tau \wedge t$ and taking $\lim_{t\to\infty} \mathbb{E}(\cdots|\mathcal{F}_0)$. □

Proof of Theorem 5. The case $\alpha \in \mathbb{N} \setminus \{0\}$ follows from the analogous statement for Brownian motion (for which we refer to [22], page 37, for general[19] Φ and dimension one; for higher dimension see our Section 9): let us show this in detail. If $n \in \mathbb{N} \setminus \{0\}$, B is a n-dimensional Brownian motion started at B_0 and $X := \|B\|_{\mathbb{R}^n}$ then $[X]_t = \|B_0\|^2_{\mathbb{R}^n} + t$ and $[B]_t = \|B_0\|^2_{\mathbb{R}^n} + nt$, so $[B]/n \leq [X] \leq [B]$. Thus, since $X^*_t = \sup_{s\leq t} \|B_s\|_{\mathbb{R}^n}$, the BDG inequalities applied to B imply those for X.

From now on, we can then assume $\alpha > 1$. Since X solves (20), we can apply Lemma 6 to X with $C = 0$ and $\gamma = \alpha - 1$, and taking expectations gives the thesis for $\Phi(t) = t$. So, let us consider the case of general Φ and $\alpha > 1$. If $\sigma \leq \hat{\sigma}$ are *finite* stopping times, then $\hat{W}_t :=$

[18] Use the monotone convergence theorem.

[19] The statement in the special case $\Phi(t) = t^p$ can be found in most books on stochastic calculus

$W_{\sigma+t} - W_\sigma$ (resp. $\hat\sigma - \sigma$) is a $\hat{\mathcal{F}}_t := \mathcal{F}_{\sigma+t}$ Brownian motion (resp. stopping time), and $\hat X_t := X_{\sigma+t}$ trivially satisfies $d\hat X = \frac{\alpha-1}{2\hat X}\,dt + d\hat W$. We can then apply (23) with $(X, W, \mathcal{F}, \tau, C, \gamma) := (\hat X, \hat W, \hat{\mathcal{F}}, \hat\sigma - \sigma, 0, \alpha - 1)$ and combine this with the bounds $\hat X^*_t \le X^*_{\sigma+t}$ and

$$\sqrt{[X]_{\sigma+t}} - \sqrt{[X]_\sigma 1_{\{\sigma>0\}}} \le \sqrt{[X]_{\sigma+t} - [X]_\sigma 1_{\{\sigma>0\}}} = \sqrt{[\hat X]_t - X_\sigma^2 1_{\{\sigma>0\}}} \le \sqrt{[\hat X]_t}$$

to obtain

$$\mathbb{E}\big(\sqrt{[X]_{\hat\sigma}} - \sqrt{[X]_\sigma}1_{\{\sigma>0\}}|\mathcal{F}_\sigma\big) \le 3\mathbb{E}\big(\hat X^*_{\hat\sigma-\sigma}|\hat{\mathcal{F}}_0\big) \le 3\mathbb{E}\big(X^*_{\hat\sigma}|\mathcal{F}_\sigma\big). \tag{25}$$

Define the localizing sequence $\sigma_n := \inf\{t \ge 0 : X^*_t + [X]_t \ge n\} \wedge n$ and, given stopping times $\tau, \hat\tau, \theta$ with $\tau \le \hat\tau$, define $\hat\sigma := \sigma_n \wedge \theta \wedge \hat\tau$, $\sigma := \sigma_n \wedge \theta \wedge \tau$. Since $\sigma \le \hat\sigma < \infty$ and $\sqrt{[X^{\sigma_n\wedge\theta}]_\tau}1_{\{\sigma>0\}} \le \sqrt{[X^{\sigma_n\wedge\theta}]_\tau}1_{\{\tau>0\}}$ we can apply (25) to get

$$\mathbb{E}\Big(\sqrt{\big[X^{\sigma_n\wedge\theta}\big]_{\hat\tau}} - \sqrt{\big[X^{\sigma_n\wedge\theta}\big]_\tau}1_{\{\tau>0\}}|\mathcal{F}_\sigma\Big) \le 3\mathbb{E}\big(X^*_{\hat\sigma}|\mathcal{F}_\sigma\big).$$

Since $\sqrt{[X^{\sigma_n\wedge\theta}]_{\hat\tau}} - \sqrt{[X^{\sigma_n\wedge\theta}]_\tau}1_{\{\tau>0\}}$ and $X^*_{\hat\sigma}$ are $\mathcal{F}_{\sigma_n\wedge\theta}$ measurable and $\mathbb{E}(Y|\mathcal{F}_\sigma) = \mathbb{E}(\mathbb{E}(Y|\mathcal{F}_{\sigma_n\wedge\theta})|\mathcal{F}_\tau)$ for any positive r.v. Y, we obtain that

$$\mathbb{E}\Big(\sqrt{\big[X^{\sigma_n\wedge\theta}\big]_{\hat\tau}} - \sqrt{\big[X^{\sigma_n\wedge\theta}\big]_\tau}1_{\{\tau>0\}}|\mathcal{F}_\tau\Big) \le 3\mathbb{E}\big(X^*_{\hat\sigma}|\mathcal{F}_\tau\big).$$

Integrating this over $\{\hat\tau > \tau\} \in \mathcal{F}_\tau$ gives, since $\sqrt{[X^{\sigma_n\wedge\theta}]_{\hat\tau}}1_{\{\hat\tau>0\}} \le \sqrt{[X^{\sigma_n\wedge\theta}]_{\hat\tau}}$,

$$\mathbb{E}\Big(\sqrt{\big[X^{\sigma_n\wedge\theta}\big]_{\hat\tau}}1_{\{\hat\tau>0\}} - \sqrt{\big[X^{\sigma_n\wedge\theta}\big]_\tau}1_{\{\tau>0\}}\Big) \le 3\mathbb{E}\big(\big(X^{\sigma_n\wedge\theta}\big)^*_{\hat\tau}1_{\{\hat\tau>\tau\}}\big).$$

It follows from [22], Lemma 1.1, that for some constant c (depending only on Φ)

$$\mathbb{E}\Phi\big(\sqrt{[X]_{\sigma_n\wedge\theta}}\big) = \mathbb{E}\Phi\Big(\sqrt{\big[X^{\sigma_n\wedge\theta}\big]_\infty}\Big) \le c\mathbb{E}\Phi\big(X^{\sigma_n\wedge\theta}\big)^*_\infty = c\mathbb{E}\Phi\big(X^*_{\sigma_n\wedge\theta}\big),$$

and so taking $n \to \infty$ gives $\mathbb{E}\Phi(\sqrt{[X]_\theta}) \le c\mathbb{E}\Phi(X^*_\theta)$.

To prove the opposite inequality unfortunately, we cannot work analogously with $\hat X$, because we would need to use that $\sqrt{[\hat X]_t} \le \sqrt{[X]_{\sigma+t}}$, which is not true (indeed $[\hat X]_t = [X]_{\sigma+t} - [X]_\sigma + X_\sigma^2$). We consider instead[20] $Y_t := X_{\sigma+t} - X_\sigma 1_{\{\sigma>0\}}$, so that

$$dY = \frac{\alpha - 1}{2(Y + X_\sigma 1_{\{\sigma>0\}})}\,dt + d\hat W. \tag{26}$$

[20]Analogously we cannot consider Y instead of $\hat X$ in the previous proof, since we could only prove the corresponding inequality in Lemma 6 when $C = 0$.

