

Martingale inequalities

Pavel Zorin-Kranich

January 21, 2022

University of Bonn, winter term 2021

Literature There are two books whose titles include our topic [Gar73; Lon93]. They can be useful for some of the basics, but are overall outdated.

For vector-valued martingales, modern references are [Hyt+16; Pis16].

The only book about sharp constants in martingale inequalities is [Osę12].

One stochastic analysis text that pays attention to our topic is [Kal21].

Contents

0	Review of martingale basics	2
1	Maximal and square functions	4
1.1	Predictable square function	10
2	Lépingle inequality	13
2.1	Martingale convergence	15
3	Vector-valued inequalities	17
3.1	Weighted and vector-valued Doob inequalities	19
4	Rough paths	21
4.1	Young integration with Hölder functions	21
4.2	Young integration with V^r functions	22
4.3	Rough paths	25
4.3.1	Rough differential equations	26
4.3.2	Operations on controlled paths	26
4.3.3	Contractive iteration	28
4.3.4	Lipschitz dependence on data	31
5	Martingale transforms	32
5.1	Davis decomposition	32
5.2	Vector-valued maximal paraproduct estimate	33
5.3	Stopping time construction	36
6	Itô integration	39
6.1	Quadratic covariation	42
7	A sharp inequality for the square function	44
7.1	A Bellman function	44
7.2	Sharp constant in the Davis inequality for the square function	46
	References	47

0 Review of martingale basics

Random variables will be usually defined on a filtered probability space $(\Omega, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mu)$. We denote by \mathcal{F}_∞ the σ -algebra generated by $\cup_{n \in \mathbb{N}} \mathcal{F}_n$. The following examples are useful to keep in mind.

Example 0.1 (Dyadic filtration). $\Omega = [0, 1]$, μ Lebesgue measure, \mathcal{F}_n is the σ -algebra generated by the dyadic intervals of length 2^{-n} , that is, intervals of the form $[2^{-n}k, 2^{-n}(k+1)]$ with $k \in \mathbb{Z}$.

A higher dimensional version involves dyadic cubes. One can also construct similar filtrations on more general manifolds, or even metric measure spaces.

Example 0.2 (Atomic filtrations). $\Omega = [0, 1]$, μ Lebesgue measure, \mathcal{F}_n is a σ -algebra generated by finitely many intervals.

We recall that an *atom* for a measure μ on a σ -algebra \mathcal{F} is a set $A \in \mathcal{F}$ such that $\mu(A) > 0$ and, for every $A' \in \mathcal{F}$ with $A' \subseteq A$, we have $\mu(A') \in \{0, \mu(A)\}$. In an atomic σ -algebra, every measurable set is a finite union of atoms.

One can view martingale analysis as analysis of atomic filtrations, if we do not allow any constants to depend on the filtration. All results that we are interested in can be transferred from atomic filtrations to general filtrations. However, it is technically convenient to always use general filtrations, since they can appear in applications.

Example 0.3. $\Omega = [0, 1]^{\mathbb{N}}$, μ the product of Lebesgue measures,

$$\mathcal{F}_n = \{B \times [0, 1]^{\{n, n+1, \dots\}} \mid B \subseteq [0, 1]^{\{0, \dots, n-1\}} \text{ Borel}\}. \quad (0.1)$$

This filtration appears in the analysis of independent random variables, and gives a good idea of how a general filtration looks like.

An *adapted process* is a sequence of functions (f_n) such that, for every $n \in \mathbb{N}$, the function f_n is \mathcal{F}_n -measurable.

An adapted process f is called *predictable* if, for every $n > 0$, the function f_n is \mathcal{F}_{n-1} -measurable.

For nested σ -algebras $\mathcal{F}' \subseteq \mathcal{F}$ on Ω , the *conditional expectation* is the orthogonal projection

$$\mathbb{E}(\cdot | \mathcal{F}') : L^2(\Omega, \mathcal{F}, \mu) \rightarrow L^2(\Omega, \mathcal{F}', \mu).$$

The conditional expectation has the following properties. Here and later, all identities and inequalities are meant to hold almost surely, unless mentioned otherwise.

1. $\mathbb{E}(\mathbf{1} | \mathcal{F}') = \mathbf{1}$.
2. For every $p \in [1, \infty]$, \mathbb{E} extends to a contraction $L^p(\mathcal{F}) \rightarrow L^p(\mathcal{F}')$
3. $\int f \mathbb{E}(f | \mathcal{F}') = \int f$ for every $f \in L^1(\mathcal{F})$.
4. Positivity: $f \geq 0 \implies \mathbb{E}(f | \mathcal{F}') \geq 0$.
5. Assume that $f \in L^1(\mathcal{F})$, $g \in L^0(\mathcal{F}')$, and either $fg \in L^1(\mathcal{F})$ or $f \geq 0$, $\mathbb{E}(f | \mathcal{F}')g \in L^1(\mathcal{F}')$. Then

$$\mathbb{E}(fg | \mathcal{F}') = \mathbb{E}(f | \mathcal{F}')g \text{ in } L^1(\mathcal{F}').$$

Suppose that (X, \mathcal{F}, μ) is a regular measure space and $\mathcal{F}' \subseteq \mathcal{F}$ is a sub- σ -algebra. Then there exists an essentially unique measurable map $(X, \mathcal{F}') \rightarrow \mathcal{M}(X)$, $y \mapsto \mu_y$ such that, for every $f \in L^1(\mathcal{F})$, for μ -a.e. $y \in X$, we have $f \in L^1(X, \mu_y)$, and $\int f d\mu_y = \mathbb{E}(f | \mathcal{F}')(y)$. This map is called a *measure disintegration*. A measure disintegration satisfies $\mu_x = \mu_y$ for μ -a.e. y and μ_y -a.e. x .

Example 0.4. If \mathcal{F}' is an atomic filtration, then we can choose a collection of disjoint atoms $\mathcal{A} \subseteq \mathcal{F}'$ with $\mu(\cup_{A \in \mathcal{A}} A) = 1$. Then, for each $x \in A \in \mathcal{A}$, we can set $d\mu_x := \mu(A)^{-1} \mathbf{1}_A d\mu$, and

$$\mathbb{E}(f|\mathcal{F}')(x) = \mu(A)^{-1} \int_A f(x') d\mu(x').$$

Example 0.5. Let $\Omega = [0, 1]^2$ with the Lebesgue measure, \mathcal{F} the Borel σ -algebra, and \mathcal{F}' the Borel σ -algebra in the first variable, like in (7.5). Then, we can choose $\mu_{(x,y)}$ to be the Lebesgue measure on $\{x\} \times [0, 1]$, and

$$\mathbb{E}(f|\mathcal{F}')(x, y) = \int_0^1 f(x, y') dy'.$$

A *martingale* is an adapted process with values in \mathbb{C} (later also in a Banach space) such that, for every $m \leq n$, the function f_n is integrable, and we have

$$f_m = \mathbb{E}(f_n|\mathcal{F}_m).$$

Example 0.6. Sums of independent random variables

Example 0.7 (Dyadic martingale). Let f be an integrable function on $[0, 1]$ with the Lebesgue measure and $f_n = \mathbb{E}(f|\mathcal{F}_n)$ (the same definition works for any filtration).

This construction can often be used to transfer results from martingales to a real analysis setting. The main difference between the analysis of dyadic martingales and general martingales is that the Calderón-Zygmund decomposition does not work for general martingales. A substitute that does work for general martingales is the Gundy decomposition.

The main reason why this construction does not produce all possible martingales is that it may happen that $\lim_{n \rightarrow \infty} \|f_n\| = \infty$, in which case there might be no function f with $f_n = \mathbb{E}(f|\mathcal{F}_n)$. A well-known example involves a doubling betting strategy.

Example 0.8. Any integrable process can be written as the sum of a process with predictable jumps and a martingale.

A *stopping time* is a function $\tau : \Omega \rightarrow \bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ such that, for every $n \in \mathbb{N}$, we have $\{\tau \leq n\} \in \mathcal{F}_n$.

Example 0.9. Any constant function is a stopping time.

Example 0.10 (Hitting time). If f is an adapted process with values in a metric space X and $B \subseteq X$ is a Borel set, then

$$\tau := \inf\{t \mid f_t \in B\}$$

is a stopping time, called the *first hitting time* of B .

If σ, τ are stopping times, then $\sigma \wedge \tau$ and $\sigma \vee \tau$ are also stopping times.

For a stopping time τ and a process f , the *stopped process* f^τ is defined by

$$f_t^\tau := f_{t \wedge \tau}.$$

If f is a martingale, then the stopped process f^τ is again a martingale.

If τ is a stopping time, the corresponding σ -algebra is defined by

$$\mathcal{F}_\tau := \{A \in \mathcal{F}_\infty \mid (\forall n) A \cap \{\tau \leq n\} \in \mathcal{F}_n\}.$$

The function τ is \mathcal{F}_τ -measurable. We abbreviate $\mathbb{E}_\tau f := \mathbb{E}(f|\mathcal{F}_\tau)$.

The *optional sampling* theorem says that, for every discrete time martingale f , bounded stopping time τ , and another stopping time σ , we have

$$f_{\sigma \wedge \tau} = \mathbb{E}_\sigma f_\tau.$$

1 Maximal and square functions

You are probably already familiar with the Lebesgue differentiation theorem. The best proof of that theorem uses the maximal operator to absorb error terms. This idea is used in many places in analysis, and the study of martingales is one of these places.

Definition 1.1 (Maximal operator). For an adapted process f with values in a normed space, we write

$$Mf_n = (Mf)_n := \sup_{k \leq n} |f_k|.$$

Remark. Since our martingales are indexed by a countable set, we can use a pointwise supremum here. In continuous time, we would need a lattice supremum.

A *submartingale* is an adapted process with values in $\mathbb{R}_{\geq 0}$ such that, for every $m \leq n$, we have

$$f_m \leq \mathbb{E}_m f_n.$$

Example 1.2. If (f_n) is a martingale, then $(|f_n|)$ is a submartingale.

Example 1.3. If f is any adapted process, then Mf is a submartingale.

If f is a submartingale, then, for any $p \in [1, \infty]$ and $m \leq n$, we have

$$\|f_m\|_p \leq \|f_n\|_p.$$

We define the L^p norm of a (sub-)martingale by

$$\|f\|_p := \sup_n \|f_n\|_p.$$

With this definition, the maximal operator is clearly bounded on L^∞ . The next result looks a lot like an $L^1 \rightarrow L^{1,\infty}$ bound for M , but is in fact stronger and more convenient to use.

Lemma 1.4. *Let f be a submartingale with values in $\mathbb{R}_{\geq 0}$. Then, for every $\lambda > 0$ and $n \in \mathbb{N}$, we have*

$$\lambda |\{Mf_n > \lambda\}| \leq \int_{\{Mf_n > \lambda\}} f_n \, d\mu.$$

Proof. For a fixed λ , define the stopping time

$$\tau := \inf\{k \mid f_k > \lambda\}.$$

Then, $\{Mf_n > \lambda\} = \{\tau \leq n\}$. Hence,

$$\begin{aligned} \lambda |\{Mf_n > \lambda\}| &= \lambda \sum_{k \leq n} |\{\tau = k\}| \\ &\leq \sum_{k \leq n} \int_{\{\tau = k\}} f_k \, d\mu \\ \text{(submartingale property)} &\leq \sum_{k \leq n} \int_{\{\tau = k\}} f_n \, d\mu \\ &= \int_{\{Mf_n > \lambda\}} f_n \, d\mu. \quad \square \end{aligned}$$

Remark ($L^{1,\infty}$ norm). By Chebychev's inequality, for any measurable function g , we have

$$\sup_{\lambda > 0} \lambda^{-1} |\{g > \lambda\}| \leq \|g\|_1.$$

The converse inequality is false with any constant, as shown by the example $g(x) = 1/x$ on the measure space $\mathbb{R}_{>0}$ with the Lebesgue measure. The left-hand side of the above inequality is the so-called $L^{1,\infty}$ seminorm of g . As suggested by the notation, $L^{1,\infty}$ is part of a larger family of Lorentz spaces, which is in turn contained in Orlicz spaces, but we will not talk about such generalities.

Theorem 1.5 (Doob's maximal inequality). *Let f be a submartingale with values in $\mathbb{R}_{\geq 0}$. Then, for every $p \in (1, \infty]$, we have*

$$\|Mf\|_p \leq p' \|f\|_p.$$

Here and later, p' denotes the Hölder conjugate: $1/p + 1/p' = 1$.

Proof. By the layer cake formula and Lemma 1.4, we write

$$\begin{aligned} \|Mf_n\|_p^p &= \int_0^\infty p\lambda^{p-1} \mu(Mf_n > \lambda) \, d\lambda \\ &\leq \int_0^\infty p\lambda^{p-2} \int_{\{Mf_n > \lambda\}} f_n \, d\mu \, d\lambda \\ &= \int_\Omega \int_0^{Mf_n} p\lambda^{p-2} f_n \, d\lambda \, d\mu \\ &= \frac{p}{p-1} \int_\Omega (Mf_n)^{p-1} f_n \, d\mu \\ &\leq \frac{p}{p-1} \left(\int_\Omega f_n^p \, d\mu \right)^{1/p} \left(\int_\Omega (Mf_n)^p \, d\mu \right)^{1-1/p}. \end{aligned}$$

Since $f_n \in L^p$ implies $f_k \in L^p$ for all $k \leq n$ by L^p contractivity of conditional expectations, we have $Mf_n \in L^p$. Hence, we can cancel suitable powers of $\|Mf_n\|_p$ on both sides and obtain the claim. \square

There are other ways to deduce Theorem 1.5 from Lemma 1.4, such as real interpolation with the L^∞ estimate. But the above proof does not rely on the L^∞ estimate, only on the L^1 -like estimate in Lemma 1.4.

It is often the case in martingale theory that inequalities near L^1 are the most powerful ones. One can justify this by the observation that L^1 is the minimal assumption required to even define a martingale.

If f is a martingale, then the increments df_k are orthogonal in the Hilbert space $L^2(\Omega)$. Indeed, more generally, for any n and $g \in L^0(\mathcal{F}_{n-1})$, we have

$$\mathbb{E}(gdf_n) = \mathbb{E}(\mathbb{E}(gdf_n|\mathcal{F}_{n-1})) = \mathbb{E}(g\mathbb{E}(df_n|\mathcal{F}_{n-1})) = \mathbb{E}(g \cdot 0) = 0. \quad (1.1)$$

Definition 1.6 (Square function). For a martingale f , we write

$$Sf_n := \left(\sum_{k \leq n} |df_k|^2 \right)^{1/2}.$$

In continuous time, the analog of Sf is denoted by $[f]$ and called the *quadratic variation*.

Here and later, for notational simplicity, we will consider martingales with $f_0 = 0$. Since martingale increments df_k are orthogonal, we have

$$\|f_n\|_2 = \|Sf_n\|_2.$$

Next, we will see that S is bounded from L^1 to $L^{1,\infty}$. The proof that we present for this fact uses a summation by parts identity of a kind that also appears e.g. in the Itô formula.

Lemma 1.7. *Let f be a martingale, $\lambda > 0$, and $\tau := \inf\{n \mid |f_n| > \lambda\}$. Then,*

$$\sum_k \mathbb{E}(|df_k|^2 \mathbf{1}_{\tau > k}) \leq 2\lambda \|f\|_1.$$

On the left-hand side of the above estimate, we are taking the expectation of $S_{\tau-1}$. Note that $\tau - 1$ is not a stopping time. Such predictability emulation will also be important in the BDG inequality below.

Proof. By the monotone convergence theorem, it suffices to consider the case $f_N = f_{N+1} = \dots$ for an arbitrarily large N . Then, we can replace τ by $\tau \wedge N + 1$ without changing the inequality. This is necessary to apply the optional sampling theorem.