Applying (22) with $(X, W, \mathcal{F}, \tau, C, \gamma) := (Y, \hat{W}, \hat{\mathcal{F}}, \hat{\sigma} - \sigma, X_\sigma 1_{\{\sigma>0\}}, \alpha - 1)$, and combining this with the bounds $\sqrt{[Y]_t} \leq \sqrt{[X]_{\sigma+t}}$ and

$$X^*_{\sigma+t} - X^*_\sigma 1_{\{\sigma>0\}} \leq Y^*_t$$

gives, for $m := (6 + 2(\alpha - 1))$,

$$\mathbb{E}\big(X^*_{\hat{\sigma}} - X^*_\sigma 1_{\{\sigma>0\}}|\mathcal{F}_\sigma\big) \leq m\mathbb{E}\big(\sqrt{[Y]_{\hat{\sigma}-\sigma}}|\hat{\mathcal{F}}_0\big) \leq m\mathbb{E}\big(\sqrt{[X]_{\hat{\sigma}}}|\mathcal{F}_\sigma\big), \tag{27}$$

which is the equivalent of (25) (with X^* and $\sqrt{[X]}$ reversed). The proof now continues exactly as above. □

5. Random exponent

We now prove some BDG like inequalities in the case in which the exponent $p \geq 1$ is a random variable. We work on a given (arbitrary) underlying filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$.

Theorem 7. *If τ is a stopping time, $p : \Omega \to [1, \infty)$ is $\mathcal{F}_\tau$-measurable and $c_p := (6p)^p$, then any cadlag local martingale $(X_t)_{t\geq 0}$ satisfies*

$$\mathbb{E}\Big[\sqrt{[X]^p_\infty} - \sqrt{[X]^p_{\tau-}}|\mathcal{F}_\tau\Big] \leq \mathbb{E}\big[c_p\big(X^*_\infty\big)^p|\mathcal{F}_\tau\big],$$

$$\mathbb{E}\big[\big(X^*_\infty\big)^p - \big(X^*_{\tau-}\big)^p|\mathcal{F}_\tau\big] \leq \mathbb{E}\Big[c_p\sqrt{[X]^p_\infty}|\mathcal{F}_\tau\Big].$$

Of course, after taking expectations Theorem 7 reduces to (3) when $\tau \equiv 0$. The idea behind the proof of the BDG inequalities with random exponent is easily understood in this sub-case; indeed, let g_n, h_n be the Borel functions of $(p, x_0, x_1, \ldots, x_n)$ provides us by the discrete time BDG inequality [5], Theorem 6.3. If p is $\mathcal{F}_0$-measurable and $(X_n)_n$ is a martingale, then $G_n := g_n(p, X_0, X_1, \ldots, X_n)$ is predictable and so $\mathbb{E}[(G \cdot X)_n] = 0$ (and analogously for H_n), and so (3) follow from the pathwise BDG inequalities. The actual proof is a little more complicated, mostly because one needs to prove that $(G \cdot X)_n$ is integrable.

To prove Theorem 7, we will need the following lemma, which generalizes [5], Theorem 6.3, from $m = 0$ to $m \in \mathbb{N}$ and affords the $p > 1$ equivalent of what [5], Lemma 6.1, is for $p = 1$; we will henceforth write

$$(h \cdot x)^n_i := \sum_{j=i}^{n-1} h_j(x_{j+1} - x_j).$$

Lemma 8. *Let $x_0, \ldots, x_{m+N}$ be real numbers, $p > 1$, $m, N \in \mathbb{N}$ and define*

$$h^{(m)}_n := \sum_{i=m}^{n} p^2\Big(\sqrt{[x]^{p-1}_i} - \sqrt{[x]^{p-1}_{i-1}}\Big) f^{(i)}_n, \qquad g^{(m)}_n := \sum_{i=m}^{n} p^2\big(\big(x^*_i\big)^{p-1} - \big(x^*_{i-1}\big)^{p-1}\big) f^{(i)}_n$$

for $n = m, \dots, m+N-1$, *where for* $i \in \mathbb{N}, i \le n$

$$f_n^{(i)} := \frac{x_n - x_{i-1}}{\sqrt{[x]_n - [x]_{i-1} + \max_{i \le k \le n}(x_k - x_{i-1})^2}}.$$

Then

$$\sqrt{[x]_{m+N}^p} - \sqrt{[x]_{m-1}^p} \le c_p\big(x_{m+N}^*\big)^p - \big(h^{(m)} \cdot x\big)_m^{m+N},$$

$$\big(x_{m+N}^*\big)^p - \big(x_{m-1}^*\big)^p \le c_p\sqrt{[x]_{m+N}^p} + 2\big(g^{(m)} \cdot x\big)_m^{m+N}.$$

Proof. It follows from [5], Theorem 1.2 and Lemma 6.1, that

$$\sqrt{[x]_{m+n}} - \sqrt{[x]_{m+i-1}} \le 6x_{m+n}^* - \big(f^{(m+i)} \cdot x\big)_{m+i}^{m+n}$$

which in terms of the tilded quantities

$$\tilde{a}_n := \sqrt{[x]_{m+n}}, \qquad \tilde{c}_n := 6x_{m+n}^*, \qquad \tilde{x}_n := x_{m+n}, \qquad \tilde{f}_n^{(i)} := f_{m+n}^{(m+i)}$$

reads $\tilde{a}_n - \tilde{a}_{i-1} \le \tilde{c}_n + (\tilde{f}^{(i)} \cdot \tilde{x})_i^n$. Applying[21] [5], Lemma 6.2, gives[22] the first inequality in the thesis, since trivially $\tilde{h}_j := \sum_{i=0}^j p^2(\tilde{a}_i^{p-1} - \tilde{a}_{i-1}^{p-1})\tilde{f}_j^{(i)}$ equals

$$\sum_{i=0}^{j} p^2\Big(\sqrt{[x]_{m+i}^{p-1}} - \sqrt{[x]_{m+i-1}^{p-1}}\Big) f_{m+j}^{(m+i)} = h_{m+j}^{(m)}$$

and $(\tilde{h} \cdot \tilde{x})_N = (h^{(m)} \cdot x)_m^{m+N}$. The second inequality is proved analogously. □

Proof of Theorem 7. We will prove only the first inequality, as the proof of the second one is completely analogous. Using the monotone convergence theorem, by localization we can assume that X is a martingale and is constant after time k. Exactly as when p is not random, one can easily[23] reduce the cadlag to the discrete time case[24] by passing to the limit as done in Lowther's blog [24].

Apply Lemma 8 on $\{p > 1\}$, and [5], Lemma 6.1, on $\{p = 1\}$, to obtain that

$$\sqrt{[X]_{\tau+N}^p} - \sqrt{[X]_{\tau-1}^p} \le c_p\big(X_{\tau+N}^*\big)^p - \big(H(p) \cdot X\big)_\tau^{\tau+N} \qquad \text{on } \{\tau < \infty\} \tag{28}$$

[21]The value $(p-1)^{p-1}$ in [5], Lemma 6.2, is incorrect, and should be replaced by p^p, which is what one gets applying Young's inequality $ab \le a^p/p + b^q/q$ with $a = c_n p^{1/q}$ and $b = a_n^{p-1}/p^{1/q}$ at the end of the proof of Lemma 6.2.
[22]More precisely, one needs to slightly generalize [5], Lemma 6.2, to bound $a_n^p - a_{-1}^p$ in the case where a_{-1} is not necessarily zero. This is achieved replacing the identity at the beginning of that proof with the following one: $a_n^p - a_{-1}^p = p(p-1)\int_{a_{-1}}^{a_n} s^{p-2}(a_n - s)\,ds$.
[23]The key is to be able to use the dominated convergence theorem. This is achieved through a localization procedure, using the fact that $\sqrt{[X]}^p$ is locally integrable iff $(X^*)^p$ is such.
[24]In which the term $\tau-$ is replaced by the term $\tau - 1$.