We use the identity

$$\sum_{k < \tau} |df_k|^2 + |f_{\tau-1}|^2 = 2f_\tau f_{\tau-1} - 2 \sum_{k \leq \tau} f_{k-1} df_k.$$

Note that

$$\{k \leq \tau\} = \Omega \setminus \{\tau < k\} = \Omega \setminus \{\tau \leq k-1\} \in \mathcal{F}_{k-1}.$$

Therefore, by (1.1), we have

$$\mathbb{E}(\mathbf{1}_{k \leq \tau} f_{k-1} df_k) = 0.$$

Hence,

$$\mathbb{E} \sum_{k < \tau} |df_k|^2 + \mathbb{E}|f_{\tau-1}|^2 = 2\mathbb{E}(f_\tau f_{\tau-1}).$$

Therefore,

$$\sum_k \mathbb{E}(|df_k|^2 \mathbf{1}_{\tau > k}) \leq 2\mathbb{E}|f_\tau f_{\tau-1}| \leq 2\mathbb{E}|f_\tau| \lambda \leq 2\lambda \|f\|_1,$$

where the last step follows from the optional sampling theorem, which gives in particular $f_\tau = \mathbb{E}_\tau f_{N+1}$. \square

Corollary 1.8. *Let f be a martingale and $\lambda > 0$. Then,*

$$|\{Sf > \lambda\}| \leq 3\lambda^{-1} \|f\|_1.$$

Proof. Let $\tau := \inf\{n \mid |f_n| > \lambda\}$. Then

$$|\{Sf > \lambda\}| \leq |\{Sf_{\tau-1} > \lambda\}| + \{\tau < \infty\} \leq \lambda^{-2} \|Sf_{\tau-1}\|_2^2 + \lambda^{-1} \|f\|_1.$$

By Lemma 1.7, the first summand is $\leq 2\lambda^{-1} \|f\|_1$. \square

Lemma 1.9 (Davis decomposition). *For every martingale (f_n) , there exists a decomposition $f = f^{\text{pred}} + f^{\text{bv}}$ as a sum of two martingales such that f^{pred} has a predictable bound on jumps:*

$$|df_n^{\text{pred}}| \leq 2Mdf_{n-1}, \quad (1.2)$$

and f^{bv} has bounded variation:

$$\mathbb{E} \sum_n |df_n^{\text{bv}}| \leq 2\mathbb{E}Mdf. \quad (1.3)$$

Proof. Let

$$\begin{aligned} dg_n &:= \min\left(1, \frac{Mdf_{n-1}}{|df_n|}\right) df_n, \\ df_n^{\text{pred}} &:= dg_n - \mathbb{E}_{n-1}(dg_n), \\ dh_n &:= df_n - dg_n = \max\left(0, 1 - \frac{Mdf_{n-1}}{|df_n|}\right) df_n, \\ df_n^{\text{bv}} &:= dh_n - \mathbb{E}_{n-1}(dh_n). \end{aligned}$$

Then, by definition,

$$|dg_n| \leq Mdf_{n-1},$$

and, by positivity of conditional expectation, also

$$|\mathbb{E}_{n-1}(dg_n)| \leq \mathbb{E}_{n-1}(|dg_n|) \leq \mathbb{E}_{n-1}(Mdf_{n-1}) = Mdf_{n-1}.$$

This implies (1.2).

On the other hand, we have the telescoping bound

$$|dh_n| = \max(0, |df_n| - Mdf_{n-1}) = Mdf_n - Mdf_{n-1},$$

which implies (1.3). \square

Theorem 1.10 (Davis inequalities). *Let f be a martingale with $f_0 = 0$. Then,*

$$\mathbb{E}Sf \sim \mathbb{E}Mf.$$

Here and later, $A \lesssim B$ means that $A \leq CB$ with an absolute constant C , and $A \sim B$ means that $A \lesssim B$ and $B \lesssim A$.

Proof. It suffices to consider martingales with $f_N = f_{N+1} = \dots =: f_\infty$ for some N , as long as we show bounds independent of N .

Let $f = f^{\text{pred}} + f^{\text{bv}}$ be a Davis decomposition as in Lemma 1.9. Then, for every $n > 0$, we have the predictable bound

$$|f_n^{\text{pred}}| \leq |f_{n-1}^{\text{pred}}| + |df_n^{\text{pred}}| \leq |f_{n-1}^{\text{pred}}| + 2|Mdf_{n-1}| =: \rho_{n-1}.$$

Let $\lambda > 0$ and $\tau := \inf\{n \mid \rho_n > \lambda\}$. Then,

$$\begin{aligned} |\{Sf^{\text{pred}} > \lambda\}| &\leq |\{Sf^{\text{pred}} > \lambda, \tau = \infty\}| + \{\tau < \infty\} \\ &\leq |\{Sf^{\text{pred},(\tau)} > \lambda\}| + \{\tau < \infty\} \\ &\leq \lambda^{-2} \|Sf^{\text{pred},(\tau)}\|_2^2 + \{\tau < \infty\} \\ &= \lambda^{-2} \|f_\tau^{\text{pred}}\|^2 + \{\tau < \infty\}. \end{aligned}$$

On the set $\{\tau = \infty\}$, we use the bound $Mf^{\text{pred}} \leq M\rho \leq \lambda$, while on the set $\{\tau < \infty\}$, we use the bound $|f_t^{\text{pred}}| \leq \rho_{\tau-1} \leq \lambda$ (at this point, predictability is essential). This gives

$$\begin{aligned} |\{Sf^{\text{pred}} > \lambda\}| &\leq \lambda^{-2} \int_{\{\tau=\infty\}} |f_t^{\text{pred}}|^2 + \lambda^{-2} \int_{\{\tau<\infty\}} |f_t^{\text{pred}}|^2 + |\{\tau < \infty\}| \\ &\leq \lambda^{-2} \int_{\{Mf^{\text{pred}} \leq \lambda\}} |Mf^{\text{pred}}|^2 + |\{\tau < \infty\}| + |\{\tau < \infty\}|. \end{aligned}$$

Note that $\{\tau < \infty\} = \{M\rho > \lambda\}$. Inserting this in the above inequality and integrating in λ , we obtain

$$\begin{aligned} \mathbb{E}Sf^{\text{pred}} &= \int_0^\infty |\{Sf^{\text{pred}} > \lambda\}| d\lambda \\ &\leq \int_0^\infty \lambda^{-2} \int_{\{Mf^{\text{pred}} \leq \lambda\}} |Mf^{\text{pred}}|^2 d\lambda + 2 \int_0^\infty |\{M\rho > \lambda\}| d\lambda \\ &= \mathbb{E}|Mf^{\text{pred}}|^2 \int_{Mf^{\text{pred}}}^\infty \lambda^{-2} d\lambda + 2\mathbb{E}M\rho \\ &= \mathbb{E}Mf^{\text{pred}} + 2\mathbb{E}M\rho \\ &\leq 3\mathbb{E}Mf^{\text{pred}} + 4\mathbb{E}Mdf \\ &\leq 3\mathbb{E}Mf + 3\mathbb{E} \sum_n |df_n^{\text{bv}}| + 4\mathbb{E}Mdf \\ &\leq 3\mathbb{E}Mf + 10\mathbb{E}Mdf. \end{aligned}$$

This, together with the simple bound

$$\mathbb{E}Sf^{\text{bv}} \leq \mathbb{E} \sum_n |df_n^{\text{bv}}| \leq 2\mathbb{E}Mdf,$$

implies $\mathbb{E}Sf \lesssim \mathbb{E}Mf$.

The proof of the converse inequality is similar and uses

$$Sf_n^{\text{pred}} = ((Sf_{n-1}^{\text{pred}})^2 + (df_n^{\text{pred}})^2)^{1/2} \leq |Sf_{n-1}^{\text{pred}}| + |df_n^{\text{pred}}| \leq |Sf_{n-1}^{\text{pred}}| + 2|Mdf_{n-1}| =: \rho_{n-1}.$$

□

Lemma 1.11 (Garsia–Neveu). *Let W, Z be positive random variables such that, for every $\lambda > 0$, we have*

$$\mathbb{E}(\mathbf{1}_{W>\lambda}(W - \lambda)) \leq \mathbb{E}(\mathbf{1}_{W>\lambda}Z). \quad (1.4)$$

Then, for every $p \geq 1$, we have

$$\|W\|_p \leq p\|Z\|_p.$$

Proof. For $p = 1$, it suffices to take $\lambda \rightarrow 0$, so assume $p > 1$. Suppose first that W is bounded by Λ and (1.4) holds for all $\lambda \in (0, \Lambda)$. We will use the formula

$$t^p = p(1-p) \int_0^t (t-\lambda)\lambda^{p-2} d\lambda.$$

It yields

$$\begin{aligned}
\mathbb{E}W^p &= p(p-1)\mathbb{E}\int_0^W (W-\lambda)\lambda^{p-2}d\lambda \\
&= p(p-1)\int_0^\Lambda \mathbb{E}((W-\lambda)\mathbf{1}_{W>\lambda})\lambda^{p-2}d\lambda \\
&\leq p(p-1)\int_0^\Lambda \mathbb{E}(Z\mathbf{1}_{W>\lambda})\lambda^{p-2}d\lambda \\
&= p(p-1)\mathbb{E}\int_0^W Z\lambda^{p-2}d\lambda \\
&= p\mathbb{E}ZW^{p-1} \\
&\leq p(\mathbb{E}Z^p)^{1/p}(\mathbb{E}W^p)^{1-1/p}.
\end{aligned}$$

Since W is bounded, we can cancel a suitable power of $\mathbb{E}W^p$ on both sides and obtain the claim.

For general W , we apply the bounded case to $\min(W, \Lambda)$ and let $\Lambda \rightarrow \infty$. \square

Corollary 1.12. *Let (A_t) be an increasing predictable process with $A_0 = 0$ and ξ a positive random variable such that, for every t , we have*

$$\mathbb{E}_t(A_\infty - A_t) \leq \mathbb{E}_t(\xi).$$

Then, for every $p \geq 1$, we have

$$\|A_\infty\|_p \leq p\|\xi\|_p.$$

Proof. For $\lambda > 0$, let $\tau := \inf\{t \mid A_{t+1} > \lambda\}$. Then,

$$\mathbb{E}\max(A_\infty - \lambda, 0) = \mathbb{E}(A_\infty - \lambda)\mathbf{1}_{\tau < \infty} \leq \mathbb{E}(A_\infty - A_\tau)\mathbf{1}_{\tau < \infty} \leq \mathbb{E}(\xi\mathbf{1}_{\tau < \infty}) \leq \mathbb{E}(\xi\mathbf{1}_{A_\infty > \lambda}),$$

so we can apply Lemma 1.11. \square

Corollary 1.13 (Burkholder–Davis–Gundy inequalities). *For every $p \in [1, \infty)$ and every martingale f , we have*

$$\|Sf\|_p \sim \|Mf\|_p.$$

Proof. For any t_0 and $B \in \mathcal{F}_{t_0}$, we can apply Theorem 1.10 to the martingale $\mathbf{1}_B(f - f^{(t_0)})_{t \geq t_0}$, and obtain

$$\mathbb{E}_{t_0}(M(f - f^{(t_0)})) \sim \mathbb{E}_{t_0}(S(f - f^{(t_0)})).$$

Note that

$$Mf - Mf_{t_0} \leq M(f - f^{(t_0)}) \leq 2Mf, \quad Sf - Sf_{t_0} \leq S(f - f^{(t_0)}) \leq Sf,$$

so the conditions of Corollary 1.12 hold for instance for $\tilde{A}_t = Mf_t$ and $\xi = Sf$, with exception of the fact that \tilde{A}_t is not predictable. This can be remedied by taking $A_t = Mf_{t-1}$ and $\xi = Sf + Mdf \leq 3Sf$. Corollary 1.12 now shows that $\|Mf\|_p \lesssim p\|Sf\|_p$.

The proof of the converse inequality is similar. □

1.1 Predictable square function

Definition 1.14 (Predictable square function). For a martingale f , let

$$sf_n := \left(\sum_{k=1}^n \mathbb{E}_{k-1}(|df_k|^2) \right)^{1/2}.$$

The continuous time version of sf is denoted by $\langle f \rangle$ and called the *predictable quadratic variation*. We mainly discuss sf because the continuous time analog is sometimes used in stochastic analysis, and in order to present a few techniques.

The next lemma looks similar to L^p contractivity of conditional expectation, but it cannot be proved by using this contractivity for each individual summand.

Lemma 1.15. *Let z_1, z_2, \dots be positive random variables. Then, for every $p \in [1, \infty)$, we have*

$$\mathbb{E} \left(\sum_{k=1}^{\infty} \mathbb{E}_{k-1}(z_k) \right)^p \leq p^p \mathbb{E} \left(\sum_{k=1}^{\infty} z_k \right)^p.$$

Proof. The hypothesis of Corollary 1.12 holds with equality for $A_n = W_n$ and $\xi = Z_\infty$, where

$$W_n := \sum_{k=1}^n \mathbb{E}_{k-1}(z_k), \quad Z_n := \sum_{k=1}^n z_k. \tag{1.5}$$

□

Corollary 1.16. *Let $p \in [2, \infty)$ and f be a martingale. Then, we have*

$$\|sf\|_p \leq (p/2)^{1/2} \|Sf\|_p.$$

Lemma 1.17. *Let Z, W be positive random variables such that, for some $C < \infty$ and all $\lambda > 0$, we have*

$$\mathbb{E}(Z \wedge \lambda) \leq C \mathbb{E}(W \wedge \lambda).$$

Then, for every $p \in (0, 1]$, we have

$$\mathbb{E}Z^p \leq C \mathbb{E}W^p.$$

Proof. For $p = 1$, it suffices to let $\lambda \rightarrow \infty$ in the hypothesis.

For $p < 1$, we use the formula

$$t^p = p(1-p) \int_0^\infty (t \wedge \lambda) \lambda^{p-2} d\lambda.$$

Changing the order of integration and applying the hypothesis for each λ , we obtain the claim. \square

Corollary 1.18. *Let z_1, z_2, \dots be positive random variables. Then, for every $p \in (0, 1]$, we have*

$$\mathbb{E}\left(\sum_{k=1}^{\infty} z_k\right)^p \leq 2\mathbb{E}\left(\sum_{k=1}^{\infty} \mathbb{E}_{k-1}(z_k)\right)^p.$$

Thus, we see that, for $p \in (0, 2]$ and any martingale f , we have

$$\|Sf\|_p \lesssim_p \|sf\|_p.$$

Proof. Define W, Z as in (5.17). Let $\tau := \inf\{n \mid W_{n+1} > \lambda\}$. Then,

$$\begin{aligned} \mathbb{E}(Z_\infty \wedge \lambda) &\leq \mathbb{E}(Z_\tau + \lambda \mathbf{1}_{\tau < \infty}) \\ &= \sum_k \mathbb{E}(\mathbf{1}_{k \leq \tau} z_k) + \lambda \mathbb{P}\{W_\infty > \lambda\} \\ &\leq \sum_k \mathbb{E}(\mathbf{1}_{k \leq \tau} \mathbb{E}_{k-1} z_k) + \mathbb{E}(W_\infty \wedge \lambda) \\ &\leq 2\mathbb{E}(W_\infty \wedge \lambda). \end{aligned}$$

We conclude by Lemma 1.17. \square

Corollary 1.19. *Let $p \in (0, 2]$ and f be a martingale. Then*

$$\|Mf\|_p \leq 5^{1/p} \|sf\|_p.$$

Proof. Let $\tau := \inf\{n \mid sf_{n+1} > \lambda\}$. Then,

$$\begin{aligned} \mathbb{E}((Mf)^2 \wedge \lambda^2) &\leq \mathbb{E}[(Mf^\tau)^2] + \lambda^2 \mathbb{E}\mathbf{1}_{\tau < \infty} \\ (\text{Doob's inequality}) &\leq 4\mathbb{E}[(f^\tau)^2] + \lambda^2 \mathbb{E}\mathbf{1}_{sf_\infty > \lambda} \\ &\leq 5\mathbb{E}((sf_\infty)^2 \wedge \lambda^2). \end{aligned}$$

We conclude by Lemma 1.17. \square

Lemma 1.20 (Good- λ inequality). *Let f, g be positive random variables, $p \in (0, \infty)$, and suppose that we are given $\beta > 1$ and $\delta, \epsilon \in \mathbb{R}_{>0}$ with $\beta^p \epsilon < 1$. Assume that, for every $\lambda > 0$, we have*

$$\mu\{g > \beta\lambda, f \leq \delta\lambda\} \leq \epsilon \mu\{g > \lambda\}.$$

Then,

$$\mathbb{E}g^p \leq \frac{\delta^{-p}}{\beta^{-p} - \epsilon} \mathbb{E}f^p.$$

Proof. The hypothesis implies

$$\begin{aligned} \mu\{g > \beta\lambda\} &= \mu\{g > \beta\lambda, f \leq \delta\lambda\} + \mu\{g > \beta\lambda, f > \delta\lambda\} \\ &\leq \epsilon \mu\{g > \lambda\} + \mu\{f > \delta\lambda\}. \end{aligned}$$

Using the formula $t^p = p \int_0^t \lambda^{p-1} d\lambda$, we obtain

$$\begin{aligned} \mathbb{E}g^p &= p \mathbb{E} \int_0^g \lambda^{p-1} d\lambda \\ &= p \int_0^\infty \mu\{g > \lambda\} \lambda^{p-1} d\lambda. \end{aligned}$$

Inserting the above estimate, we obtain

$$\mathbb{E}(g/\beta)^p \leq \epsilon \mathbb{E}g^p + \mathbb{E}(f/\delta)^p.$$

Rearranging, we obtain the claim. \square

Lemma 1.21. *Let z_1, z_2, \dots be positive random variables. Then, for every $p \in (0, \infty)$, we have*

$$\mathbb{E}\left(\sum_{k=1}^{\infty} z_k\right)^p \lesssim_p \mathbb{E}\left(Mz \vee \sum_{k=1}^{\infty} \mathbb{E}_{k-1}(z_k)\right)^p.$$

Thus, we see that, for $p \in [2, \infty)$ and any martingale f , we have

$$\|Sf\|_p \lesssim_p \|sf \vee Mdf\|_p.$$

Proof. We continue using notation (5.17). We will verify the hypothesis of Lemma 1.20. Let $\beta, \delta \in \mathbb{R}_{>0}$ with $\beta > \delta + 1$. Consider the stopping times

$$\tau := \inf\{n \mid Z_n > \lambda\}, \tau_\beta := \inf\{n \mid Z_n > \beta\lambda\}, \sigma := \inf\{n \mid (W_{n+1} \vee z_n) > \delta\lambda\}$$

and the process

$$h_n = \sum_{k \leq n} \mathbf{1}_{\tau < k \leq \tau_\beta \wedge \sigma} z_k.$$

Then, if $Z > \beta\lambda$ and $W \vee Mz \leq \delta\lambda$, we have $\sigma = \infty$, so that

$$h_\infty = \sum_{\tau < k \leq \tau_\beta} z_k \geq \beta\lambda - \sum_{k < \tau} z_k - z_\tau \geq \beta\lambda - \lambda - \delta\lambda = (\beta - 1 - \delta)\lambda.$$

Hence,

$$\begin{aligned} \mu\{Z > \beta\lambda, W \vee Mz \leq \delta\lambda\} &\leq \mu\{h_\infty > (\beta - 1 - \delta)\lambda\} \\ &\leq \frac{1}{(\beta - 1 - \delta)\lambda} \mathbb{E}h_\infty \\ &= \frac{1}{(\beta - 1 - \delta)\lambda} \sum_k \mathbb{E}(\mathbf{1}_{\tau < k \leq \tau_\beta \wedge \sigma} z_k) \\ &= \frac{1}{(\beta - 1 - \delta)\lambda} \sum_k \mathbb{E}(\mathbf{1}_{\tau < k \leq \tau_\beta \wedge \sigma} \mathbb{E}_{k-1} z_k) \\ &\leq \frac{1}{(\beta - 1 - \delta)\lambda} \mathbb{E}(\mathbf{1}_{\tau < \infty} \delta\lambda) \\ &= \frac{\delta}{\beta - 1 - \delta} \mu\{Z > \lambda\}. \end{aligned}$$

Now it suffices to take any $\beta > 1$ and δ sufficiently small depending on β . \square

Remark. Most of the material in this section is from [Bur73].

2 Lépingle inequality

For $0 < r < \infty$ and a sequence of random variables $f = (f_n)_n$, the r -variation of f on the interval $[t', t]$ is defined by

$$V^r f_{t',t} := \sup_{J, t' \leq u_0 < \dots < u_J \leq t} \left(\sum_{j=1}^J |f_{u_{j-1}} - f_{u_j}|^r \right)^{1/r}, \quad (2.1)$$

where the supremum is taken over arbitrary increasing sequences. Analogously, $V^\infty f_{t',t} := \sup_{t' \leq u' < u \leq t} |f_{u'} - f_u|$.

We abbreviate $V^r f_t := V^r f_{0,t}$ and $V^r f := V^r f_{0,\infty}$.

The following result is a quantitative version of the martingale convergence theorem.

Theorem 2.1 (Lépingle inequality). *For every $p \in [1, \infty)$, there exists a constant $C_p < \infty$ such that, for every $r > 2$, and every martingale f , we have*

$$\|V^r f\|_{L^p} \leq C_p \frac{r}{r-2} \|Mf\|_{L^p}. \quad (2.2)$$

The main difficulty here is that the supremum in the definition of V^r is taken over not necessarily adapted partitions. We will remedy this by a greedy selection algorithm, stopping as soon as we see a large jump. We have to use stopping rules depending on a parameter m in order to capture jumps of all possible sizes.

Proof. By the monotone convergence theorem, we may assume that f becomes constant after some time N .

Let $M_t := \sup_{t'' \leq t' \leq t} |f_{t'} - f_{t''}|$. For each $m \in \mathbb{N}$, define an adapted partition by

$$\tau_0^{(m)} := 0, \quad \tau_{j+1}^{(m)} := \inf\{t \geq \tau_j^{(m)} \mid |f_t - f_{\tau_j}| \geq 2^{-m} M_t\}. \quad (2.3)$$

Claim. Let $0 \leq t' < t < \infty$ and $m \geq 2$. Suppose that

$$2 < |f_{t'} - f_t| / (2^{-m} M_t) \leq 4. \quad (2.4)$$

Then there exists j with $t' < \tau_j^{(m)} \leq t$ and

$$|f_{t'} - f_t| \leq 8 |f_{\tau_{j-1}^{(m)}} - f_{\tau_j^{(m)}}|. \quad (2.5)$$

Proof of the claim. Let j be the largest integer with $\tau_j^{(m)} \leq t$. We claim that $\tau' > t'$. Suppose for a contradiction that $\tau' < t'$ (the case $\tau' = t'$ is similar but easier). By the hypothesis (2.4) and the assumption that t, t' are not stopping times, we obtain

$$2 \cdot 2^{-m} M_t < |f_{t'} - f_t| \leq |f_{\tau'} - f_{t'}| + |f_{\tau'} - f_t| < 2^{-m} M_{\tau'} + 2^{-m} M_t \leq 2 \cdot 2^{-m} M_t,$$

a contradiction. This shows $\tau' > t'$.

It remains to verify (2.5). Assume that $M_{\tau'} < M_t/2$. Then, for some $\tau'' \in (\tau, t]$, we have $|f_{\tau'} - f_{\tau''}| \geq M_t/2 \geq 2^{-m} M_{\tau''}$, contradicting maximality of τ' . It follows that

$$|f_{\tau_{j-1}^{(m)}} - f_{\tau_j^{(m)}}| \geq 2^{-m} M_{\tau'} \geq 2^{-m} M_t/2 \geq |f_{t'} - f_t|/8. \quad \square$$

Next, for any $0 < \rho < r < \infty$, we will show the pathwise inequality

$$V_t^r(f_t)^r \leq 8^\rho \sum_{m=2}^{\infty} (2^{-(m-2)} M_\infty)^{r-\rho} \sum_{j=1}^{\infty} |f_{\tau_{j-1}}^{(m)} - f_{\tau_j}^{(m)}|^\rho. \quad (2.6)$$

Let (u_l) be any increasing sequence. For each l with $|f_{u_l} - f_{u_{l+1}}| \neq 0$, let $m = m(l) \geq 2$ be such that

$$2 < |f_{u_l} - f_{u_{l+1}}| / (2^{-m} M_{u_{l+1}}) \leq 4.$$

Such m exists because the distance is bounded by $M_{u_{l+1}}$.

Let $j = j(l)$ be given by the Claim above with $t' = u_l$ and $t = u_{l+1}$. Then

$$|f_{u_l} - f_{u_{l+1}}| \leq 8^\rho |f_{\tau_{j-1}}^{(m)} - f_{\tau_j}^{(m)}|^\rho \cdot (4 \cdot 2^{-m} M_{u_{l+1}})^{r-\rho}.$$

Since each pair (m, j) occurs for at most one l , this implies

$$\sum_l |f_{u_l} - f_{u_{l+1}}|^r \leq 8^\rho \sum_{m,j} |f_{\tau_{j-1}}^{(m)} - f_{\tau_j}^{(m)}|^\rho \cdot (2^{-(m-2)} M_\infty)^{r-\rho}.$$

Taking the supremum over all increasing sequences (u_l) , we obtain (2.6).

Since we assumed that f_n becomes independent of n for sufficiently large n , we have

$$M_\infty \leq V_t^r(f_t) < \infty.$$

Substituting this inequality in (2.6) and canceling $V_t^r(f_t)^{r-2}$ on both sides, we obtain

$$V_t^r(f_t)^\rho \leq 8^\rho \sum_{m=2}^{\infty} 2^{-(m-2)(r-\rho)} \sum_{j=1}^{\infty} |f_{\tau_{j-1}}^{(m)} - f_{\tau_j}^{(m)}|^\rho. \quad (2.7)$$

By the optional sampling theorem, for each m , the sequence $(f_{\tau_j}^{(m)})_j$ is a martingale with respect to the filtration $(\mathcal{F}_{\tau_j}^{(m)})_j$. Let

$$S_{(m)} := \left(\sum_{j=1}^{\infty} |f_{\tau_{j-1}}^{(m)} - f_{\tau_j}^{(m)}|^2 \right)^{1/2}$$

denote the square function of this martingale. Then (2.7) implies

$$V^r f \leq 8 \left(\sum_{m=2}^{\infty} 2^{-(m-2)(r-2)} S_{(m)}^2 \right)^{1/2} \leq 8 \sum_{m=2}^{\infty} 2^{-(m-2)(r-2)/2} S_{(m)}$$

By Minkowski's inequality, this implies

$$\|V^r f\|_{L^p} \leq 8 \sum_{m=2}^{\infty} 2^{-(m-2)(r-2)/2} \|S_{(m)}\|_{L^p}.$$

Applying the BDG inequality for the martingales $(f_{\tau_j}^{(m)})_j$, we obtain

$$\begin{aligned} \|V^r f\|_{L^p} &\lesssim_p \sum_{m=2}^{\infty} 2^{-(m-2)(r-2)/2} \|\sup_j |f_{\tau_j}^{(m)}|\|_{L^p} \\ &\lesssim \frac{r}{r-2} \|Mf\|_{L^p}. \end{aligned} \quad \square$$

Remark. The dependence of the constant in Theorem 2.1 on the variation exponent r can be improved to $\sqrt{\frac{r}{r-2}}$ using a vector-valued BDG inequality. In fact, a slightly more careful version of the above argument already shows this for $p \in [2, \infty)$.

2.1 Martingale convergence

Corollary 2.2 (Martingale convergence). *Let f be a martingale with $\|f\|_1 < \infty$. Then, for every $r > 2$, $V^r f$ is finite a.s. In particular, the sequence $(f_n)_n$ converges a.s.*

Proof. Fixing $\lambda < \infty$ and $N \in \mathbb{N}$, consider the stopping time

$$\tau := \inf\{n \mid |f_n| > \lambda\}.$$

Then $Mf^{\tau \wedge N} \leq \lambda \vee |f_{\tau \wedge N}|$. Since $\tau \wedge N$ is a bounded stopping time, by the optional sampling theorem, we have

$$\|Mf^{\tau \wedge N}\|_1 \leq \lambda + \|f_{\tau \wedge N}\|_1 \leq \lambda + \|f\|_1.$$

By Lépingle's inequality, for $r > 2$, we obtain

$$\|V^r f^{\tau \wedge N}\|_1 \lesssim \lambda + \|f\|_1.$$

Since the right-hand side does not depend on N and by the monotone convergence theorem, we obtain

$$\|V^r f^\tau\|_1 \lesssim \lambda + \|f\|_1.$$

In particular, the function $V^r f^\tau$ is finite a.s. On the set $\{Mf < \lambda\}$, we have $f = f^\tau$, so that $V^r f$ is finite a.s. on this set.

Taking the union over $\lambda \in \mathbb{N}$, we find that $V^r f$ is finite a.s. on the set $\{Mf < \infty\}$. By Doob's maximal inequality, this set has full probability. \square

We have shown that, for an L^1 martingale f , the limit $f_\infty := \lim_{n \rightarrow \infty} f_n$ exists a.s. We recall the condition under which this convergence also holds in L^1 . We start with an example that shows that this is not always the case.

Example 2.3. Consider $\Omega = [0, 1]$ with the dyadic filtration. The sequence

$$f_n = 2^n \mathbf{1}_{[0, 2^{-n}]}$$

is a martingale and converges pointwise a.s. to 0. On the other hand, $\mathbb{E}f_n = 1 \neq 0 = \mathbb{E}0$.

Definition 2.4. A set F of real random variables is called *uniformly integrable* if

$$\lim_{\lambda \rightarrow \infty} \sup_{f \in F} \mathbb{E}(\mathbf{1}_{|f| > \lambda} |f|) = 0.$$

Example 2.5. If $f \in L^1$, then the singleton $\{f\}$ is uniformly integrable.

Uniform integrability is useful because it is the minimal hypothesis under which an analogue of the dominated convergence theorem holds. We recall this analogue.

Proposition 2.6 (Dominated convergence). *For every sequence of random variables $f = (f_n)$, the following are equivalent:*

1. f is Cauchy in L^1 ,
2. f is uniformly integrable and Cauchy in probability, that is,

$$(\forall \epsilon > 0)(\exists N \in \mathbb{N})(\forall m, n \geq N)\mu\{|f_n - f_m| > \epsilon\} < \epsilon.$$

In relation with martingales, it is important that uniform integrability is preserved under conditional expectation. This can be seen using the following characterization.

Lemma 2.7. *Let F be a set of random variables. The following are equivalent.*

1. F is uniformly integrable.

2. $\sup_{f \in F} \mathbb{E}|f| < \infty$ and

$$\lim_{\mu(A) \rightarrow 0} \sup_{f \in F} \mathbb{E}(\mathbf{1}_A |f|) = 0. \quad (2.8)$$

Proof. Without loss of generality, all r.v. in F are positive.

\implies : For every $\lambda > 0$, we have

$$\mathbb{E}f \leq \lambda + \mathbb{E}(\mathbf{1}_{f > \lambda} f).$$

Taking the supremum over $f \in F$ on both sides and choosing λ such that $\sup_{f \in F} \mathbb{E}(\mathbf{1}_{f > \lambda} f) < \infty$, we see that $\sup_{f \in F} \mathbb{E}|f| < \infty$.

Moreover,

$$\mathbb{E}(\mathbf{1}_A f) \leq \mu(A)\lambda + \mathbb{E}(\mathbf{1}_{f > \lambda} f).$$

Taking first the supremum over $f \in F$, and then $\lim_{\mu(A) \rightarrow 0}$, we obtain

$$\lim_{\mu(A) \rightarrow 0} \sup_{f \in F} \mathbb{E}(\mathbf{1}_A f) \leq \sup_{f \in F} \mathbb{E}(\mathbf{1}_{f > \lambda} f).$$

Taking $\lim_{\lambda \rightarrow 0}$, we obtain (2.8).

\impliedby : We have $\mu\{f > \lambda\} \leq \lambda^{-1} \mathbb{E}f$. Taking the supremum over $f \in F$, we obtain

$$\sup_{f \in F} \mu\{f > \lambda\} \leq \lambda^{-1} \sup_{f \in F} \mathbb{E}f.$$

The RHS converges to 0 as $\lambda \rightarrow \infty$. The second hypothesis now implies that F is uniformly integrable. \square

Lemma 2.8. *Let $F \subset L^1(\Omega, \mathcal{F})$ be uniformly integrable. Then the set*

$$\{\mathbb{E}(f|\mathcal{F}') \mid f \in F, \mathcal{F}' \subseteq \mathcal{F} \text{ sub-}\sigma\text{-algebra}\}$$

is uniformly integrable.

Proof. We verify the characterization above. We have

$$\mathbb{E}|\mathbb{E}(f|\mathcal{F}')| \leq \mathbb{E}\mathbb{E}(|f|\mathcal{F}') = \mathbb{E}|f|.$$

This verifies the first condition.

Fixing $\lambda > 0$, we also have

$$\mathbb{E}(\mathbf{1}_A |\mathbb{E}(f|\mathcal{F}')|) \leq \mathbb{E}(\mathbf{1}_A \mathbb{E}(|f|\mathcal{F}')) = \mathbb{E}(\mathbb{E}(\mathbf{1}_A |\mathcal{F}')|f|) \leq \lambda \mathbb{E}|f| + \mathbb{E}(\mathbf{1}_{\mathbb{E}(\mathbf{1}_A |\mathcal{F}') > \lambda} |f|)$$

Since $\mu\{\mathbb{E}(\mathbf{1}_A |\mathcal{F}') > \lambda\} \leq \lambda^{-1} \mu(A)$ and using (2.8) for the set F , we obtain

$$\lim_{\mu(A) \rightarrow 0} \sup_{f \in F, \mathcal{F}' \subseteq \mathcal{F}} \mathbb{E}(\mathbf{1}_A |\mathbb{E}(f|\mathcal{F}')|) \leq \lambda \mathbb{E}|f|.$$

Since $\lambda > 0$ was arbitrary, this implies (2.8) for the set of conditional expectations. \square

The above results can now be summarized as follows.

Proposition 2.9 (Martingale closure). *For every martingale $f = (f_n)_{n \in \mathbb{N}}$, the following are equivalent.*

1. The set $\{f_n\}_{n \in \mathbb{N}}$ is uniformly integrable.
2. The sequence $(f_n)_{n \in \mathbb{N}}$ converges in L^1 .
3. There exists $f_\infty \in L^1$ such that, for every $n \in \mathbb{N}$, we have $f_n = \mathbb{E}_n f_\infty$.