for some process $H(p)$ such that $H(p)_i$ is Borel-measurable function of p and $(X_k)_{k\le i}$; since p is $\mathcal{F}_\tau$-measurable, $H(p)_{\tau+i}$ is $\mathcal{F}_{\tau+i}$-measurable. Fix an arbitrary $A\in\mathcal{F}_\tau$, and on $\{\tau<\infty\}$ write $(H(p)\cdot X)_\tau^{\tau+N}$ as $\sum_{i=0}^{N-1}\Gamma_i\Delta_i$, where

$$\Delta_i := (X_{\tau+i+1}-X_{\tau+i})1_{A\cap\{\tau<\infty\}}, \qquad \Gamma_i := H(p)_{\tau+i}1_{A\cap\{\tau<\infty\}}.$$

If we show that each $\Gamma_i\Delta_i$ is in $L^1(\mathbb{P})$ and has zero expectation, from (28) we get

$$\mathbb{E}\Big[\Big(\sqrt{[X]^p_{\tau+N}}-\sqrt{[X]^p_{\tau-1}}\Big)1_{A\cap\{\tau<\infty\}}\Big] \le \mathbb{E}\big[c_p\big(X^*_{\tau+N}\big)^p 1_{A\cap\{\tau<\infty\}}\big] \tag{29}$$

and the proof is concluded taking the limit $N\to\infty$ (by monotone convergence). We will assume that $A\in\mathcal{F}_\tau$ is such that $c_p(X^*_{\tau+N})^p1_{A\cap\{\tau<\infty\}}$ is in $L^1(\mathbb{P})$: otherwise (29) is trivially satisfied and there is nothing to prove.

We can (and will) assume w.l.o.g. that p is bounded: indeed to obtain (29) for general p we can apply (29) to $p\wedge j$ and take limits as $j\to\infty$, using Fatou's lemma on the l.h.s and the dominated convergence theorem[25] on the r.h.s.

Since $(X^*_{\tau+N})^p1_{A\cap\{\tau<\infty\}}\in L^1(\mathbb{P})$ and 2^p is bounded, $|\Delta_i|^p\in L^1(\mathbb{P})$. Clearly, $\Gamma_i\Delta_i1_{\{p=1\}}\in L^1(\mathbb{P})$ since[26] $|\Gamma_i|\le 1$ and $|\Delta_i|=|\Delta_i|^p$ on $\{p=1\}$. To show that $\Gamma_i\Delta_i1_{\{p>1\}}\in L^1(\mathbb{P})$ apply on $\{p>1\}$ Young's inequality (where $q=p/(p-1)$)

$$|\Gamma_i\Delta_i|\le|\Gamma_i|^q/q+|\Delta_i|^p/p\le|\Gamma_i|^q+|\Delta_i|^p$$

and use that $|\Gamma_i|^q1_{\{p>1\}}\in L^1(\mathbb{P})$: indeed $H(p)$ corresponds, for $p>1$, to $h^{(m)}$ from Lemma 8, and since $|f_n^{(i)}|\le 1$ and p^2 is bounded, it is enough to show that

$$\Big(\sqrt{[X]^{p-1}_{\tau+i}}-\sqrt{[X]^{p-1}_{\tau+i-1}}\Big)^q 1_{A\cap\{\tau<\infty\}}1_{\{p>1\}}\in L^1 \qquad \text{for } i\le N, \tag{30}$$

which follows from $(X^*_{\tau+N})^p1_{A\cap\{\tau<\infty\}}\in L^1(\mathbb{P})$ since[27] $\sqrt{[X]_n}\le r_kX^*_n$ for all n and some r_k. Thus, $\Gamma_i\Delta_i\in L^1(\mathbb{P})$. Since $H(p)_{\tau+i}$ is $\mathcal{F}_{\tau+i}$-measurable, so is Γ_i; since $(X_{\tau+i})_i$ is a $(\mathcal{F}_{\tau+i})_i$-martingale $\mathbb{E}[\Delta_i|\mathcal{F}_{\tau+i}]=0$, and so

$$\mathbb{E}[\Gamma_i\Delta_i]=\mathbb{E}\big[\Gamma_i\mathbb{E}[\Delta_i|\mathcal{F}_{\tau+i}]\big]=0. \qquad \square$$

6. Random times

In this section, we discuss for which random times τ and martingales M do the BDG inequalities hold. We recall that $\mathbb{H}^1$ is defined in (4) and a random time τ is called a *pseudo stopping time* if $\mathbb{E}M_\tau=0$ holds for any $M\in\mathbb{H}^1$.

[25] As domination one can use that $((p\wedge j)6X^*_\infty)^{p\wedge j}1_A\le(1+(p6X^*_\infty)^p)1_A\in L^1$, which holds since $p\mapsto y^p$ is increasing if $y\ge1$, and $y^p\le1$ if $0<y\le1$.

[26] Indeed for $p=1$ the integrand $H(p)$ corresponds to $f^{(i)}$ given by [5], Lemma 6.1, which is bounded by 1.

[27] Indeed $(X_i)_{i\ge0}$ is a process constant after time k and, since on $\mathbb{R}^{k+1}$ all norms are equivalent, there exists a constant r_k such that $\sqrt{[x]_k}\le r_kx^*_k$ for all $(x_0,\dots,x_k)\in\mathbb{R}^{k+1}$.

From Theorem 3 and Corollary 4 (resp. Theorem 1 and Corollary 2), it follows that the BDG inequalities (2) hold for any pseudo stopping time τ and local martingale X if Φ is as in Theorem 3 (resp. if $\Phi(t) = t$); to see this, one first has to localize X so as to make $H \cdot X$ and $G \cdot X$ in $\mathbb{H}^1$, then take expectations and then limits (using the monotone convergence theorem).

Although the above extension to pseudo stopping times had already been proved in [27], Proposition[28] 2 with change of filtration techniques, it is convenient that it follows *immediately* from the pathwise inequalities; moreover, as mentioned in the introduction, we are able to obtain yet another setting in which (2) holds, and this seems to be new. Indeed, one can go the other way around and, given an arbitrary random time τ, study the subspace $\mathcal{S}_1(\tau)$ of $M \in \mathbb{H}^1$ for which $\mathbb{E}M_\tau = 0$. The above discussion shows that if[29] $H \cdot X \in \mathcal{S}_1(\tau)$, where X is a local martingale and H is as in Theorem 1, then (2) hold for $\Phi(t) = t$; analogously for H, G from Theorem 3 and a correspondingly general Φ. As we already said, this can be useful since $\mathcal{S}_1(\tau)$ is 'large' (it has co-dimension at most 1 in $\mathbb{H}^1$) and can be quite explicitly characterized: see [26], Section 3. This works out particularly well when τ is an honest time; for example, one can show that if $(\mathcal{F}_t)_t$ is the filtration generated by a one dimensional Brownian motion B and $\tau := \sup\{t < \sigma : B_t = 0\}$, where σ is the first time B hits 1, then τ is an honest time and if $(L_t)_t$ denotes the local time at zero of B and $\mathcal{L}^n(x)$ is the Laguerre polynomial $\frac{e^x}{n!}\frac{d^n}{dx^n}(x^n e^{-x})$ then $M_t := \mathbb{E}[\mathcal{L}_n(L_\sigma)|\mathcal{F}_t]$ is in $\mathcal{S}_1(\tau)$ whenever $n \neq 1$: see [26], Example 3.7.

7. Randomized stopping times

In this section, we prove that for Φ as in[30] Theorem 3 (and also for $\Phi(t) = t$) we have (5) for any local martingale X and randomized stopping time A; but first, we need some more definitions. We say that A is a randomized stopping time if it is a cadlag increasing adapted process with $A_0 = 0$ and $\lim_{t\to\infty} A_t \le 1$. Breaking from our conventions, in this section we allow A to have a jump at infinity: we will write $A_{\infty-}$ for $\lim_{t\to\infty} A_t$, and we define $A_\infty := 1$ (however X^*_∞ is defined as usual as $\lim_{t\to\infty} X^*_t$, and analogously for $[X]_\infty$) and if $f : [0,\infty] \to [0,\infty]$ is Borel then $\int_0^\infty f(s)\,dA_s := \int_{(0,\infty]} f(s)\,dA_s$. If τ is a stopping time (not necessarily finite) $A_t := 1_{[\tau,\infty)}(t)$ is a randomized stopping time, and $f(\tau) = \int_0^\infty f(s)\,dA_s$.