If the above conditions hold, then we can take $f_\infty = \lim_{n \rightarrow \infty} f_n$ a.s. and in L^1 .

Remark. We have seen that, for a martingale f , we have

$$Mf \in L^1 \implies f \text{ is uniformly integrable} \implies f \in L^1.$$

The second implication cannot be reversed by Example 2.3. The following example shows that the first implication also cannot be reversed, and also that M is not bounded on L^1 .

Example 2.10. Let $\Omega = [0, 1]$ with the dyadic filtration and

$$f_\infty := \sum_{m=0}^{\infty} (m+1)^{-2} 2^m \mathbf{1}_{[2^{-m-1}, 2^{-m}]}$$

and $f_n := \mathbb{E}_n f_\infty$. Then, for $x \in (2^{-l-1}, 2^{-l})$, we have

$$Mf(x) = f_l(x) = 2^l \sum_{m=l}^{\infty} \int_0^{2^{-l}} (m+1)^{-2} 2^m \mathbf{1}_{[2^{-m-1}, 2^{-m}]}(y) dy = 2^{l-1} \sum_{m=l}^{\infty} (m+1)^{-2} \sim (l+1)^{-1} 2^l.$$

We see that $Mf \notin L^1$.

3 Vector-valued inequalities

Remark. The norm on a Banach space X will be sometimes denoted by $Xf := \|f\|_X$. This is consistent with operator notation, in that every norm is also a sublinear operator.

Lemma 3.1. *Let X be a Banach space. Let T be a subadditive operator that maps X -valued martingales starting at 0 to $\mathbb{R}_{\geq 0}$ -valued increasing adapted processes starting at 0 such that, for every stopping time τ , we have*

$$Tf_\tau = T(f^\tau)_\tau,$$

and for every time t we have

$$Tf_t \leq Tf_{t-1} + Xdf_t.$$

Let U be another operator satisfying the same properties as T .

Assume that, for some $r \in (1, \infty)$ and every X -valued martingale f , we have

$$\mathbb{P}\{Tf > \lambda\} \lesssim \lambda^{-r} (L^r Uf)^r. \quad (3.1)$$

Then, for every $q \in [1, \infty)$ and every X -valued martingale f , we have

$$L^q Tf \lesssim_q L^q Uf + L^q MXdf. \quad (3.2)$$

The formulation is flexible enough to apply to any combination $\{T, U\} = \{XM, XS\}$, and also some other operators, such as martingale transforms or r -variation norms.

Proof. Consider first the case $q = 1$. We use the Davis decomposition $f = f^{\text{pred}} + f^{\text{bv}}$. Of the two possible generalizations of the Davis decomposition to X -valued martingales, we use the one in which the absolute value is replaced by the X -norm (the alternative would be to make a decomposition for each k separately). Specifically, we have

$$Xdf_n^{\text{pred}} \leq 2MXdf_{n-1}, \quad \mathbb{E} \sum_n Xdf_n^{\text{bv}} \leq 2\mathbb{E}MXdf.$$

For $\lambda > 0$, define the stopping time

$$\tau := \inf\{t \mid Uf_t^{\text{pred}} > \lambda \text{ or } MXdf_t > \lambda\}.$$

We claim that

$$Uf_\tau^{\text{pred}} \leq Uf^{\text{pred}} \wedge 3\lambda. \quad (3.3)$$

Indeed, the first bound is trivial. The second bound is trivial if $\tau = \infty$, so assume $\tau \in (0, \infty)$. Then, by properties of the Davis decomposition, we have

$$Uf_\tau^{\text{pred}} \leq Uf_{\tau-1}^{\text{pred}} + Xdf_\tau^{\text{pred}} \leq \lambda + 2MXdf_{\tau-1} \leq 3\lambda.$$

Also,

$$\begin{aligned} \{Tf^{\text{pred}} > \lambda\} &\subseteq \{Tf_\tau^{\text{pred}} > \lambda\} \cup \{\tau < \infty\} \\ &\subseteq \{Tf_\tau^{\text{pred}} > \lambda\} \cup \{Uf^{\text{pred}} > \lambda\} \cup \{MXdf_\tau > \lambda\}. \end{aligned}$$

By the layer cake formula,

$$\begin{aligned} \|Tf^{\text{pred}}\|_{L^1} &= \int_0^\infty \mathbb{P}\{Tf^{\text{pred}} > \lambda\} d\lambda \\ &\leq \int_0^\infty \mathbb{P}\{Tf_\tau^{\text{pred}} > \lambda\} d\lambda + \int_0^\infty \mathbb{P}\{Uf^{\text{pred}} > \lambda\} d\lambda \\ &\quad + \int_0^\infty \mathbb{P}\{MXdf > \lambda\} d\lambda =: I + II + III. \end{aligned}$$

The term *III* is easy to bound by the layer cake formula.

In the term *II*, we use the properties of the Davis decomposition to estimate

$$\begin{aligned} II &= \|Uf^{\text{pred}}\|_{L^1} \\ &\leq \|Uf\|_{L^1} + \|Uf^{\text{bv}}\|_{L^1} \\ &\leq \|Uf\|_{L^1} + \left\| \sum_n Xdf_n^{\text{bv}} \right\|_{L^1} \\ &\leq \|Uf\|_{L^1} + 2\|MXdf\|_{L^1}. \end{aligned}$$

Using the hypothesis (3.1) and (3.3), we bound the first term by

$$\begin{aligned} I &\lesssim \int_0^\infty \lambda^{-r} \|Uf_\tau^{\text{pred}}\|_{L^r}^r d\lambda \\ &\leq \int_0^\infty \lambda^{-r} \|Uf^{\text{pred}} \wedge 3\lambda\|_{L^r}^r d\lambda \\ &= \mathbb{E} \int_0^\infty \min(\lambda^{-r} (Uf^{\text{pred}})^r, 3^r) d\lambda \\ &\lesssim \mathbb{E} Uf^{\text{pred}} = II, \end{aligned}$$

and we reuse the previously established estimate for *II*.

We have shown (3.2) for $q = 1$, and we will now extend this claim to $q > 1$. In doing so, we may replace Uf by $(Uf + MXdf)/2$. This has the effect that we may omit the second summand on the RHS of (3.2).

For $q > 1$, we use the Garsia-Neveu lemma. For a martingale f , let ${}^\tau f := f - f^\tau$ be the martingale f started at (stopping) time τ .

For any $t \in \mathbb{N}$ and $B \in \mathcal{F}_t$, we can apply the case $q = 1$ to the martingale $\mathbf{1}_B {}^t f$, and obtain

$$\mathbb{E}_t T {}^t f \lesssim \mathbb{E}_t U {}^t f.$$

Note that

$$\begin{aligned} Tf_\infty - Tf_t &= Tf_\infty - T(f^t)_\infty \leq T(f - f^t)_\infty = T {}^t f_\infty, \\ U {}^t f_\infty &\leq (Uf + Uf^t)_\infty \leq 2Uf_\infty, \end{aligned}$$

so the conditions of Corollary 1.12 hold for $A_t = Tf_{(t-1) \vee 0}$ and $\xi \sim Uf_\infty$. Corollary 1.12 now shows the claim. \square

Monotonicity of Uf is used in estimates for both I and II above.

Corollary 3.2 (ℓ^r valued BDG inequality). *For every $k \in \mathbb{N}$, let $f_k = (f_{k,n})_n$ be a real-valued martingale starting at 0. Then, for every $q, r \in [1, \infty)$, we have*

$$L^q \ell_k^r M f_k \lesssim_q L^q \ell_k^r S f_k.$$

Proof. For $r = q$, this follows from the scalar-valued Doob's inequality and Fubini. For general q , we apply Lemma 3.1 with $X = \ell^r$. \square

Toy example: better constant in Lépingle.

3.1 Weighted and vector-valued Doob inequalities

Let $w \in L^1(\Omega, \mathcal{F}_\infty)$ be a positive function, that we will call a *weight*. We write $w_k := \mathbb{E}_k w$, $w^* := \sup_k w_k$, and $wB := \int_B w \, d\mu$ for $B \in \mathcal{F}_\infty$.

Lemma 3.3. *Let f be an $\mathbb{R}_{\geq 0}$ -valued submartingale (recall that this means $f_k \leq \mathbb{E}_k f_n$ for $k \leq n$). Then, for every $\lambda > 0$ and $n \in \mathbb{N}$, we have*

$$\lambda w\{Mf_n > \lambda\} \leq \int_{\{Mf_n > \lambda\}} f_n M w_n \, d\mu.$$

Proof. For a fixed λ , define the stopping time

$$\tau := \inf\{k \mid f_k > \lambda\}.$$

Then, $\{Mf_n > \lambda\} = \{\tau \leq n\}$. Hence,

$$\begin{aligned} \lambda w\{Mf_n > \lambda\} &= \lambda \sum_{k \leq n} w\{\tau = k\} \\ &\leq \sum_{k \leq n} \int_{\{\tau = k\}} f_k w_k \, d\mu \\ (\text{submartingale property}) &\leq \sum_{k \leq n} \int_{\{\tau = k\}} f_n w_k \, d\mu \\ &\leq \int_{\{Mf_n > \lambda\}} f_n w^* \, d\mu. \end{aligned} \quad \square$$

We recall that, for a function f on a measure space (Ω, μ) and $p \in [1, \infty)$, the $L^{p, \infty}$ quasinorm is defined by

$$\|f\|_{L^{p, \infty}} := \sup_{\lambda > 0} \lambda \mu\{|f| > \lambda\}^{1/p}.$$

For $p = \infty$, we have $\|f\|_{L^{p, \infty}} = \|f\|_{L^\infty}$.

Theorem 3.4 (Marcinkiewicz/real interpolation, see e.g. [Gra14]). *Let (Ω, μ) and $(\tilde{\Omega}, \tilde{\mu})$ be σ -finite measure spaces. Let $0 < p_0 < p_1 \leq \infty$. Let*

$$T : L^{p_0}(\Omega) + L^{p_1}(\Omega) \rightarrow L^0(\tilde{\Omega})$$

be a subadditive operator such that, for every function f , we have

$$\|Tf\|_{L^{p_j, \infty}(\tilde{\Omega})} \leq A_j \|f\|_{L^{p_j}(\Omega)}, \quad j = 0, 1.$$

Then, for every $p \in (p_0, p_1)$, we have

$$\|Tf\|_{L^p(\tilde{\Omega})} \lesssim_p \|f\|_{L^p(\Omega)}.$$

Lemma 3.5. *Let $p \in (1, \infty)$, f be a submartingale, and w a weight. Then,*

$$\|Mf\|_{L^p(w)} \lesssim_p \|f\|_{L^p(w^*)}$$

Proof. We will show the following more precise estimate:

$$\|Mf_N\|_{L^p(w)} \lesssim_p \|f_N\|_{L^p(Mw_N)}, \quad (3.4)$$

from which the claim follows by letting $N \rightarrow \infty$. The main advantage of this formulation is that the right-hand side increases in N , which would not be the case if we would use the weight w^* instead of Mw_N .

Lemma 3.3 implies in particular

$$\|Mf_N\|_{L^{1,\infty}(w)} \lesssim \|f_N\|_{L^1(Mw_N)}.$$

The estimate

$$\|Mf_N\|_{L^\infty(w)} \lesssim \|f_N\|_{L^\infty(Mw_N)}$$

is easy to see. By real interpolation, these two bounds imply (3.4). \square

Proposition 3.6 (Vector-valued maximal inequality). *Let $p \in (1, \infty)$ and $r \in (1, \infty]$. Then, for any sequence of martingales $f_k = (f_{k,n})_n$, we have*

$$L^p \ell_k^r M f_k \lesssim_{q,r} L^p \ell_k^r f_k.$$

Proof. The case $r = \infty$ follows from the scalar case, because $\ell^\infty f_k$ is a submartingale.

The case $p = r$ also follows from the scalar case and Fubini.

For $1 < r < p < \infty$, let $\rho := (p/r)' = p/(p-r)$. Then, for some weight $w \in L^\rho$ with $\|w\|_\rho \leq 1$, we have

$$\begin{aligned} \|\ell_k^r M f_k\|_p^r &\lesssim \left(\mathbb{E} \left(\sum_k (M f_k)^r \right)^{p/r} \right)^{r/p} \\ &= \mathbb{E} \sum_k (M f_k)^r w \\ &\lesssim \mathbb{E} \sum_k |f_k|^r w^* \end{aligned}$$

and we conclude using Hölder's inequality and boundedness of the maximal operator on L^ρ .

For $1 < p < r < \infty$, let $s \in (1, p)$. Then, for some sequence of weights with

$$\|\ell_k^{(r/s)'} w_k\|_{(p/s)'} \sim 1,$$

we have

$$\begin{aligned} \|\ell_k^r M f_k\|_p^s &= \|\ell_k^{r/s} (M f_k)^s\|_{p/s} \\ &\sim \int \sum_k w_k (M f_k)^s \\ &\lesssim \int \sum_k w_k^* |f_k|^s \\ &\leq \|\ell_k^{(r/s)'} w_k^*\|_{(p/s)'} \|\ell_k^{r/s} |f_k|^s\|_{p/s}. \end{aligned}$$

Since $(r/s)' < (p/s)'$, from the previously shown case, we obtain

$$\|\ell_k^{(r/s)'} w_k^*\|_{(p/s)'} \lesssim \|\ell_k^{(r/s)'} w_k\|_{(p/s)'} \sim 1. \quad \square$$

Remark. Most of this section is from [Hyt+16, Section 3].

4 Rough paths

In this section, we discuss how to define $\int a dg$ for not very regular functions a, g . The intended application is that g is a sample path of a martingale, which only has bounded r -variation for $r > 2$, so that we cannot use Stieltjes integration.

The function a is assumed to have the same regularity as g , because we want the theory to be suitable for solving equations like $f(t) = \int_0^t a(s, f(s)) dg(s)$.

4.1 Young integration with Hölder functions

We begin with a criterion for convergence of Riemann-like sums. Let $\Delta := \{(s, t) \mid 0 \leq s \leq t \leq T\}$. For any $\Xi : \Delta \rightarrow \mathbb{R}$ and a partition $0 = \pi_0 < \dots < \pi_J = T$, write

$$\mathcal{J}^\pi \Xi_{0,T} := \sum_{j=1}^J \Xi_{\pi_{j-1}, \pi_j}.$$

We will discuss sufficient conditions for the convergence of these sums along the directed (by set inclusion) set of partitions.

For a mapping $\Xi : \Delta \rightarrow E$ (E a vector space), we write

$$\delta \Xi_{s,u,t} := \Xi_{s,t} - \Xi_{s,u} - \Xi_{u,t}.$$

A *control* is a function $\omega : \Delta \rightarrow [0, \infty)$ that is superadditive in the sense that

$$\omega(s, t) + \omega(t, u) \leq \omega(s, u)$$

for all $s \leq t \leq u$. This implies in particular that $\omega(t, t) = 0$ for all t .

Lemma 4.1. *Let ω be a control and $\theta > 1$. Let $(E, |\cdot|)$ be a Banach space and $\Xi : \Delta \rightarrow E$ such that, for every $0 \leq s \leq t \leq u \leq T$, we have*

$$|\delta \Xi_{s,t,u}| \leq \omega(s, u)^\theta. \quad (4.1)$$

Then, for every partition $0 = \pi_0 \leq \dots \leq \pi_J = T$, we have

$$|\Xi_{0,T} - \mathcal{J}^\pi \Xi_{0,T}| \lesssim (\theta - 1)^{-1} \omega(0, T)^\theta.$$

Proof. By induction on the partition size J , we will show the bound

$$|\Xi_{0,T} - \mathcal{J}^\pi \Xi_{0,T}| \leq \sum_{k=1}^{J-1} (2/k)^\theta \omega(0, T)^\theta.$$

For $J = 1$, we have $\mathcal{J}^\pi \Xi_{0,T} = \Xi_{0,T}$, which serves as the induction base.