Let us now prove (5); as we just said, the integrals are over $(0,\infty]$. If the underlying space is $(\Omega, (\mathcal{F}_t)_t, \mathbb{P})$, consider the enlargement $\bar\Omega := \Omega \otimes [0,1]$ endowed with the product probability $\bar{\mathbb{P}} := \mathbb{P} \otimes \mathcal{L}^1$, and the usual augmentation $(\bar{\mathcal{F}}_t)_t$ of the filtration $(\mathcal{F}_t \otimes \mathcal{B})_t$ (where $\mathcal{L}^1$ denotes the Lebesgue measure and $\mathcal{B}$ the Borel subsets of $[0,1]$). If τ_n is a localizing sequence for X and Y a process on Ω, we extend them to $\bar\Omega$ by setting for all t

$$\bar\tau_n(\omega, s) := \tau_n(\omega), \qquad \bar Y_t(\omega, s) := Y_t(\omega) \qquad \text{for all } \omega \in \Omega, s \in [0,1].$$

Let $C(\omega, s) := \inf\{t : A_t(\omega) > s\}$ be the cad inverse of A; then trivially $\bar\tau_n$ are $(\bar{\mathcal{F}}_t)_t$ stopping times localizing the $(\bar{\mathcal{F}}_t)_t$ local martingale $\bar X$, and C is a $(\bar{\mathcal{F}}_t)_t$ stopping time since $\{C < t\} =$

[28] This has a typo: p should be ≥ 1. Only if M is continuous one can take any $p > 0$.

[29] Because of Corollary 2 it is clear that $H \cdot X$ is in $\mathbb{H}^1$ if one (and thus both) of the quantities in (12) are finite; what it not clear is whether $\mathbb{E}(H \cdot X)_\tau = 0$.

[30] If X is continuous then one can even take Φ as in Theorem 5

$\{(\omega, s) : s < A_{t-}(\omega)\}$ is in $\mathcal{F}_t \otimes \mathcal{B}$. In particular, we can apply the BDG inequalities (12) and (18) to the local martingale $\bar{X}$ stopped at C, and then use the change of time formula

$$\bar{\mathbb{E}}\bar{Y}_C := \int \bar{Y}_C \, d\bar{\mathbb{P}} = \mathbb{E}\int_0^1 Y_{C_s}\, ds = \mathbb{E}\int_0^\infty Y_s \, dA_s$$

with $Y = \Phi(\sqrt{[X]})$ and $Y = \Phi(X^*)$ to obtain (5).

8. Change of measure

As we mentioned, the proof of the equivalence $M \in \mathrm{BMO}(\mathbb{P}) \Leftrightarrow$ (6) is not simple; in this section we show how to easily get the following weaker[31] statement.

Theorem 9. *For every $s, T \geq 0$ there exist c, C such that*

$$c\mathbb{E}\sqrt{[X]_\tau} \leq \mathbb{E}X^*_\tau \leq C\mathbb{E}\sqrt{[X]_\tau} \tag{31}$$

holds for all stopping times τ and X satisfying[32] $dX = \sigma X(dW + \mu\, dt)$ *on* $[0, \tau]$ *for some predictable σ, μ such that $|\mu|$ and $|\sigma\mu|$ are bounded by s and $\mu = 0$ on (T, ∞).*

Proof. Since X satisfies $dX = \sigma X(dW + \mu\, dt)$, X is continuous and $\int_0^{t\wedge\tau} \sigma^2 X^2 \, ds < \infty$ a.s. for all $t < \infty$ (otherwise $\int_0^t \sigma X\, dW$ is not defined on $[0, \tau]$). Thus, if $H_t := X_t/\sqrt{[X]_t + (X^*_t)^2} \in [-1, 1]$, $M_t := \int_0^{t\wedge\tau} H\sigma X\, dW$ is well defined and a local martingale, and so there exist a localizing sequence $(\tau_n)_n$ for it and we get

$$\left|\mathbb{E}(H\cdot X)_{\tau_n\wedge\tau}\right| = \left|\mathbb{E}\int_0^{\tau_n\wedge\tau} H\sigma X\mu\, dt\right| \leq s\mathbb{E}\int_0^{T\wedge\tau} |X|\, dt \leq sT\mathbb{E}X^*_\tau,$$

and thus applying (8), localizing, taking expectations and passing to the limit we conclude $\mathbb{E}\sqrt{[X]_\tau} \leq (3 + sT)\mathbb{E}X^*_\tau$. Analogously since by Holder inequality

$$\left|\mathbb{E}\int_0^{\tau_n\wedge\tau} H\sigma X\mu\, dt\right| \leq s\mathbb{E}\int_0^\infty 1_{[0,T]}\big(1_{[0,\tau]}|\sigma X|\big)\, dt \leq s\mathbb{E}\sqrt{T}\sqrt{\int_0^\tau \sigma^2 X^2\, dt}$$

and $\int_0^\tau \sigma^2 X^2\, dt = [X]_\tau$, we can conclude that $\mathbb{E}X^*_\tau \leq (6 + 2s\sqrt{T})\mathbb{E}\sqrt{[X]_\tau}$. □

The reader may be interested in knowing that there are several conditions equivalent to (6) which one can impose on Z: the reverse Holder inequality, the Muckenhoupt (A_p) condition

[31] In this statement the role of $\mathbb{P}$ and $\hat{\mathbb{P}}$ is reversed, and $\hat{\mathbb{P}}$ is chosen so that $d\hat{W} := dW + \mu\, dt$ is a $\hat{\mathbb{P}}$-Brownian motion. Given $\hat{M} := \mu \cdot \hat{W}$, the local $\hat{\mathbb{P}}$-martingale $\exp(\hat{M} - [\hat{M}]/2)$ is in $\mathrm{BMO}(\hat{\mathbb{P}})$ since it has bounded quadratic variation, so this theorem is a special case of the result above.

[32] We are not assuming that this SDE has solution with explosion time strictly bigger than τ: this will depend on σ. We are saying that, *if* $(X_t)_{t\in\mathbb{R}_+}$ is a such a solution, then (31) holds.

and yet another unnamed (B_p) condition (see Theorem 3.4, Corollary 3.4 and Theorem[33] 2.4 in [19]). Moreover, in [7] one can find (under additional conditions) an extension of the implication $M \in \mathrm{BMO}(\mathbb{P}) \Rightarrow$ (6) in the case where not every martingale is continuous.

9. Higher dimension

The pathwise BDG inequalities (8) and (15), so far stated and proved in dimension 1, automatically hold in any finite dimension (with worse constants). Indeed, since on $\mathbb{R}^n$ all norms are equivalent, if $\|\cdot\|_{\mathbb{R}^n}$ denotes the Euclidian norm there exist $0 < \alpha_n < \beta_n$ such that for every n-dimensional semimartingale $X = (X^i)_{i=1,\dots,n}$

$$\alpha_n \sum_{i=1}^{n} \left(X^i\right)_t^* = \alpha_n \sum_{i=1}^{n} \sup_{s \le t} \left|X_s^i\right| \le X_t^* := \sup_{s \le t} \|X\|_{\mathbb{R}^n} \le \beta_n \sum_{i=1}^{n} \left(X^i\right)_t^*.$$

Thus, the fact that Davis inequalities (8) hold for X follows summing over i the corresponding inequalities for X^i (since $[X] = \sum_i [X^i]$ and $(H \cdot X) := \sum_i (H^i \cdot X^i)$). Similarly one obtains (15), using also the fact that for all $t_i \ge 0$

$$\frac{1}{n} \sum_{i=1}^{n} \Phi(t_i) \le \Phi\left(\sum_{i=1}^{n} t_i\right) \le n^{p-1} \sum_{i=1}^{n} \Phi(t_i);$$

these inequalities hold since Φ is increasing, convex and satisfies $\Phi(nt) \le n^p \Phi(t)$.