Suppose that the claim is known for all partitions of size J and let π be a partition of size $J + 1$. By superadditivity of ω , we have

$$\sum_{k=0}^{J-1} \omega(\pi_k, \pi_{k+2}) = \sum_{k \leq J-1 \text{ even}} \omega(\pi_k, \pi_{k+2}) + \sum_{k \leq J-1 \text{ odd}} \omega(\pi_k, \pi_{k+2}) \leq 2\omega(0, T).$$

Hence, there exists k with $\omega(\pi_k, \pi_{k+2}) \leq 2\omega(0, T)/J$. Let $\pi' := \pi \setminus \{\pi_{k+1}\}$. Then,

$$|\mathcal{J}^\pi \Xi_{0,T} - \mathcal{J}^{\pi'} \Xi_{0,T}| = |\delta \Xi_{\pi_k, \pi_{k+1}, \pi_{k+2}}| \leq \omega(\pi_k, \pi_{k+2})^\theta \leq (2/J)^\theta \omega(0, T).$$

Applying the inductive hypothesis to π' , we obtain the claim. \square

Proposition 4.2 (Sewing). *Let ω be a control such that*

$$\limsup_{\pi} \sup_j \omega(\pi_j, \pi_{j+1}) = 0. \quad (4.2)$$

Let $\Xi : \Delta \rightarrow E$ be such that (4.1) holds. Then, the limit

$$\mathcal{J}\Xi_{0,T} := \lim_{\pi} \mathcal{J}^{\pi}\Xi_{0,T}$$

exists and satisfies

$$|\mathcal{J}\Xi_{0,T} - \Xi_{0,T}| \lesssim (\theta - 1)^{-1} \omega(0, T)^{\theta}.$$

Proof. For any partitions $\pi \subseteq \pi'$, by Lemma 4.1, we have

$$\begin{aligned} |\mathcal{J}^{\pi}\Xi_{s,t} - \mathcal{J}^{\pi'}\Xi_{s,t}| &\leq \sum_j |\Xi_{\pi_j, \pi_{j+1}} - \mathcal{J}^{\pi'}\Xi_{\pi_j, \pi_{j+1}}| \\ &\lesssim \sum_j \omega(\pi_j, \pi_{j+1})^{\theta} \\ &\lesssim \sup_j \omega(\pi_j, \pi_{j+1})^{\theta-1} \sum_j \omega(\pi_j, \pi_{j+1}) \\ &\lesssim \omega(s, t) \sup_j \omega(\pi_j, \pi_{j+1})^{\theta-1}. \end{aligned}$$

The hypothesis (4.2) implies that $\pi \mapsto \mathcal{J}^{\pi}\Xi_{s,t}$ is a Cauchy net. □

Example 4.3 (Young integral). Let $a, g : [0, T] \rightarrow \mathbb{R}$ be α -Hölder functions and

$$\Xi_{s,t} := a_s(g_t - g_s).$$

Then,

$$\begin{aligned} \delta\Xi_{s,t,u} &= \Xi_{s,u} - \Xi_{s,t} - \Xi_{t,u} \\ &= a_s(g_u - g_s) - a_s(g_t - g_s) - a_t(g_u - g_t) \\ &= a_s(g_u - g_t) + a_s(g_t - g_s) - a_s(g_t - g_s) - a_t(g_u - g_t) \\ &= (a_s - a_t)(g_u - g_t). \end{aligned}$$

In particular, $\Xi_{s,t,u} \lesssim |s - t|^{\alpha} |u - t|^{\alpha} \leq |u - s|^{2\alpha}$. Hence, provided that $2\alpha > 1$, the hypothesis of Proposition 4.2 holds with $\theta = 2\alpha$ and $\omega(s, t) \sim |t - s|$.

4.2 Young integration with V^r functions

If the function $g : [0, T] \rightarrow \mathbb{R}$ has bounded r -variation, then

$$\omega_{g,r}(s, t) := (V^r g_{s,t})^r = \sup_{s \leq \pi_0 \leq \dots \leq \pi_J \leq t} \sum_{j=1}^J |g_{\pi_j} - g_{\pi_{j-1}}|^r$$

is a control. If both $a, g \in V^r$, then, in the situation of Example 4.3, we have

$$|\delta\Xi_{s,t,u}| \leq \omega_{a,r}(s, t)^{1/r} \omega_{g,r}(t, u)^{1/r}. \quad (4.3)$$

This implies in particular (4.1) with $\omega = \omega_{a,r} + \omega_{g,r}$ and $\theta = 2/r$. However, (4.2) only holds for this control if a, g are also continuous. In order to integrate functions with jumps, we will show a version of the sewing lemma under the condition (4.3).

Definition 4.4. If f is a function defined on a suitable interval, we write

$$f(t+) := \lim_{t' \rightarrow t, t' > t} f(t'), \quad f(t-) := \lim_{t' \rightarrow t, t' < t} f(t'),$$

provided that the respective limits exist.

Lemma 4.5. *Let ω be a control. Then, for every $\epsilon > 0$, there exists a partition π such that*

$$\max_j (\omega(\pi_{j-1}+, \pi_j) \wedge \omega(\pi_{j-1}, \pi_j-)) \leq \epsilon. \quad (4.4)$$

Proof. We select the partition greedily starting with $\pi_0 = 0$. It will be clear from the construction that the claim (4.4) holds. If π_j has been already selected and $\pi_j < T$, we distinguish two cases.

Case 1: if $\omega(\pi_j, \pi_j+) < \epsilon$, then we let

$$\pi_{j+1} := \sup\{t > \pi_j \mid \omega(\pi_j, t) < \epsilon\}.$$

Then $\omega(\pi_j, \pi_{j+1}-) \leq \epsilon$.

Case 2: if $\omega(\pi_j, \pi_j+) \geq \epsilon$, then we choose any $\pi_{j+1} \in (\pi_j, T]$ with $\omega(\pi_j+, \pi_{j+1}) \leq \epsilon$.

To see that such π_{j+1} exists, suppose for a contradiction that, for every $t \in (\pi_j, T]$, we have $\omega(\pi_j+, t) > \epsilon$. Choose recursively $t_0 = T$ and, given t_k , let $t_{k+1} \in (\pi_j, t_k)$ be such that $\omega(t_{k+1}, t_k) > \epsilon$. By superadditivity, we obtain

$$\omega(\pi_j, T) \geq \sum_k \omega(t_{k+1}, t_k) = +\infty,$$

a contradiction.

The selection of π_j 's ends after finitely many steps, because otherwise we would have $\omega(\pi_j, \pi_{j+1}+) \geq \epsilon$ for every j , which in turn implies $\omega(\pi_j, \pi_{j+2}) \geq \epsilon$, and summing over even j we obtain a contradiction with the superadditivity of ω . \square

Theorem 4.6 (Sewing with jumps). *Let $\omega_{1,n}, \omega_{2,n}$ be controls and $\alpha_{1,n}, \alpha_{2,n} \geq 0$ with $\alpha_{1,n} + \alpha_{2,n} > 1$ for all $n \in \{1, \dots, N\}$.*

Let $\Xi : \Delta \rightarrow E$. Assume

$$|\delta \Xi_{s,u,t}| \leq \sum_{n=1}^N \omega_{1,n}^{\alpha_{1,n}}(s, u) \omega_{2,n}^{\alpha_{2,n}}(u, t),$$

Then the following net limit exists,

$$\mathcal{J}\Xi_{0,T} := \lim_{\pi} \mathcal{J}^{\pi}\Xi_{0,T},$$

and one has the following estimate:

$$|\mathcal{J}\Xi_{0,T} - \Xi_{0,T}| \lesssim \sum_{n=1}^N \omega_{1,n}^{\alpha_{1,n}}(0, T-) \omega_{2,n}^{\alpha_{2,n}}(0+, T),$$

with C depending only on $\min_n \alpha_{1,n} + \alpha_{2,n}$.

Proof. Let

$$\theta := \min_n \alpha_{1,n} + \alpha_{2,n} > 1$$

and

$$\omega(s, t) := \sum_{n=1}^N \omega_{1,n}^{\alpha_{1,n}/\theta}(s, t-) \omega_{2,n}^{\alpha_{2,n}/\theta}(s+, t).$$

Note that the functions $(s, t) \mapsto \omega_{1,n}(s, t-)$ and $(s, t) \mapsto \omega_{2,n}(s+, t)$ are also controls. The function ω is a control, because, by Hölder's inequality,

$$\begin{aligned} \omega_{1,n}^{\alpha_{1,n}/\theta}(s, u) \omega_{2,n}^{\alpha_{2,n}/\theta}(s, u) &\geq (\omega_{1,n}(s, t) + \omega_{1,n}(t, u))^{\alpha_{1,n}/\theta} (\omega_{2,n}(s, t) + \omega_{2,n}(t, u))^{\alpha_{2,n}/\theta} \\ &\geq \omega_{1,n}(s, t)^{\alpha_{1,n}/\theta} \omega_{2,n}(s, t)^{\alpha_{2,n}/\theta} + \omega_{1,n}(t, u)^{\alpha_{1,n}/\theta} \omega_{2,n}(t, u)^{\alpha_{2,n}/\theta}. \end{aligned}$$

It follows from Lemma 4.5 applied to the control $\sum_n (\omega_{1,n} + \omega_{2,n})$ that (4.2) holds. Since $|\delta \Xi_{s,t,u}| \leq \omega(s, u)^\theta$, we can apply Proposition 4.2. \square

Theorem 4.6 allows us to integrate under the hypothesis (4.3), but only for $r < 2$. By the law of the iterated logarithm, paths of the Brownian motion are not in V^r for any $r \leq 2$.

There are two options to refine the above reasoning in order to obtain an integration theory suitable for martingales.

- The classical stochastic integration uses orthogonality between martingale increments. We will return to that.
- Rough integration theory replaces Riemann–Stieltjes sums by higher order quadrature schemes. Higher order quadrature improves the order of convergence (under suitable regularity assumptions). In the rough setting, it improves from having no convergence to having convergence.

Another motivation for us to look at the rough integration theory is that it informs us about the estimates that we will want when we return to stochastic integration.

4.3 Rough paths

Throughout this section, we fix some $r \in [2, 3)$.

An r -rough path (with values in \mathbb{R}^d) consists of

$$X : [0, T] \rightarrow \mathbb{R}^d, \quad \mathbb{X} : \Delta \rightarrow \mathbb{R}^{d \times d}$$

such that $X \in V^r, \mathbb{X} \in V^{r/2}$, where

$$V^r \mathbb{X}_{s,t} = \sup_{s \leq u_0 \leq \dots \leq u_J \leq t} \left(\sum_j |\mathbb{X}_{u_j, u_{j+1}}|^r \right)^{1/r}.$$

and *Chen's relation* holds for $s < t < u$:

$$\mathbb{X}_{s,u} = \mathbb{X}_{s,t} + \mathbb{X}_{t,u} + (X_t - X_s) \otimes (X_u - X_t). \quad (4.5)$$

Remark. If X is sufficiently regular for the integral to be defined, one can take

$$\mathbb{X}_{s,t} := \int_s^t (X_{u-} - X_s) \otimes dX_u.$$

Next, we introduce the spaces in which one can state and solve differential equations driven by rough paths.

Definition 4.7 (Controlled path). Let $\mathbf{X} = (X, \mathbb{X})$ be an r -rough path in \mathbb{R}^d . An \mathbf{X} -controlled r -rough path in a vector space E consists of $Y : [0, T] \rightarrow E, Y' : [0, T] \rightarrow L(\mathbb{R}^d, E)$ such that $Y' \in V^r$ and

$$R_{s,t}^{\mathbf{Y}, \mathbf{X}} := \delta Y_{s,t} - Y'_s \delta X_{s,t} \in V^{r/2}.$$

Although $R^{\mathbf{Y}, \mathbf{X}}$ depends on all of Y, Y', X , it is commonly abbreviated to R^Y , since the other dependencies are usually unambiguous.

It may be helpful to think about the scalar case $d = d' = 1$, and we will use the scalar notation. All arguments will be, however, formulated in such a way that they work in the vector-valued case, which is important e.g. because one may wish to incorporate time as an additional coordinate in \mathbf{X} .

Example 4.8. $(X, 1)$ is an X -controlled path.

Lemma 4.9 (Rough integral). Let $\mathbf{X} = (X, \mathbb{X})$ be an r -rough path and $\mathbf{Y} = (Y, Y')$ an \mathbf{X} -controlled path in $L(\mathbb{R}^d, \mathbb{R}^{d'})$. Let $\Xi_{s,t} := Y_s \delta X_{s,t} + Y'_s \mathbb{X}_{s,t}$. Then,

$$\int_0^T \mathbf{Y} d\mathbf{X} := \lim_{\pi} \mathcal{J}^\pi \Xi_{0,T}$$

exists and satisfies

$$\left| \int_s^t \mathbf{Y} d\mathbf{X} - \Xi_{s,t} \right| \lesssim V^{r/2} R_{s,t} - V^r X_{s+,t} + V^r Y'_{s,t} - V^{r/2} \mathbb{X}_{s+,t}.$$

Proof.

$$\begin{aligned} \delta \Xi_{s,t,u} &= Y_s \delta X_{s,u} + Y'_s \mathbb{X}_{s,u} - (Y_s \delta X_{s,t} + Y'_s \mathbb{X}_{s,t}) - (Y_t \delta X_{t,u} + Y'_t \mathbb{X}_{t,u}) \\ &= Y_s \delta X_{t,u} + Y'_s (\mathbb{X}_{s,u} - \mathbb{X}_{s,t}) - (Y_t \delta X_{t,u} + Y'_t \mathbb{X}_{t,u}) \\ &= (Y_s - Y_t) \delta X_{t,u} + Y'_s (\mathbb{X}_{t,u} + \delta X_{s,t} \otimes \delta X_{t,u}) - Y'_t \mathbb{X}_{t,u} \\ &= (Y_s - Y_t) \delta X_{t,u} + (Y'_s - Y'_t) \mathbb{X}_{t,u} + Y'_s (\delta X_{s,t} \otimes \delta X_{t,u}) \\ &= -R_{s,t}^{\mathbf{Y}, \mathbf{X}} \delta X_{t,u} + (Y'_s - Y'_t) \mathbb{X}_{t,u}. \end{aligned} \quad (4.6)$$

Hence, we can apply Theorem 4.6. □

Example 4.10. $\int_0^T (X - X_0, 1) \otimes dX = \mathbb{X}_{0,T}$. Indeed, in this case $\delta\Xi = 0$ in (4.6).

This gives even a $V^{r/3}$ approximation of the integral. However, we drop the second order term and only use a $V^{r/2}$ approximation, since we want to run the iteration in the simplest possible space for the given regularity.

4.3.1 Rough differential equations

We consider an initial value problem that is informally described by $Y_0 = y_0$, $dY = \phi(Y) dX$. A rigorous formulation is the following: given a rough path \mathbf{X} in \mathbb{R}^d and a function $\phi : \mathbb{R}^{d'} \rightarrow L(\mathbb{R}^d, \mathbb{R}^{d'})$, we are looking for an \mathbf{X} -controlled path \mathbf{Y} such that

$$Y_T = y_0 + \int_0^T \phi(Y) \mathbf{X}. \quad (4.7)$$

We still have to define what we mean by $\phi(\mathbf{Y})$.

We will find that this problem has a unique solution if $\phi \in C_b^{2,1}$, the space of twice differentiable functions with $\phi, D\phi, D^2\phi$ bounded and $D^2\phi$ Lipschitz.

The proof proceeds by a fixed point argument. First, we need to identify a metric space in which this argument will be run. We abbreviate $\|\mathbf{X}\|_r := V^r X_{0,T}$.

4.3.2 Operations on controlled paths

We have deliberately omitted the condition $Y \in V^r$ from the definition of the controlled path, because it is implicit in the other conditions.

Lemma 4.11 (Implicit bound). *Let \mathbf{X} be a rough path and \mathbf{Y} an \mathbf{X} -controlled path. Then*

$$\|\mathbf{Y}\|_r \leq \|Y'\|_{\text{sup}} \|\mathbf{X}\|_r + \|R^Y\|_{r/2}.$$

Lemma 4.12 (Integration). *Let \mathbf{X} be a rough path and \mathbf{Y} an \mathbf{X} -controlled path. Let $Z_t := \int_0^t Y d\mathbf{X}$ and $Z'_t := Y_t$. Then, $\mathbf{Z} = (Z, Z')$ is an \mathbf{X} -controlled path, and it satisfies*

$$\|R^{\mathbf{Z}}\|_{r/2} \lesssim \|R^Y\|_{r/2} \|\mathbf{X}\|_r + \|Y'\|_r \|\mathbb{X}\|_{r/2} + \|Y'\|_{\text{sup}} \|\mathbb{X}\|_{r/2}. \quad (4.8)$$

Proof.

$$\begin{aligned} R_{s,t}^{\mathbf{Z}} &= \delta Z_{s,t} - Z'_s \delta X_{s,t} \\ &= \int_0^t Y d\mathbf{X} - \int_0^s Y d\mathbf{X} - Y_s \delta X_{s,t} \\ &= \int_s^t Y d\mathbf{X} - Y_s \delta X_{s,t} - Y'_s \mathbb{X}_{s,t} + Y'_s \mathbb{X}_{s,t}. \end{aligned}$$

The last term corresponds to the last term in the conclusion, and the remaining difference is estimated by Lemma 4.9. \square

Lemma 4.13 (Composition with a smooth function). *Let \mathbf{X} be a rough path and \mathbf{Y} an \mathbf{X} -controlled path. Suppose $\phi \in C_b^{1,1}$. Then $\phi(\mathbf{Y}) := (\phi(Y), D\phi(Y)Y')$ is an \mathbf{X} -controlled path, and we have*

$$\|\phi(\mathbf{Y})'\|_r = \|D\phi(Y)Y'\|_r \leq \|D\phi\|_{\text{sup}} \|Y'\|_r + \|D\phi\|_{\text{Lip}} \|Y\|_r \|Y'\|_{\text{sup}}, \quad (4.9)$$

$$\|R^{\phi(\mathbf{Y})}\|_{r/2} \leq \|D\phi\|_{\text{sup}} \|R^Y\|_{r/2} + \frac{1}{2} \|D\phi\|_{\text{Lip}} \|Y\|_r^2. \quad (4.10)$$

Interestingly, this estimate does not depend on \mathbf{X} .

Proof. Adding and subtracting $D\phi(Y_t)Y'_s$, we obtain

$$D\phi(Y_t)Y'_t - D\phi(Y_s)Y'_s = D\phi(Y_t)(Y'_t - Y'_s) + (D\phi(Y_t) - D\phi(Y_s))Y'_s. \quad (4.11)$$

This implies the first estimate.

By Taylor's formula,

$$\begin{aligned} |R_{s,t}^{\phi(Y)}| &= |\phi(Y_t) - \phi(Y_s) - D\phi(Y_s)Y'_s X_{s,t}| \\ &\leq |\phi(Y_t) - \phi(Y_s) - D\phi(Y_s)Y_{s,t}| + |D\phi(Y_s)R_{s,t}^Y| \\ &\leq \frac{1}{2}\|D\phi\|_{Lip}|Y_{s,t}|^2 + \|D\phi\|_{\sup}|R_{s,t}^Y|, \end{aligned}$$

which implies second estimate. \square

The above discussion allows us to construct a space in which the equation (4.23) makes sense.

Lemma 4.14 (Solution space). *For any $\phi \in C_b^{1,1}$ and any $A \in (0, \infty)$, there exists $\epsilon > 0$ such that if $\|X\|_r + \|\mathbb{X}\|_{r/2} < \epsilon$, then the set of \mathbf{X} -controlled paths*

$$\mathcal{Y} = \mathcal{Y}(\mathbf{X}, A) := \{\mathbf{Y} \mid \|Y'\|_r \leq A, \|R^Y\|_{r/2} \leq A^2, \|Y\|_r \leq A/\|\phi\|_{Lip}, \|Y'\|_{\sup} \leq \|\phi\|_{\sup}\} \quad (4.12)$$

is invariant under the mapping

$$\text{Step} : \mathbf{Y} \mapsto (y_0 + \int_0^\cdot \phi(\mathbf{Y}) \, d\mathbf{X}, \phi(\mathbf{Y})) \quad (4.13)$$

for any $y_0 \in \mathbb{R}^{d'}$.

Proof. Implicit constants in this proof are allowed to depend on the $C_b^{1,1}$ norm of ϕ .

Suppose $\mathbf{Y} \in \mathcal{Y}$. Direct estimates show

$$\|\phi(\mathbf{Y})\|_r \leq \|\phi\|_{Lip}\|Y\|_r \leq A, \quad \|\phi(\mathbf{Y})\|_{\sup} \leq \|\phi\|_{\sup}.$$

By Lemma 4.13, we have

$$\|\phi(\mathbf{Y})'\|_r \lesssim A, \quad \|R^{\phi(\mathbf{Y})}\|_{r/2} \lesssim A^2. \quad (4.14)$$

By Lemma 4.12, we obtain

$$\|R^{\int_0^\cdot \phi(\mathbf{Y}) \, d\mathbf{X}}\|_{r/2} \lesssim \epsilon(A^2 + A + 1).$$

By Lemma 4.11, this implies

$$\left\| \int_0^\cdot \phi(\mathbf{Y}) \, d\mathbf{X} \right\|_r \lesssim \epsilon(A^2 + A + 1).$$

Choosing ϵ sufficiently small, we obtain the claim. \square

Remark. Global bounds on ϕ can be replaced by growth conditions, and then ϵ would also depend on y_0 and these growth conditions. There is also a ‘‘rough Gronwall lemma’’ for concatenating local solutions in such a setting.

4.3.3 Contractive iteration

In this section, we will show that the iteration (4.13) is contractive with respect to a suitable metric on the space (4.12). This implies existence and uniqueness of solutions.

For the later purpose of proving continuous dependence of the solution on data, the estimates will involve two rough paths $\mathbf{X}, \tilde{\mathbf{X}}$ and paths $\mathbf{Y}, \tilde{\mathbf{Y}}$ controlled by the respective rough paths. We abbreviate

$$\begin{aligned} \Delta Y &= Y - \tilde{Y}, & \Delta Y' &= Y' - \tilde{Y}', & \Delta R^Y &= R^Y - R^{\tilde{Y}}, & \Delta X &= X - \tilde{X}, & \Delta \mathbb{X} &= \mathbb{X} - \tilde{\mathbb{X}}, \\ \Delta \phi(Y) &= \phi(Y) - \phi(Y'), & \Delta R^{\phi(Y)} &= R^{\phi(Y)} - R^{\phi(\tilde{Y})}. \end{aligned}$$

Lemma 4.15 (Stability of composition). *Suppose $\phi \in C_b^{2,1}$. Let $\mathbf{X}, \tilde{\mathbf{X}}$ be rough paths, \mathbf{Y} a \mathbf{X} -controlled path, and $\tilde{\mathbf{Y}}$ a $\tilde{\mathbf{X}}$ -controlled path. Then,*

$$\|\phi(Y) - \phi(\tilde{Y})\|_r \leq \|D\phi\|_{\text{sup}} \|\Delta Y\|_r + \|D\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}} \|\tilde{Y}\|_r, \quad (4.15)$$

$$\begin{aligned} \|\phi(Y)' - \phi(\tilde{Y})'\|_r &\leq \|D\phi\|_{\text{sup}} \|\Delta Y'\|_r + \|D\phi\|_{\text{Lip}} \|Y\|_r \|\Delta Y'\|_{\text{sup}} \\ &+ \|D\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}} \|\tilde{Y}'\|_r + \|D^2\phi\|_{\text{sup}} \|\Delta Y'\|_r \|\tilde{Y}'\|_{\text{sup}} + \|D^2\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}} \|\tilde{Y}\|_r \|\tilde{Y}'\|_{\text{sup}}, \end{aligned} \quad (4.16)$$

$$\begin{aligned} \|R^{\phi(Y)} - R^{\phi(\tilde{Y})}\|_{r/2} &\leq \|D\phi\|_{\text{sup}} \|\Delta R^Y\|_{r/2} + \|D\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}} \|R^{\tilde{Y}}\|_{r/2} \\ &+ \frac{1}{2} \|D^2\phi\|_{\text{sup}} (\|Y\|_r + \|\tilde{Y}\|_r) \|\Delta Y\|_r + \frac{1}{2} \|D^2\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}} \|\tilde{Y}\|_r^2. \end{aligned} \quad (4.17)$$

Lemma 4.15 almost recovers Lemma 4.13 upon setting $\tilde{\mathbf{Y}} = 0$.

Proof. In order to show (4.15), we write

$$\begin{aligned} &(\phi(Y_t) - \phi(Y_s)) - (\phi(\tilde{Y}_t) - \phi(\tilde{Y}_s)) \\ &= \int_0^1 D\phi(Y_s + r(Y_t - Y_s))(Y_t - Y_s) dr - \int_0^1 D\phi(\tilde{Y}_s + r(\tilde{Y}_t - \tilde{Y}_s))(\tilde{Y}_t - \tilde{Y}_s) dr \\ &= \int_0^1 \left(D\phi(Y_s + r(Y_t - Y_s))(Y_t - Y_s - (\tilde{Y}_t - \tilde{Y}_s)) \right. \\ &\quad \left. + (D\phi(Y_s + r(Y_t - Y_s)) - D\phi(\tilde{Y}_s + r(\tilde{Y}_t - \tilde{Y}_s)))(\tilde{Y}_t - \tilde{Y}_s) \right) dr, \end{aligned} \quad (4.18)$$

and estimate

$$\begin{aligned} &|D\phi(Y_s + r(Y_t - Y_s)) - D\phi(\tilde{Y}_s + r(\tilde{Y}_t - \tilde{Y}_s))| \\ &\leq \|D\phi\|_{\text{Lip}} |(Y_s + r(Y_t - Y_s)) - (\tilde{Y}_s + r(\tilde{Y}_t - \tilde{Y}_s))| \\ &\leq \|D\phi\|_{\text{Lip}} (r|Y_t - \tilde{Y}_t| + (1-r)|Y_s - \tilde{Y}_s|) \\ &\leq \|D\phi\|_{\text{Lip}} \|\Delta Y\|_{\text{sup}}. \end{aligned}$$

In order to show (4.16), we write

$$\begin{aligned} &(D\phi(Y_t)Y_t' - D\phi(\tilde{Y}_t)\tilde{Y}_t') - (D\phi(Y_s)Y_s' - D\phi(\tilde{Y}_s)\tilde{Y}_s') \\ &= D\phi(Y_t)((Y_t' - \tilde{Y}_t') - (Y_s' - \tilde{Y}_s')) + (D\phi(Y_t) - D\phi(Y_s))(Y_s' - \tilde{Y}_s') \\ &+ (D\phi(Y_t) - D\phi(\tilde{Y}_t))(\tilde{Y}_t' - \tilde{Y}_s') + ((D\phi(Y_t) - D\phi(Y_s)) - (D\phi(\tilde{Y}_t) - D\phi(\tilde{Y}_s)))\tilde{Y}_s'. \end{aligned}$$

The first 3 terms contribute the first 3 terms to (4.16). The 4-fold difference in the last term can be written as in (4.18) with $D\phi$ in place of ϕ , which gives the last 2 terms in (4.16).

Now we show (4.17). By inserting $D\phi(Y_s)Y_{s,t}$ and $D\phi(\tilde{Y}_s)\tilde{Y}_{s,t}$, one has

$$\begin{aligned} R_{s,t}^{\phi(Y)} - R_{s,t}^{\phi(\tilde{Y})} &= \phi(Y)_{s,t} - D\phi(Y_s)Y'_s X_{s,t} - (\phi(\tilde{Y})_{s,t} - D\phi(\tilde{Y}_s)\tilde{Y}'_s \tilde{X}_{s,t}) \\ &= \phi(Y)_{s,t} - D\phi(Y_s)Y_{s,t} - (\phi(\tilde{Y})_{s,t} - D\phi(\tilde{Y}_s)\tilde{Y}_{s,t}) \\ &\quad + D\phi(Y_s)R_{s,t}^Y - D\phi(\tilde{Y}_s)R_{s,t}^{\tilde{Y}} \end{aligned} \quad (4.19)$$

By Taylor's formula, the first line in (5.12) equals

$$\int_0^1 (1-r)(D^2\phi(Y_s + rY_{s,t})Y_{s,t}^{\otimes 2} - D^2\phi(\tilde{Y}_s + r\tilde{Y}_{s,t})\tilde{Y}_{s,t}^{\otimes 2}) dr.$$

The integrand can be written in the form

$$D^2\phi(Y_s + rY_{s,t})(Y_{s,t}^{\otimes 2} - \tilde{Y}_{s,t}^{\otimes 2}) + (D^2\phi(Y_s + rY_{s,t}) - D^2\phi(\tilde{Y}_s + r\tilde{Y}_{s,t}))\tilde{Y}_{s,t}^{\otimes 2},$$

and this contributes the last line to (4.17).

The second line in (5.12) equals

$$D\phi(Y_s)R_{s,t}^Y - D\phi(\tilde{Y}_s)R_{s,t}^{\tilde{Y}} \lesssim D\phi(Y_s)(R_{s,t}^Y - R_{s,t}^{\tilde{Y}}) + (D\phi(Y_s) - D\phi(\tilde{Y}_s))R_{s,t}^{\tilde{Y}}$$

which completes the proof. \square

Lemma 4.16 (Stability of rough integration). *Let $\mathbf{X}, \tilde{\mathbf{X}}$ be rough paths, \mathbf{Y} a \mathbf{X} -controlled path, and $\tilde{\mathbf{Y}}$ a $\tilde{\mathbf{X}}$ -controlled path. Define $\mathbf{Z}, \tilde{\mathbf{Z}}$ as in Lemma 4.12. Then,*

$$\begin{aligned} \|\Delta R^Z\|_{r/2} &\lesssim_r \|\Delta Y'\|_{\text{sup}}\|\otimes\|_{r/2} + \|\Delta Y'\|_r\|\otimes\|_{r/2} + \|\Delta R^Y\|_{r/2}\|X\|_r \\ &\quad + \|\tilde{Y}'\|_{\text{sup}}\|\Delta\otimes\|_{r/2} + \|\tilde{Y}'\|_r\|\Delta\otimes\|_{r/2} + \|R^{\tilde{Y}}\|_{r/2}\|\Delta X\|_r. \end{aligned} \quad (4.20)$$

Note that (4.20) with $\tilde{\mathbf{Y}} = 0$ recovers (4.8).

Proof. Let $\Xi_{s,t} := Y_s X_{s,t} + Y'_s \otimes_{s,t} - (\tilde{Y}_s \tilde{X}_{s,t} + \tilde{Y}'_s \tilde{\otimes}_{s,t})$. Then one has,

$$\begin{aligned} |R_{s,t}^{I_{\mathbf{X}}(Y)} - R_{s,t}^{I_{\tilde{\mathbf{X}}}(\tilde{Y})}| &= \left| \int_s^t \mathbf{Y} d\mathbf{X} - Y_s X_{s,t} - \left(\int_s^t \tilde{\mathbf{Y}} d\tilde{\mathbf{X}} - \tilde{Y}_s \tilde{X}_{s,t} \right) \right| \\ &\leq |Y'_s \otimes_{s,t} - \tilde{Y}'_s \tilde{\otimes}_{s,t}| + |I(\Xi)_{s,t} - \Xi_{s,t}|. \end{aligned}$$

We estimate the first term by

$$|\Delta Y'_s \otimes_{s,t}| + |\tilde{Y}'_s \Delta \otimes_{s,t}|.$$

We estimate the last term by the sewing lemma (Theorem 4.6) with (as the calculation (4.6) shows)

$$\delta(\Xi)_{\tau,u,\nu} = R_{\tau,u}^Y \delta X_{u,\nu} + \delta Y'_{\tau,u} \otimes_{u,\nu} - (R_{\tau,u}^{\tilde{Y}} \delta \tilde{X}_{u,\nu} + \delta \tilde{Y}'_{\tau,u} \tilde{\otimes}_{u,\nu}).$$

and

$$|\delta(\Xi)_{\tau,u,\nu}| \leq |\Delta \delta Y'_{\tau,u} \otimes_{u,\nu}| + |\delta \tilde{Y}'_{\tau,u} \Delta \otimes_{u,\nu}| + |\Delta R_{\tau,u}^Y \delta X_{u,\nu}| + |R_{\tau,u}^{\tilde{Y}} \Delta \delta X_{u,\nu}|.$$

\square

Lemma 4.17 (Stability of implicit bound).

$$\|Y - \tilde{Y}\|_r \leq \|\Delta Y'\|_{\text{sup}}\|X\|_r + \|\tilde{Y}'\|_{\text{sup}}\|\Delta X\|_r + \|\Delta R^Y\|_{r/2}.$$

Proof. This follows from writing

$$\begin{aligned} Y_{s,t} - \tilde{Y}_{s,t} &= (Y_s' X_{s,t} + R_{s,t}^Y) - (\tilde{Y}_s' \tilde{X}_{s,t} + R_{s,t}^{\tilde{Y}}) \\ &= (Y_s' - \tilde{Y}_s') X_{s,t} + \tilde{Y}_s' (X_{s,t} - \tilde{X}_{s,t}) + (\Delta R_{s,t}^Y). \quad \square \end{aligned}$$

Lemma 4.18 (Contractivity of the iteration). *For every $A < \infty$ and $\phi \in C_b^{2,1}$, there exists $\epsilon > 0$ such that, for every rough path \mathbf{X} with $\|\mathbf{X}\|_r < \epsilon$ and $\|\otimes\|_{r/2} < \epsilon$, the map (4.13) is strictly contractive on the set*

$$\{Y \in \mathcal{Y}(\mathbf{X}, A) \mid Y_0 = y_0, Y'_0 = \phi(y_0)\} \quad (4.21)$$

with respect to the metric

$$d(\mathbf{Y}, \tilde{\mathbf{Y}}) = \max(\|\Delta R^Y\|_{r/2}, \|\Delta Y'\|_r, 2(\|D\phi\|_{\text{sup}} + A\|D\phi\|_{\text{Lip}})\|\Delta Y\|_r). \quad (4.22)$$

Proof. The implicit constants here are allowed to depend on A in the definition of \mathcal{Y} .