We should say that although throughout the paper we make no effort to get good constants, the results of [3] imply that the optimal constants for the pathwise BDG inequalities for $\mathbb{R}^n$-valued semimartingales are equal to (or smaller than[34]) the optimal constants for the classical BDG inequalities $\mathbb{R}^n$-valued martingales; the problem of finding these is important and still mostly[35] open.

Unfortunately, all this falls short of what one can do with the classic BDG inequalities, for which one can not only apply the above reasoning, but also prove that automatically they hold for every martingale M with values in a Hilbert space $\mathbb{H}$, and with the same constant as for $\mathbb{R}^2$. In fact, one can easily construct (possibly on an enlarged probability space) a $\mathbb{R}^2$-valued martingale N such that $\|M_t\|_{\mathbb{H}} = \|N_t\|_{\mathbb{R}^2}$ and $[M]_t = [N]_t$ for all $t \ge 0$. For the simple proof of this nice yet not so well-known result of [17] in the discrete time case, see [21], Proposition 5.8.3. For the general (much harder) cadlag case, one can consult [17]; notice however that one does *not* need this to prove the BDG inequalities for cadlag martingales, as these follow from their discrete time version!

[33] This theorem has a typo, it should be $p > 1$ not $p \ge 1$.
[34] We expect of course the optimal constants to be equal; this would easily follow if one could show that the stochastic integral term is $\mathbb{P}$-integrable.
[35] If $\Phi(t) = t^p/p$ for some values of p the optimal value of c or C is known, see [29].

10. Davis inequality for continuous local martingales

We give here an alternative statement and derivation of pathwise *Davis* inequalities for *continuous* semimartingales; the following treatment builds on [5], Theorem 5.1, where the easier of the two inequalities was proved for continuous local martingales starting at zero. Unlike the discrete time proof, the proof in this section has the advantage of being a relatively straightforward application of Ito's formula; this has an early precedent in the proofs of some subcases of BDG inequalities found in [31], Chapter 4, Proposition 4.3 and 4.4. One could probably analogously prove pathwise BDG inequalities for continuous semimartingales and $\Phi(t) = t^p/p$ for any $p \geq 1$; we leave this pursuit to other researchers. We remind the reader that the classical BDG inequalities also admit a very simple proof in the case of *continuous* martingale: see [32], Chapter 6, Theorem 42.1.

Theorem 10. *If X is a continuous semimartingale, then $\mathbb{P}$ a.s. for all $t \geq 0$*

$$X_t^* - 4\sqrt{[X]_t} \leq \left(\frac{2X_t}{\sqrt{[X]_t} \vee X_t^*} \cdot X_t \right)_t \leq 3X_t^* - 2\sqrt{[X]_t}.$$

Notice that the functional form of the integrand in Theorem 10 is slightly different[36] from the one obtained in Theorem 1. The next lemma is just a slight modification of the arguments after equation (4.3) in [5].

Lemma 11. *Let $f, g : \mathbb{R}^+ \to \mathbb{R}^+$ be continuous increasing and such that $f(0) \vee g(0) > 0$. Then on $\mathbb{R}^+$*

$$2g - 3f \leq \frac{g^2}{f \vee g} + \int_0^{\cdot} \frac{g^2}{f^2 \vee g^2}\, d(f \vee g) - \int_0^{\cdot} \frac{1}{f \vee g}\, df^2 \leq 3g - 2f. \tag{32}$$

Proof. Since $d\frac{1}{x} = -\frac{1}{x^2}\, dx$, by a change of variables $\int \frac{g^2}{f^2 \vee g^2}\, d(f \vee g) = \int g^2\, d\frac{-1}{f \vee g}$ and integrating the latter by parts we obtain that the middle term in (32) equals

$$\int_0^{\cdot} \frac{d(g^2 - f^2)}{f \vee g}. \tag{33}$$

As easily shown with the arguments[37] after equation (4.3) in [5], (33) is always smaller than $3g - 2f$. Applying this inequality with the role of f and g reversed and then multiplying by -1 we see that (33) is bigger than $-(3f - 2g)$. □

[36] We conjecture that [5], Theorem 1, from which Theorem 1 follows, also holds with the integrand $h_n = x_n / (\sqrt{[x]_n} \vee x_n^*)$ and potentially different constants, although proving this would require much longer computations.
[37] Unlike [5], our $f \vee g$ is strictly positive; however this has only the effect of slightly simplifying the calculations (one can consider directly $\int_0^t$ instead of $\lim_{\delta \to} \int_\delta^t$).

Proof of Theorem 10. For $\varepsilon \geq 0$ define $H^\varepsilon := \frac{2X}{\sqrt{[X]+\varepsilon} \vee X^*}$ and $I_t^\varepsilon := (H^\varepsilon \cdot X)_t$. When $\varepsilon > 0$ applying Ito's formula to X^2 and $1/(\sqrt{[X]+\varepsilon} \vee X^*)$ we find

$$d\frac{X^2}{\sqrt{[X]+\varepsilon} \vee X^*} = \frac{-X^2 d(\sqrt{[X]+\varepsilon} \vee X^*)}{(\sqrt{[X]+\varepsilon} \vee X^*)^2} + \frac{2X\,dX + d[X]}{\sqrt{[X]+\varepsilon} \vee X^*}.$$

In other words for $\varepsilon > 0$ the integral $I_t^\varepsilon = (H^\varepsilon \cdot X)_t$ equals[38]

$$\frac{X_s^2}{\sqrt{[X]+\varepsilon} \vee X_s^*}\Bigg|_{s=0}^{s=t} + \int_0^t \frac{X^2 d(\sqrt{[X]+\varepsilon} \vee X^*)}{(\sqrt{[X]+\varepsilon} \vee X^*)^2} - \int_0^t \frac{d([X]+\varepsilon)}{\sqrt{[X]+\varepsilon} \vee X^*}. \tag{34}$$

Since $X_0^2 = X_0^{*2}$, the quantity in (34) trivially gets bigger if we replace each occurrence of X by X^*, so applying Lemma 11 to $f(t) = \sqrt{[X]_t + \varepsilon}$, $g(t) = X_t^*$ gives

$$I^\varepsilon \leq 3X^* - 2\big(\sqrt{[X]+\varepsilon}\big).$$

To pass to the limit as $\varepsilon \to 0$ notice that if $\sqrt{[X]_t} \vee X_t^* = 0$ then $X_t = 0$, so $H_t^\varepsilon = 0$ and $H_t^0 = 0$ (since by our definition $0/0 = 0$); if instead $\sqrt{[X]_t} \vee X_t^* > 0$ then $H_t^\varepsilon \to H_t^0$ is trivially true. In summary, the stochastic dominated convergence theorem gives that $I^\varepsilon \to I^0$ uniformly on compacts in probability as $\varepsilon \to 0$, so there exists some $\varepsilon_n \to 0$ for which $I_t^{\varepsilon_n} \to I_t^0$ a.s. for all t, proving that $I^0 \leq 3X^* - 2\sqrt{[X]}$.

To prove the opposite inequality, replace X with X^* in (34) in both occurrences, and call J^ε the resulting quantity; then Lemma 11 applied to $f = \sqrt{[X]+\varepsilon}$, $g = X^*$ yields $J^\varepsilon \geq 2X^* - 3(\sqrt{[X]+\varepsilon})$. The thesis $I^0 \geq X^* - 4\sqrt{[X]}$ then follows taking $\varepsilon_n \to 0$ as above if we can show that $I^\varepsilon - J^\varepsilon \geq -X^* - \sqrt{[X]+\varepsilon}$. To prove the latter, let us bound separately the two terms C^ε, D^ε whose sum gives $I^\varepsilon - J^\varepsilon$. First,

$$C^\varepsilon := \frac{X_t^2 - X_t^{*2}}{\sqrt{[X]_t + \varepsilon} \vee X_t^*} \geq \frac{-X_t^{*2}}{\sqrt{[X]_t + \varepsilon} \vee X_t^*} \geq -X_t^*.$$

For the second term D^ε, notice that if f, g are continuous increasing then

$$d(f \vee g) = 1_{\{f \geq g\}}\, df + 1_{\{f < g\}}\, dg; \tag{35}$$

applying this to $f = \sqrt{[X]+\varepsilon}$ and $g = X^*$, and using that $X_t^2 = (X_t^*)^2$ holds[39] for dX_t^* a.e. t, we get that D^ε equals

$$\int_0^t \frac{X^2 - X^{*2}}{(\sqrt{[X]+\varepsilon} \vee X^*)^2} d\big(\sqrt{[X]+\varepsilon} \vee X^*\big) = \int_0^t \frac{X^2 - X^{*2}}{[X]+\varepsilon} 1_{\{\sqrt{[X]+\varepsilon} \geq X^*\}}\, d\sqrt{[X]+\varepsilon},$$

which is bounded below by $-\sqrt{[X]+\varepsilon}$ since the integrand on the right hand side is bounded below by -1. □

[38] We also use the trivial fact that $d[X] = d([X]+\varepsilon)$.

[39] Indeed $O := \{t > 0 : X_t^* > |X_t|\}$ is open in $\mathbb{R}$, so it can be written as a countable union of open intervals; on each of these $X_\cdot^*$ is constant, so $dX_\cdot^*$ is supported by $\mathbb{R}_+ \setminus O = \{t \geq 0 : X_t^* = |X_t|\}$.

Appendix

The appendix is devoted to proving Theorem 3 and Corollary 4. The perspicacious reader will wonder why in the proof of Theorem 3 we start from the *time-changed* pathwise Davis inequalities (and get (38)); the reason is that a crucial point in the proof is (42), where we make use of the stochastic Fubini theorem. It seems at first that one could get around this problem by introducing a more general stochastic Fubini theorem of the form

$$\int_0^T \left(\int_s^T K(s,t)\,dX_t\right) dV_s = \int_0^T \left(\int_0^{t-} K(s,t)\,dV_s\right) dX_t, \tag{36}$$

where X is a semimartingale and V a cadlag adapted process of finite variation. However, it is hard to determine for which class of integrands K the above equality holds,[40] and even for which K the family of random variables $\{\int_s^T K(s,t)\,dX_t : s \in [0,T]\}$ can be well-defined as a process.

Proof of Theorem 3. *Step 1: H, G are cag predictable.*

Since $\{C_s < t\} = \{s < \sqrt{[X]}_{t-}\}$, $F_t^{(C.)}$ is defined on $[0, \sqrt{[X]}_{t-})$ and setting $F_t^{(s)} := 0$ for $s \geq t$ we get[41]

$$H_t := \int_0^\infty p F_t^{(C_s)}\,d\phi(s), \qquad G_t := \int_0^\infty p F_t^{(D_s)}\,d\phi(s).$$

Denote with $\mathcal{F}_t \otimes \mathcal{B}$ the product sigma algebra of $\mathcal{F}_t$ with the Borel sets $\mathcal{B}$ of $\mathbb{R}_+$; we will now prove that $(F_t^{(C_s)})_{s\in\mathbb{R}_+}$ is $\mathcal{F}_t \otimes \mathcal{B}$ measurable, so that H_t is $\mathcal{F}_t$ measurable, that is, H is adapted. Since $\sqrt{[X]}$ is adapted, C_s is a stopping time and so $C_s \wedge t$ is $\mathcal{F}_t$ measurable, and so since $C.$ is cad the map $Z(\omega, s) := (\omega, C_s(\omega) \wedge t)$ is $\mathcal{F}_t \otimes \mathcal{B}/\mathcal{F}_t \otimes \mathcal{B}$ measurable. Analogously (but using *left* continuity[42] of $(F_t^{(s)})_{s\in[0,t)}$ and the fact that $F_t^{(s)} = 0$ for $s \geq t$) the map $F_t^{(\cdot)}$ is $\mathcal{F}_t \otimes \mathcal{B}$ measurable, and thus so is the composition $F_t^{(C.)} = F_t^{(\cdot)} \circ Z$. Since $F_t^{(C_s)} \in [-1,1]$, the dominated convergence theorem implies that H is cag (so predictable): indeed if $t_n \uparrow t$ and $M_t^{C_s} := \sup_{C_s \leq u < t} |X_u - X_{C_s-}| > 0$ then trivially $M_{t_n}^{C_s} \to M_t^{C_s}$ and $F_{t_n}^{(C_s)} \to F_t^{(C_s)}$, whereas if $M_t^{C_s} = 0$ then trivially $M_{t_n}^{C_s} = 0$ and so $0 = F_{t_n}^{(C_s)} \to F_t^{(C_s)} = 0$. The proof that G is adapted and cag is analogous.

Step 2: X satisfies $\Phi(\sqrt{[X]}) \leq c_p \Phi(X^*) - (H \cdot X)$.

Since C_s is a stopping time $Y_t^{(s)} := X_{C_s+t} - X_{C_s-}$ is a semimartingale (w.r.t. the time changed filtration $\mathcal{F}_t^{(s)} := \mathcal{F}_{C_s+t}$) and satisfies

$$\sqrt{[X]_{C_s+t}} - \sqrt{[X]_{C_s-}} \leq \sqrt{[X]_{C_s+t} - [X]_{C_s-}} = \sqrt{\left[Y^{(s)}\right]_t}. \tag{37}$$

[40] By integration by parts it holds at least when K is a finite sum of terms of the form $\theta 1_{(a,b]}(s) 1_{(c,d]}(t)$, where $a < b \leq c < d$ and θ is a $\mathcal{F}_a$-measurable random variable.

[41] We write $\int_0^\infty$ for $\int_{(0,\infty)}$, which is the same as $\int_{(0,\infty]}$ since by definition $\Phi(\infty) = \lim_{t\to\infty} \Phi(t)$.

[42] We warn the reader that for $s \uparrow t$ the limit of $F_t^{(s)}$ may not exist (so we specified $s \in [0,t)$).

Define $f(x,q,s) := x/\sqrt{q+s^2}$ and apply (8) to $Y^{(s)}$ (instead of X) to get

$$\sqrt{\big[Y^{(s)}\big]} \le 3Y^{(s)*} - \big(H^{(s)} \cdot Y^{(s)}\big) \qquad \text{for } H^{(s)} := f\big(Y^{(s)}_-, \big[Y^{(s)}\big]_-, Y^{(s)*}_-\big). \tag{38}$$

Writing the integrals as limits (in probability, uniformly on compacts) of Riemann sums we get

$$\big(H^{(s)} \cdot Y^{(s)}\big)_\cdot = \int_{C_s}^{C_s+\cdot} F_u^{(C_s)}\, dX_u,$$

so since $Y_t^{(s)*} \le 2X^*_{C_s+t}$, using (37), (38) we get

$$\sqrt{[X]_{C_s+\cdot}} - \sqrt{[X]_{C_s-}} \le 6X^*_{C_s+\cdot} - \int_{C_s}^{C_s+\cdot} F_u^{(C_s)}\, dX_u. \tag{39}$$