Suppose $\mathbf{Y}, \tilde{\mathbf{Y}} \in \mathcal{Y}(\mathbf{X}, A)$ with $d(\mathbf{Y}, \tilde{\mathbf{Y}}) = \alpha$ and $Y_0 = \tilde{Y}_0 = y_0$. Then,

$$\|\Delta Y\|_{\text{sup}} \leq \|\Delta Y'\|_r \lesssim \alpha.$$

By Lemma 4.15, we obtain

$$\|\phi(Y) - \phi(\tilde{Y})\|_r \leq \|D\phi\|_{\text{sup}}\|\Delta Y\|_r + \|D\phi\|_{\text{Lip}}\|\Delta Y\|_{\text{sup}}\|\tilde{Y}\|_r \leq \alpha/2.$$

$$\begin{aligned} \|\phi(Y)' - \phi(\tilde{Y})'\|_r &\lesssim_{\phi} \|\Delta Y'\|_r + \|Y\|_r \|\Delta Y'\|_{\text{sup}} \\ &\quad + \|\Delta Y\|_{\text{sup}}\|\tilde{Y}'\|_r + \|\Delta Y'\|_r \|\tilde{Y}'\|_{\text{sup}} + \|\Delta Y\|_{\text{sup}}\|\tilde{Y}\|_r \|\tilde{Y}'\|_{\text{sup}} \lesssim \alpha(1 + A), \end{aligned}$$

$$\begin{aligned} \|R^{\phi(Y)} - R^{\phi(\tilde{Y})}\|_{r/2} &\lesssim \|\Delta R^Y\|_{r/2} + \|\Delta Y\|_{\text{sup}}\|R^{\tilde{Y}}\|_{r/2} \\ &\quad + (\|Y\|_r + \|\tilde{Y}\|_r)\|\Delta Y\|_r + \|\Delta Y\|_{\text{sup}}\|\tilde{Y}\|_r^2 \lesssim \alpha(1 + A^2). \end{aligned}$$

Let $\mathbf{Z} = \text{Step}(\mathbf{Y})$, $\tilde{\mathbf{Z}} = \text{Step}(\tilde{\mathbf{Y}})$. Then

$$\|\Delta Z'\|_r = \|\phi(Y) - \phi(\tilde{Y})\|_r \leq \alpha/2.$$

By Lemma 4.16,

$$\|\Delta R^Z\|_{r/2} \lesssim \epsilon(\|\Delta\phi(Y)'\|_{\text{sup}} + \|\Delta\phi(Y)'\|_r + \|\Delta R^{\phi(Y)}\|_{r/2}) \lesssim \alpha\epsilon.$$

By Lemma 4.17, this implies

$$\|\Delta Z\|_r \lesssim \|\Delta Z'\|_{\text{sup}}\|\mathbf{X}\|_r + \|\Delta R^Z\|_{r/2} \leq \|\Delta Z'\|_r \epsilon + C\epsilon\alpha \lesssim \epsilon\alpha.$$

Choosing ϵ small enough, we obtain the claim. \square

Theorem 4.19 (Existence of solutions). *For every $\phi \in C_b^{2,1}$, there exist $A < \infty$ and $\epsilon > 0$ such that, for every rough path \mathbf{X} with $\|\mathbf{X}\|_r < \epsilon$ and $\|\otimes\|_{r/2} < \epsilon$ and every y_0 , there exists a unique \mathbf{X} -controlled path $\mathbf{Y} \in \mathcal{Y}(\mathbf{X}, A)$ such that, for every $t \in [0, T]$, we have*

$$\mathbf{Y}_t = y_0 + \int_0^t \phi(\mathbf{Y}) d\mathbf{X}, \quad Y'_t = \phi(Y_t). \quad (4.23)$$

Proof. Let A, ϵ be as in Lemma 4.18. Define a sequence of controlled paths by

$$(\mathbf{Y}_0)_t = y_0, \quad (\mathbf{Y}_0)'_t = 0, \quad (\mathbf{Y}_{j+1}) = \text{Step}(\mathbf{Y}_j).$$

For $j \geq 1$, the paths (\mathbf{Y}_j) are elements of (4.21). It follows from Lemma 4.18 that this sequence is Cauchy with respect to the metric (4.22), and its limit solves the RDE (4.23). Uniqueness of the solution follows from strict contractivity of (4.13). \square

One can also show that the solution \mathbf{Y} in Theorem 4.19 is unique among all \mathbf{X} -controlled paths (not only those in $\mathcal{Y}(\mathbf{X}, A)$). To this end, we note that, given any two \mathbf{X} -controlled solutions, they are in $\mathcal{Y}(\mathbf{X}, A)$ for some large A . One then has to subdivide the time interval into smaller intervals, on each of which \mathbf{X} has sufficiently small variation norm to apply Lemma 4.18. Large jumps of \mathbf{X} have to be handled separately. We omit the tedious details.

4.3.4 Lipschitz dependence on data

Lemma 4.20 (Local Lipschitz dependence of RDE solution on the data). *For every $\phi \in C_b^{2,1}$ and every $A < \infty$, there exists $\epsilon > 0$ such that if $\mathbf{X}, \tilde{\mathbf{X}}$ are rough paths with norms $< \epsilon$, $Y \in \mathcal{Y}(\mathbf{X}, A)$ is a solution of (4.23) with initial datum y_0 , and $\tilde{Y} \in \mathcal{Y}(\tilde{\mathbf{X}}, A)$ is a solution of (4.23) with initial datum \tilde{y}_0 such that $|y_0 - \tilde{y}_0| \leq \epsilon$ and rough path $\tilde{\mathbf{X}}$, then*

$$\max(\|\Delta R^Y\|_{r/2}, \|\Delta Y'\|_r, \|\Delta Y\|_r) \lesssim_{r, \phi, A} \max(\|X - \tilde{X}\|_r, \|\mathbb{X} - \tilde{\mathbb{X}}\|_{r/2}, |y_0 - \tilde{y}_0|).$$

Proof. Let β be the RHS of the conclusion and

$$\alpha := \max(\|\Delta R^Y\|_{r/2}, \|\Delta Y'\|_r, 2(\|D\phi\|_{\text{sup}} + A\|D\phi\|_{\text{Lip}} + 1)\|\Delta Y\|_r).$$

We may assume $\alpha > \beta$, since otherwise the conclusion already holds. In this case, we have

$$\|\Delta Y\|_{\text{sup}} \leq |y_0 - \tilde{y}_0| + \|\Delta Y\|_r \lesssim \alpha, \quad \|\Delta Y'\|_{\text{sup}} \leq \|\phi\|_{\text{Lip}}\|\Delta Y\|_{\text{sup}} \lesssim \alpha.$$

By Lemma 4.15, we obtain

$$\|\phi(Y)' - \phi(\tilde{Y})'\|_r \lesssim \alpha(1 + A), \quad \|\Delta R^\phi\|_{r/2} \lesssim \alpha(1 + A^2),$$

A simple bound is

$$\|\phi(\tilde{Y})\|_r \leq \|\phi\|_{\text{Lip}}\|\tilde{Y}\|_r \leq A.$$

As in (5.15), we have

$$\|R^{\phi(\tilde{Y})}\|_{r/2} \lesssim A^2.$$

Inserting all these bounds into Lemma 4.16, we obtain

$$\begin{aligned} \|\Delta R^Y\|_{r/2} &\lesssim \epsilon(\|\Delta\phi(Y)'\|_{\text{sup}} + \|\Delta\phi(Y)'\|_r + \|\Delta R^{\phi(Y)}\|_{r/2}) + \beta(\|\phi(\tilde{Y})'\|_{\text{sup}} + \|\phi(\tilde{Y})'\|_r + \|R^{\phi(\tilde{Y})}\|_{r/2}) \\ &\lesssim \epsilon\alpha(1 + A^2) + \beta(1 + A^2). \end{aligned}$$

By Lemma 4.15, we have

$$\begin{aligned} \|\Delta Y'\|_r &= \|\phi(Y) - \phi(\tilde{Y})\|_r \leq \|D\phi\|_{\text{sup}}\|\Delta Y\|_r + \|D\phi\|_{\text{Lip}}\|\Delta Y\|_{\text{sup}}\|\tilde{Y}\|_r \\ &\leq (\|D\phi\|_{\text{sup}} + A\|D\phi\|_{\text{Lip}})\|\Delta Y\|_r + A\|D\phi\|_{\text{Lip}}|y_0 - \tilde{y}_0| \leq \alpha/2 + C\beta. \end{aligned}$$

Moreover, by Lemma 4.17, we have

$$\begin{aligned} \|\Delta Y\|_r &\leq \|\Delta Y'\|_{\text{sup}}\|X\|_r + \|\tilde{Y}'\|_{\text{sup}}\|\Delta X\|_r + \|\Delta R^Y\|_{r/2}. \\ &\leq C\alpha\epsilon + A\beta + C(\epsilon\alpha + \beta). \end{aligned}$$

Inserting these bounds into the definition of α , we obtain

$$\alpha \leq \max(C(\epsilon\alpha + \beta), \alpha/2 + C\beta, C(\epsilon\alpha + \beta)),$$

where C depends on r, ϕ, A . If $C\epsilon \leq 1/2$, then this implies $\alpha \leq C\beta$. \square

Remark. Rough paths were introduced in [Lyo98] and controlled paths in [Gub04]. For a long time, the theory concentrated on Hölder continuous paths; a good exposition of this case is in the book [FH20]. The treatment of V^r paths is adapted from [FZ18].

5 Martingale transforms

The main result of this section, Theorem 5.4, is a bound for discrete time versions of the Itô integral.

We denote ℓ^p norms by

$$\ell_k^p a_k := \left(\sum_{k \in \mathbb{N}} |a_k|^p \right)^{1/p}.$$

In order to simplify notation, we only consider martingales g with $g_0 = 0$.

5.1 Davis decomposition

We will use the following L^q bound for the Davis decomposition (constructed in Lemma 1.9).

Lemma 5.1 (Davis decomposition in L^q). *For every martingale (f_n) with values in a Banach space X , there exists a decomposition $f = f^{\text{pred}} + f^{\text{bv}}$ as a sum of two martingales adapted to the same filtration with $f_0^{\text{pred}} = 0$ such that the differences of f^{pred} have predictable majorants:*

$$Xd f_n^{\text{pred}} \leq 2MXdf_{n-1} \quad (5.1)$$

and f^{bv} has bounded variation, in an integral sense for every $q \in [1, \infty)$:

$$L^q \sum_{k \leq n} Xdf_k^{\text{bv}} \leq (q+1)L^qMXdf_n. \quad (5.2)$$

Proof. Recall from the construction in Lemma 1.9 that

$$df_n^{\text{bv}} = d\tilde{h}_n - \mathbb{E}_{n-1}(d\tilde{h}_n),$$

where \tilde{h}_n is a process such that

$$Xd\tilde{h}_n = MXdf_n - MXdf_{n-1}.$$

Therefore,

$$L^q \sum_{k \leq n} Xdf_k^{\text{bv}} \leq L^q \sum_{k \leq n} Xd\tilde{h}_k + L^q \sum_{k \leq n} \mathbb{E}_{k-1} Xd\tilde{h}_k.$$

Using Lemma 1.15 in the second summand, we obtain the claim. \square

Lemma 5.2. *Let $1 \leq q < \infty$, X be a Banach function space, elements of which are \mathbb{R} -valued maps $x(\cdot)$, and (f_n) a martingale with values in X . Then for f^{pred} given by Lemma 5.1 we have*

$$\| \|Sf^{\text{pred}}\|_X \|_{L^q} \leq (q+2) \| \|Sf\|_X \|_{L^q},$$

where the square function is given by

$$\|Sf\|_X := \|\ell_n^2(df_n(\cdot))\|_X$$

Remark. We will apply this with $X = \ell^r$, i.e. r -summable series, viewed as maps from $\mathbb{N} \rightarrow \mathbb{R}$, with the usual Banach structure.

Proof. Using (5.2) we estimate

$$\begin{aligned} \| \|Sf^{\text{pred}}\|_X \|_{L^q} &\leq \| \|Sf\|_X \|_{L^q} + \| \|Sf^{\text{bv}}\|_X \|_{L^q} \\ &\leq \| \|Sf\|_X \|_{L^q} + \| \sum_n |d_n f^{\text{bv}}| \|_X \|_{L^q} \\ &\leq \| \|Sf\|_X \|_{L^q} + \| \sum_n \|d_n f^{\text{bv}}\|_X \|_{L^q} \\ &\leq \| \|Sf\|_X \|_{L^q} + (q+1) \| \sup_n \|d_n f\|_X \|_{L^q} \\ &\lesssim \| \|Sf\|_X \|_{L^q}. \end{aligned} \quad \square$$

5.2 Vector-valued maximal paraproduct estimate

We call a process $(F_{s,t})_{s \leq t}$ depending on two time variables *adapted* if $F_{s,t}$ is \mathcal{F}_t -measurable for every $s \leq t$.

For an adapted process $(F_{s,t})$ and a martingale (g_n) , we define

$$\Pi(F, g)_{s,t} := \sum_{s < j \leq t} F_{s,j-1} dg_j \quad (5.3)$$

Note that $\Pi(F, g)_{s,\cdot}$ only depends on $(F_{s,\cdot})$.

Proposition 5.3. *Let $0 < q, q_1 \leq \infty$, $1 \leq q_0, r, r_0 < \infty$, $1 \leq r_1 \leq \infty$. Assume $1/q = 1/q_0 + 1/q_1$ and $1/r = 1/r_0 + 1/r_1$. Then, for any martingales $(g_n^{(k)})_n$, any adapted sequences $(F_{s,t}^{(k)})_{s \leq t}$, and any stopping times $\tau'_k \leq \tau_k$ with $k \in \mathbb{Z}$, we have*

$$\left\| \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k)})_{\tau'_k, t}| \right\|_q \leq C_{q_0, q_1, r_0, r_1} \left\| \ell_k^{r_1} \sup_{\tau'_k \leq t < \tau_k} |F_{\tau'_k, t}^{(k)}| \right\|_{q_1} \left\| \ell_k^{r_0} Sg_{\tau'_k, \tau_k}^{(k)} \right\|_{q_0}, \quad (5.4)$$

where $Sg_{s,t} := (\delta(Sg)^2)_{s,t}^{1/2} = (\sum_{j=s+1}^t |dg_j|^2)^{1/2}$.

Proof of Proposition 5.3. We may replace each $g^{(k)}$ by the martingale

$$\bar{g}_n^{(k)} := g_{n \wedge \tau_k}^{(k)} - g_{n \wedge \tau'_k}^{(k)} \quad (5.5)$$

without changing the value of either side of (5.4).

Consider first $q \geq 1$. For each k , the sequence

$$h_t^{(k)} := \begin{cases} 0, & t < \tau'_k, \\ \Pi(F^{(k)}, g^{(k)})_{\tau'_k, t}, & t \geq \tau'_k, \end{cases}$$

is a martingale. We may also assume $F_{\tau'_k, t} = 0$ if $t \notin [\tau'_k, \tau_k)$. By the ℓ^r valued BDG inequality (Corollary 3.2), we can estimate

$$\begin{aligned} LHS (5.4) &\lesssim \left\| \ell_k^r |Sh^{(k)}| \right\|_q \\ &= \left\| \ell_k^r \ell_j^2 |F_{\tau'_k, j-1}^{(k)} dg_j^{(k)}| \right\|_q \\ &\leq \left\| \ell_k^r MF^{(k)} \ell_j^2 |dg_j^{(k)}| \right\|_q \\ &\leq \left\| \ell_k^{r_1} MF^{(k)} \right\|_{q_1} \left\| \ell_k^{r_0} Sg^{(k)} \right\|_{q_0}. \end{aligned}$$

Here and later, we abbreviate $MF^{(k)} := \sup_j |F_{\tau'_k, j}^{(k)}|$.

Consider now $q < 1$. By homogeneity, we may assume

$$\left\| \ell_k^{r_1} MF^{(k)} \right\|_{q_1} = \left\| \ell_k^{r_0} Sg^{(k)} \right\|_{q_0} = 1, \quad (5.6)$$

and we have to show

$$\left\| \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k)})_{\tau'_k, t}| \right\|_q \lesssim 1.$$

We use the Davis decomposition $g = g^{\text{pred}} + g^{\text{bv}}$ (Lemma 5.1 with $X = \ell^{r_0}$). The

contribution of the bounded variation part is estimated as follows:

$$\begin{aligned}
& \|\ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k),bv})_{\tau'_k, t}|\|_q \\
& \leq \|\ell_k^r \sum_j |F_{\tau'_k, j-1}^{(k)}| \cdot |dg_j^{(k),bv}|\|_q \\
& \leq \|\ell_k^{r_1} MF^{(k)}\|_{q_1} \|\ell_k^{r_0} \left(\sum_j |dg_j^{(k),bv}| \right)\|_{q_0} \\
& \leq \|\ell_k^{r_1} MF^{(k)}\|_{q_1} \|\sum_j \ell_k^{r_0} |dg_j^{(k),bv}|\|_{q_0} \\
& \lesssim \|\ell_k^{r_1} MF^{(k)}\|_{q_1} \|\sup_j \ell_k^{r_0} |dg_j^{(k)}|\|_{q_0} \\
& \leq \|\ell_k^{r_1} MF^{(k)}\|_{q_1} \|\ell_k^{r_0} Sg^{(k)}\|_{q_0},
\end{aligned}$$

where we used (5.2) in the penultimate step.