Since $\sqrt{[X]_{C_s-}} \le s$, $\{\sqrt{[X]_t} \ge s\} = \{C_{s-} \le t\} \supseteq \{C_s \le t\}$ and $[X]$ is constant on $[C_{s-}, C_s)$ (when this interval is non-empty) we get

$$\big(\sqrt{[X]_t} - s\big)^+ \le \big(\sqrt{[X]_t} - \sqrt{[X]_{C_s-}}\big) 1_{\{\sqrt{[X]_t} \ge s\}} = \big(\sqrt{[X]_t} - \sqrt{[X]_{C_s-}}\big) 1_{\{C_s \le t\}}$$

which, combined with (39) and with $X_t^* 1_{\{C_s \le t\}} \le X_t^* 1_{\{\sqrt{[X]_t} \ge s\}}$, gives

$$\big(\sqrt{[X]_t} - s\big)^+ \le 1_{\{\sqrt{[X]_t} \ge s\}} 6X_t^* - 1_{\{C_s \le t\}} \int_{C_s}^t F_u^{(C_s)}\, dX_u. \tag{40}$$

Integrating (40) over $s \in \mathbb{R}_+$ with respect to $d\phi(s)$ and using the identities $\Phi(t) = \int_0^\infty (t-s)^+ d\phi(s)$ and $1_{\{C_s \le t\}} \int_{C_s}^t F_u^{(C_s)}\, dX_u = \int_0^t F_u^{(C_s)} 1_{\{C_s < u\}}\, dX_u$ gives

$$\Phi\big(\sqrt{[X]_t}\big) \le 6X_t^* \phi\big(\sqrt{[X]_t}\big) - \int_0^\infty \left(\int_0^t F_u^{(C_s)} 1_{\{C_s < u\}}\, dX_u\right) d\phi(s) \tag{41}$$

Since $\{C_s < u\} = \{s < \sqrt{[X]_{u-}}\}$, the stochastic Fubini theorem [30], Chapter 4, Theorem[43] 65, gives

$$\int_0^\infty \int_0^t F_u^{(C_s)} 1_{\{C_s < u\}}\, dX_u\, d\phi(s) = \int_0^t \int_{[\sqrt{[X]_{0-}}, \sqrt{[X]_{u-}})} F_u^{(C_s)}\, d\phi(s)\, dX_u. \tag{42}$$

Now apply the inequalities (14) to write

$$6X_t^* \phi\big(\sqrt{[X]_t}\big) \le \Phi\big(6pX_t^*\big) + \Psi\big(\phi\big(\sqrt{[X]_t}\big)/p\big) \le (6p)^p \Phi\big(X_t^*\big) + \Psi\big(\phi\big(\sqrt{[X]_t}\big)\big)/p \tag{43}$$

[43] As observed after the statement of the theorem, by passing from $d\mu$ to $f\, d\mu$ for some $f \in L^1(\mu)$ one can prove the theorem for *sigma*-finite μ. Saying it differently, one can consider the *finite* measure $d\mu(s) = \exp(-\phi(s))\, d\phi(s)$ and apply Theorem 65 to $H_s^u := \exp(\phi(s)) F_u^{(C_s)} 1_{[0, A_{u-})}(s)$.

and bound the last term using that $\Psi(\phi(s)) \le (p-1)\Phi(s)$; combine the resulting inequality with (41) and (42) to get

$$\Phi\big(\sqrt{[X]_t}\big)\left(1-\frac{p-1}{p}\right) \le (6p)^p \Phi\big(X_t^*\big) - \int_0^t \frac{H_u}{p}\, dX_u,$$

that is, the first inequality (15). If $\Phi(t) = t^p/p$, to get the better constant $(6p)^p$, instead of (43) use the following inequality, obtained applying Young's inequality $ab \le a^p/p + b^q/q$ with $a = 6X_t^* p^{1/q}$ and $b = \phi(\sqrt{[X]_t})/p^{1/q}$:

$$6X_t^* \sqrt{[X]_t}^{\,p-1} \le \big((6p)^p \big(X_t^*\big)^p/p + \sqrt{[X]_t}^{\,p}/q\big)/p.$$

Step 3: X satisfies $\Phi(X^*) \le c_p \Phi(\sqrt{[X]}) + 2(G \cdot X)$.

Proceeding analogously using X^* and D (instead of $\sqrt{[X]}$ and C) yields a $Y^{(s)}$ which satisfies

$$X^*_{D_s+t} - X^*_{D_s-} \le \sup_{u \in [D_s, D_s+t]} |X_u - X_{D_s-}| = Y_t^{(s)*}$$

and since $\sqrt{[Y^{(s)}]_\cdot} \le \sqrt{[X]_{D_s+\cdot}}$ we obtain

$$X^*_{D_s+\cdot} - X^*_{D_s-} \le 6\sqrt{[X]_{D_s+\cdot}} + 2\int_{D_s}^{D_s+\cdot} F_u^{(D_s)}\, dX_u;$$

the proof continues exactly as before, yielding the second inequality (15).

Step 4: Alternative expression for H, G.

If ϕ is continuous and $A = [X]$, the cad inverse g of $\phi \circ A$ equals $C \circ \psi$ and $(\phi \circ A)(u-) = \phi(A_{u-}-)$, so

$$\int_{(\phi\circ A)(0-)}^{(\phi\circ A)(u-)} F_u^{(g(s))}\, ds = \int_{\phi(A_{0-}-)}^{\phi(A_{u-}-)} F_u^{(C_{\psi(s)})}\, ds. \tag{44}$$

By change of variable the first integral in (44) equals $\int_{[0,u)} F_u^{(s)}\, d\phi(A_s)$, and the second one $\int_{[A_{0-}, A_{u-})} F_u^{(C_s)}\, d\phi(s)$; thus $\int_{[0,t)} pF_t^{(s)}\, d\phi(\sqrt{[X]_s})$ equals H_t. Proceed analogously for $A = X^*$, D and G.

Step 5: H and G are lad.

We will use the expression (16), writing however the integrals over $\mathbb{R}_+$ and extending $F_t^{(s)}$ to be zero for $s \ge t$. Define the quantities $\hat{M}_t^s := \sup_{s \le u \le t} |X_u - X_{s-}|$ for $s \le t$ (so that $\hat{M}_t^s = |X_t - X_{t-}|$ for $s = t$) and

$$\hat{F}_t^{(s)} := \frac{X_t - X_{s-}}{\sqrt{[X]_t - [X]_{s-} + (\hat{M}_t^s)^2}} \qquad \text{for } s \le t, \qquad \hat{F}_t^{(s)} := 0 \qquad \text{for } s > t.$$

If $t_n \downarrow t$ and $s > t$, then by definition $F^{(s)}_{t_n} = 0 = \hat{F}^{(s)}_t$ for n big enough (such that $t_n < s$). If $s \le t$, then trivially $\hat{M}^s_{t_n} \to \hat{M}^s_t$, so if $\hat{M}^s_t > 0$ we get $F^{(s)}_{t_n} \to \hat{F}^{(s)}_t$; however if $\hat{M}^s_t = 0$ it can happen[44] that $\hat{M}^s_{t_n} > 0$ for all n and that $F^{(s)}_{t_n}$ does not converge to $\hat{F}^{(s)}_t = 0$. We shall now show that this can only happen for s in a set of $d\mu := d(\phi(\sqrt{[X]_\cdot}) + \phi(X^*_\cdot))$ measure zero, so $F^{(s)}_{t_n} \to \hat{F}^{(s)}_t$ for $d\mu$ a.e. s and the dominated convergence theorem implies that H and G lad and

$$H_{t+} = \int_{[0,t]} p\hat{F}^{(s)}_t \, d\phi\big(\sqrt{[X]_s}\big), \qquad G_{t+} = \int_{[0,t]} p\hat{F}^{(s)}_t \, d\phi\big(X^*_s\big). \tag{45}$$

If the set $Z := \{s : \hat{M}^s_t = 0\}$ contains some element $\underline{s}$ then necessarily it contains the whole interval $[\underline{s}, t]$; it follows that, for some $s_0 \in [0,t]$, Z equals either $[s_0, t]$ or $(s_0, t]$. Since X is cad, if $Z \ni s_n \downarrow s$ then $s \in Z$, so $Z = [s_0, t]$. Since X_u takes the constant value X_{s_0-} for all $u \in [s_0, t]$, $[X]_u$ (resp. X^*_u) takes the constant value $[X]_{s_0-}$ (resp. $X^*_{s_0-}$) for all $u \in [s_0, t]$, so $d\mu$ gives measure 0 to $[s_0, t]$. □

Proof of Corollary 4. Since for increasing positive A, B, with cad B, we have $\int_0^t A\,dB \le A_t B_t$, and since $|F^{(s)}_t| \le 1$ implies that $H^2 \le p^2 \phi(\sqrt{[X]})^2$, using (13) we get

$$[H \cdot X] = H^2 \cdot [X] \le p^2 \phi\big(\sqrt{[X]}\big)^2 \cdot [X] \le p^2 \phi\big(\sqrt{[X]}\big)^2 [X] \le p^2 \big(p\Phi\big(\sqrt{[X]}\big)\big)^2.$$

Analogously for G we can write, for any $y > 0$

$$\sqrt{[G \cdot X]} \le \sqrt{p^2 \phi\big(X^*\big)^2 \cdot [X]} \le p\phi\big(X^*\big)\sqrt{[X]} = \big(py\sqrt{[X]}\big)\left(\frac{\phi(X^*)}{y}\right);$$

when $y \ge 1$ (and so $py \ge 1$) we can apply the inequalities (14) and get

$$\big(py\sqrt{[X]}\big)\left(\frac{\phi(X^*)}{y}\right) \le \Phi\big(py\sqrt{[X]}\big) + \Psi\left(\frac{\phi(X^*)}{y}\right) \le (py)^p \Phi\big(\sqrt{[X]}\big) + \frac{\Psi(\phi(X^*))}{y};$$

now bound the last term above using that $\Psi(\phi(s)) \le (p-1)\Phi(s)$; putting the inequalities together concludes the proof of (17).

If X is a local-martingale, working as in Corollary 2 gives the thesis: the only difference here is that $\mathbb{E}(H \cdot X)^*_\infty < \infty$ follows from $\mathbb{E}\Phi(\sqrt{[X]_\infty}) < \infty$ since (17) gives that $\mathbb{E}\sqrt{[H \cdot X]_\infty} < \infty$ and we can then apply (12) to $H \cdot X$ (instead of X). □

Acknowledgements

This paper was written while the author was at Oxford University. This research was generously supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 335421, and the Oxford-Man Institute of Quantitative Finance.

[44] For example take $s = 0$ and see footnote 16.

References

[1] Acciaio, B., Beiglböck, M., Penkner, F., Schachermayer, W. and Temme, J. (2013). A trajectorial interpretation of Doob's martingale inequalities. *Ann. Appl. Probab.* **23** 1494–1505. MR3098440

[2] Aoyama, H. (2009). Lebesgue spaces with variable exponent on a probability space. *Hiroshima Math. J.* **39** 207–216. MR2543650

[3] Beiglböck, M. and Nutz, M. (2014). Martingale inequalities and deterministic counterparts. *Electron. J. Probab.* **19** no. 95, 15. MR3272328

[4] Beiglböck, M. and Siorpaes, P. (2013). Pathwise versions of the Burkholder–Davis–Gundy inequality. arXiv preprint, arXiv:1305.6188v3.

[5] Beiglböck, M. and Siorpaes, P. (2015). Pathwise versions of the Burkholder–Davis–Gundy inequality. *Bernoulli* **21** 360–373. MR3322322

[6] Bichteler, K. (1981). Stochastic integration and L^p-theory of semimartingales. *Ann. Probab.* **9** 49–89.

[7] Bonami, A. and Lépingle, D. (1979). Fonction maximale et variation quadratique des martingales en présence d'un poids. *Sémin. Probab.* **13** 294–306.

[8] Bouchard, B. and Nutz, M. (2015). Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.* **25** 823–859.

[9] Chou, C. (1975). Les methodes d'A. Garsia en theorie des martingales. Extensions au cas continu. *Sémin. Probab.* **9** 213–225.

[10] Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.* **300** 463–520.

[11] Dellacherie, C. and Meyer, P.-A. (1982). *Probabilities and Potential. B: Theory of Martingales. North-Holland Mathematics Studies* **72**. Amsterdam: North-Holland. Translated from the French by J. P. Wilson.

[12] Graversen, S. and Peškir, G. (1998). Maximal inequalities for Bessel processes. *J. Inequal. Appl.* **1998** 621735.

[13] Gushchin, A.A. (2014). On pathwise counterparts of Doob's maximal inequalities. *Proc. Steklov Inst. Math.* **287** 118–121. MR3484325

[14] Hobson, D. (1998). Robust hedging of the lookback option. *Finance Stoch.* **2** 329–347.

[15] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer.

[16] Jiao, Y., Zhou, D., Hao, Z. and Chen, W. (2016). Martingale Hardy spaces with variable exponents. *Banach J. Math. Anal.* **10** 750–770.

[17] Kallenberg, O. and Sztencel, R. (1991). Some dimension-free features of vector-valued martingales. *Probab. Theory Related Fields* **88** 215–247. MR1096481

[18] Karandikar, R. (1995). On pathwise stochastic integration. *Stochastic Process. Appl.* **57** 11–18.

[19] Kazamaki, N. (1994). *Continuous Exponential Martingales and BMO. Lecture Notes in Math.* **1579**. Berlin: Springer. MR1299529

[20] Krasnoselskii, M. and Rutickii, Y.B. (1961). *Convex Functions and Orlicz Spaces*. Groningen, The Netherlands: Noordhoff. Translated from the first Russian edition by L. F. Boron.

[21] Kwapien, S. and Woyczynski, W.A. (1992). *Random Series and Stochastic Integrals: Single and Multiple*. Boston: Birkhäuser.

[22] Lenglart, E., Lépingle, D. and Pratelli, M. (1980). Présentation unifiée de certaines inégalités de la théorie des martingales. *Sémin. Probab.* **14** 26–48.

[23] Liu, P. and Wang, M. (2014). Burkholder–Gundy–Davis inequality in martingale Hardy spaces with variable exponent. arXiv preprint, arXiv:1412.8146.

[24] Lowther, G. (2015). https://almostsure.wordpress.com/2010/04/06/the-burkholder-davis-gundy-inequality/.
[25] Meyer, P.-A. (1976). Un cours sur les intégrales stochastiques (exposés 1 à 6). *Sémin*. *Probab*. **10** 245–400.
[26] Nikeghbali, A. (2007). Non-stopping times and stopping theorems. *Stochastic Process*. *Appl*. **117** 457–475.
[27] Nikeghbali, A. (2008). How badly are the Burkholder–Davis–Gundy inequalities affected by arbitrary random times? *Statist*. *Probab*. *Lett*. **78** 766–770.
[28] Nikeghbali, A. and Yor, M. (2005). A definition and some characteristic properties of pseudo-stopping times. *Ann*. *Probab*. 1804–1824.
[29] Osekowski, A. (2010). Sharp maximal inequalities for the martingale square bracket. *Stochastics* **82** 589–605.
[30] Protter, P.E. (2004). *Stochastic Integration and Differential Equations*, 2nd ed. *Applications of Mathematics* (*New York*) **21**. Berlin: Springer. MR2020294
[31] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Berlin: Springer. MR1725357
[32] Rogers, L.C.G. and Williams, D. (2000). *Diffusions*, *Markov Processes and Martingales*: *Vol*. 2, *Itô Calculus*. Cambridge University Press.
[33] Sekiguchi, T. (1979). BMO-martingales and inequalities. *Tohoku Math*. *J*. (2) **31** 355–358.
[34] Sekiguchi, T. (1980). Weighted norm inequalities on the martingale theory. *Mathematics Reports* **3** 37–100.

Received February 2017