It remains to consider the part g^{pred} with predictable bounds for jumps. We use the layer cake formula in the form

$$\int f^q = \int_0^\infty \mathbb{P}\{f^q > \lambda\} d\lambda = \int_0^\infty \mathbb{P}\{f > \lambda^{1/q}\} d\lambda.$$

By the layer cake formula, we have

$$\begin{aligned}
& \left\| \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k),\text{pred}})_{\tau'_k, t}| \right\|_q^q \\
& = \int_0^\infty \mathbb{P}\left\{ \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k),\text{pred}})_{\tau'_k, t}| > \lambda^{1/q} \right\} d\lambda. \quad (5.7)
\end{aligned}$$

Fix some $\lambda > 0$ and define a stopping time

$$\tau := \inf\left\{ t \mid M(\ell_k^{r_0} dg^{(k)})_t > c\lambda^{1/q_0} \text{ or } \ell_k^{r_0} Sg_t^{(k),\text{pred}} > c\lambda^{1/q_0} \text{ or } \ell_k^{r_1} \sup_{0 < j \leq t} |F_{\tau'_k, j}^{(k)}| > \lambda^{1/q_1} \right\}. \quad (5.8)$$

Define stopped martingales $\tilde{g}_t^{(k)} := g_{t \wedge \tau}^{(k),\text{pred}}$ and adapted processes

$$\tilde{F}_{t', t}^{(k)} := F_{t', t \wedge \tau - 1}^{(k)}.$$

Then, on the set $\{\tau = \infty\}$, we have

$$\Pi(F^{(k)}, g^{(k),\text{pred}})_{\tau'_k, t} = \Pi(\tilde{F}^{(k)}, \tilde{g}^{(k)})_{\tau'_k, t} \quad \text{for all } k, t.$$

Hence,

$$\begin{aligned}
& \left\{ \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(F^{(k)}, g^{(k),\text{pred}})_{\tau'_k, t}| > \lambda^{1/q} \right\} \\
& \subset \left\{ \ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(\tilde{F}^{(k)}, \tilde{g}^{(k)})_{\tau'_k, t}| > \lambda^{1/q} \right\} \\
& \cup \left\{ \ell_k^{r_0} Sg^{(k)} > \lambda^{1/q_0} \right\} \cup \left\{ \ell_k^{r_0} Sg^{(k),\text{pred}} > \lambda^{1/q_0} \right\} \\
& \cup \left\{ \ell_k^{r_1} MF^{(k)} > \lambda^{1/q_1} \right\}
\end{aligned} \quad (5.9)$$

The contributions of the latter three terms to (5.7) are $\lesssim 1$ by (5.6) and Lemma 5.2. It remains to handle the first term.

By construction, we have $\ell_k^{r_1} M\tilde{F}^{(k)} \leq \lambda^{1/q_1}$, and due to (5.1) we also have $\ell_k^{r_0} S\tilde{g}^{(k)} \leq \lambda^{1/q_0}$, provided that the absolute constant c in (5.8) is small enough. Choose an arbitrary

exponent \tilde{q} with $q_0 < \tilde{q} < \infty$. By the already known case of the Proposition with (q_0, q_1) replaced by (\tilde{q}, ∞) , we obtain

$$\begin{aligned}
& \mathbb{P}\{\ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(\tilde{F}^{(k)}, \tilde{g}^{(k)})_{\tau'_k, t}| > \lambda^{1/q}\} \\
& \leq \lambda^{-\tilde{q}/q} \|\ell_k^r \sup_{\tau'_k \leq t \leq \tau_k} |\Pi(\tilde{F}^{(k)}, \tilde{g}^{(k)})_{\tau'_k, t}|\|_{\tilde{q}}^{\tilde{q}} \\
& \lesssim_{\tilde{q}} \lambda^{-\tilde{q}/q} \|\ell_k^{r_1} M\tilde{F}^{(k)}\|_{\infty}^{\tilde{q}} \|\ell_k^{r_0} S\tilde{g}^{(k)}\|_{\tilde{q}}^{\tilde{q}} \\
& \leq \lambda^{-\tilde{q}/q_0} \|\ell_k^{r_0} Sg^{(k), \text{pred}}\| \wedge \lambda^{1/q_0} \|\tilde{q}\|.
\end{aligned} \tag{5.10}$$

This estimate no longer depends on the stopping time τ . Integrating the right-hand side of (5.10) in λ , we obtain

$$\begin{aligned}
\int_0^\infty \lambda^{-\tilde{q}/q_0} \|\ell_k^{r_0} Sg^{(k), \text{pred}}\| \wedge \lambda^{1/q_0} \|\tilde{q}\| d\lambda &= \mathbb{E} \int_0^\infty (\lambda^{-\tilde{q}/q_0} (\ell_k^{r_0} Sg^{(k), \text{pred}})^{\tilde{q}} \wedge 1) d\lambda \\
&\sim \mathbb{E} (\ell_k^{r_0} Sg^{(k), \text{pred}})^{q_0} \\
&\sim 1,
\end{aligned}$$

where we used $\tilde{q} > q_0$, Lemma 5.2 with $X = \ell^{r_0}$, and the assumption (5.6). \square

We will soon need a version of Proposition 5.3 with a supremum in both time arguments of the paraproduct. Nice estimates of such form are only available under some structural assumptions on F . In this course, we only consider $F = \delta f$.

Theorem 5.4. *Let q, q_0, q_1, r, r_1 be as in Proposition 5.3 with $r_0 = 2$, that is,*

$$0 < q, q_1 \leq \infty, 1 \leq q_0, r < \infty, 1 \leq r_1 \leq \infty, 1/q = 1/q_0 + 1/q_1, 1/r = 1/2 + 1/r_1.$$

Let f be an adapted process, g a martingale, and τ an adapted partition. Then, we have

$$\left\| \ell_k^r \sup_{\tau_{k-1} \leq s \leq t \leq \tau_k} |\Pi(\delta f, g)_{s,t}| \right\|_q \lesssim \left\| \ell_k^{r_1} \sup_{\tau_{k-1} \leq t < \tau_k} |\delta f_{\tau_{k-1}, t}| \right\|_{q_1} \|Sg\|_{q_0}. \quad (5.11)$$

Proof. For any $s \leq t \leq u$, the sums (5.3) satisfy the relation

$$\begin{aligned} \delta \Pi(F, g)_{s,t,u} &= \Pi_{s,u}(F, g) - \Pi_{s,t}(F, g) - \Pi_{t,u}(F, g) \\ &= \sum_{s < j \leq u} F_{s,j-1} dg_j - \sum_{s < j \leq t} F_{s,j-1} dg_j - \sum_{t < j \leq u} F_{t,j-1} dg_j \\ &= \sum_{t < j \leq u} (F_{s,j-1} - F_{t,j-1}) dg_j. \end{aligned}$$

In case of $F = \delta f$, the right-hand side becomes

$$\sum_{t < j \leq u} (f_t - f_s) dg_j = (f_t - f_s)(g_u - g_t).$$

Therefore, for any $\tau'_k \leq s \leq t$, we can estimate

$$|\Pi_{s,t}(\delta f, g)| \leq |\Pi_{\tau'_k, t}(\delta f, g)| + |\Pi_{\tau'_k, s}(\delta f, g)| + |\delta f_{\tau'_k, s} \delta g_{s,t}|.$$

In the first two terms, we apply Proposition 5.3 with $\tau'_k = \tau_{k-1}$, $f^{(k)} = f$, $g^{(k)} = g$ for each k . In the last term, by Hölder's inequality, we have

$$\left\| \ell_k^r \sup_{\tau_{k-1} \leq s < t \leq \tau_k} |\delta f_{\tau'_k, s} \delta g_{s,t}| \right\|_q \leq \left\| \ell_k^{r_1} \sup_{\tau_{k-1} \leq s < t \leq \tau_k} |\delta f_{\tau'_k, s}| \right\|_{q_1} \left\| \ell_k^2 \sup_{\tau_{k-1} \leq s < t \leq \tau_k} |\delta g_{s,t}| \right\|_{q_0}.$$

In the former norm, we observe that the dependence on t disappears. In the latter norm, we use the ℓ^2 valued BDG inequality (Corollary 3.2) with the martingales $h^{(k)} = \tau_{k-1} g^{\tau_k}$. \square

5.3 Stopping time construction

In this section, we estimate the r -variation of a two-parameter function by square function-like objects, like we did this for one-parameter functions in the proof of Lépingle's inequality.

For an adapted process $(\Pi_{s,t})_{s \leq t}$, let

$$\Pi_{n''}^* := \sup_{0 \leq n < n' \leq n''} |\Pi_{n,n'}|, \quad \Pi^* := \Pi_\infty^*.$$

Lemma 5.5. *For any discrete time adapted process $(\Pi_{s,t})_{s < t}$, there exist adapted partitions $\tau_j^{(m)}$ such that, for every $0 < \rho < r < \infty$, we have*

$$\sup_{u_0 < \dots < u_{l_{\max}}} \sum_{l=1}^{l_{\max}} |\Pi_{u_{l-1}, u_l}|^r \leq \frac{(\Pi^*)^r}{1 - 2^{-r}} + 2^\rho \sum_{m=0}^{\infty} (2^{-m} \Pi^*)^{r-\rho} \sum_{j=1}^{\infty} \left(\sup_{\tau_{j-1}^{(m)} \leq t < \tau_j^{(m)}} |\Pi_{t, \tau_j^{(m)}}| \right)^\rho. \quad (5.12)$$

Proof of Lemma 5.5. For $m \in \mathbb{N}$, define stopping times

$$\tau_0^{(m)} := 0,$$

and then, for $j \geq 0$, allowing values in $\mathbb{N} \cup \{\infty\}$,

$$\tau_{j+1}^{(m)} := \inf\left\{t > \tau_j^{(m)} \mid \sup_{\tau_j^{(m)} \leq t' < t} |\Pi_{t',t}| > 2^{-m-1}\Pi_t^*\right\}. \quad (5.13)$$

Fix $\omega \in \Omega$ and let $(u_l)_{l=0}^{l_{\max}}$ be a finite strictly increasing sequence. Consider $0 < \rho < r < \infty$ and split

$$\sum_{l=1}^{l_{\max}} |\Pi_{u_{l-1}, u_l}|^r = \sum_{m=0}^{\infty} \sum_{l \in L(m)} |\Pi_{u_{l-1}, u_l}|^r, \quad (5.14)$$

where

$$L(m) := \{l \in \{1, \dots, l_{\max}\} \mid 2^{-m-1}\Pi_{u_l}^* < |\Pi_{u_{l-1}, u_l}| \leq 2^{-m}\Pi_{u_l}^*\}. \quad (5.15)$$

In (5.14), we only omitted vanishing summands, since $|\Pi_{u_{l-1}, u_l}| \leq \Pi_{u_l}^*$. Let also $L'(m) := L(m) \setminus \{\sup L(m)\}$. Using (5.15), we obtain

$$\sum_{l=1}^{l_{\max}} |\Pi_{u_{l-1}, u_l}|^r \leq \sum_{m=0}^{\infty} (2^{-m}\Pi^*)^{r-\rho} \sum_{l \in L'(m)} |\Pi_{u_{l-1}, u_l}|^\rho + \sum_{m=0}^{\infty} (2^{-m}\Pi^*)^r. \quad (5.16)$$

Claim. For every $l \in L(m)$, there exists j s.t. $\tau_j^{(m)} \in (u_{l-1}, u_l]$.

Proof of the claim. Let j be maximal with $\tau_j^{(m)} \leq u_{l-1}$. Since $l \in L(m)$, by definition (5.15), we have

$$|\Pi_{u_{l-1}, u_l}| > 2^{-m-1}\Pi_{u_l}^*.$$

By the definition of stopping times (5.13), we obtain $\tau_{j+1}^{(m)} \leq u_l$. \square

Fix m . For each $l \in L'(m)$, let $j(l)$ be the largest j such that $\tau_j^{(m)} \in (u_{l-1}, u_l]$. Then all $j(l)$ are distinct, and, since $l \neq \max L(m)$, the claim shows that $\tau_{j(l)+1}^{(m)} < \infty$. Furthermore, by (5.15), the monotonicity of $t \mapsto \Pi_t^*$, and the fact that the infimum in the definition (5.13) of stopping times is in fact a minimum unless it is infinite, we have

$$|\Pi_{u_{l-1}, u_l}| \leq 2^{-m}\Pi_{u_l}^* \leq 2^{-m}\Pi_{\tau_{j(l)+1}^{(m)}}^* \leq 2 \sup_{\tau_{j(l)}^{(m)} \leq t' < \tau_{j(l)+1}^{(m)}} |\Pi_{t', \tau_{j(l)+1}^{(m)}}|. \quad (5.17)$$

Since all $j(l)$ are distinct, this implies

$$\sum_{l \in L'(m)} |\Pi_{u_{l-1}, u_l}|^\rho \leq 2^\rho \sum_{j=1}^{\infty} \sup_{\tau_{j-1}^{(m)} \leq t' < \tau_j^{(m)}} |\Pi_{t', \tau_j^{(m)}}|^\rho.$$

Substituting this into (7.5), we conclude the proof of Lemma 5.5. \square

Corollary 5.6. *Let $(\Pi_{s,t})_{s \leq t}$ be an adapted process with $\Pi_{t,t} = 0$ for all t . Then, for every $0 < \rho < r < \infty$ and $q \in (0, \infty]$, we have*

$$\|V^r \Pi\|_{L^q} \lesssim \sup_{\tau} \left\| \left(\sum_{j=1}^{\infty} \left(\sup_{\tau_{j-1} \leq t < t' \leq \tau_j} |\Pi_{t', t}| \right)^\rho \right)^{1/\rho} \right\|_{L^q}, \quad (5.18)$$

where the supremum is taken over all adapted partitions τ .

Proof. By the monotone convergence theorem, we can restrict the times in the definition of V^r to a finite set, and then apply Lemma 5.5.

The term Π^* is of the form on the right-hand side of (5.18) with $\tau_1 = \infty$. Therefore, the claim follows from the triangle inequality in L^q (if $q \geq 1$), q -convexity of L^q (if $q < 1$), and Hölder's inequality. \square

Theorem 5.7. *Let*

$$0 < q, q_1 \leq \infty, 1 \leq q_0 < \infty, 1 \leq r_1 \leq \infty, 1/q = 1/q_0 + 1/q_1, 1/r < 1/2 + 1/r_1.$$

Let f be an adapted process and g a martingale. Then, we have

$$\|V_k^r \Pi(\delta f, g)\|_q \lesssim \|V^{r_1} f\|_{q_1} \|Sg\|_{q_0}. \quad (5.19)$$

Proof. Combine Corollary 5.6 and Theorem 5.4. \square

Remark. This section is mostly copied from [FZ20].

6 Itô integration

In this section, we discuss integration with respect to a càdlàg martingale in continuous time, generalizing the paraproduct Π from the previous section. We consider stochastic processes adapted to some filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_{\geq 0}}$ indexed by positive real times. We begin with some convenient regularity assumptions. These assumptions are not restrictive, in the sense that, for any martingale, one can reparametrize time and change the filtration in such a way that they are satisfied, but we will not discuss this, since our focus is on more quantitative issues.

A filtration $(\mathcal{F}_t)_{t \in \mathbb{R}_{\geq 0}}$ is called *right-continuous* if, for every $t \geq 0$, we have

$$\mathcal{F}_t = \bigcap_{t' > t} \mathcal{F}_{t'}.$$

For a function $f : \mathbb{R}_{\geq 0} \rightarrow E$ with values in a metric space E , the left and right limits at a point t are denoted by

$$f_{t-} := \lim_{s \rightarrow t, s < t} f_s, \quad f_{t+} := \lim_{s \rightarrow t, s > t} f_s,$$

if they exist. A function $f : \mathbb{R}_{\geq 0} \rightarrow E$ is called *càdlàg* (for “continue à droite, limite à gauche”, “right continuous with left limits”; some authors use the English abbreviation “rcll”) if f_{t-} exists for every $t > 0$, and $f_t = f_{t+}$ for every $t \geq 0$. A stochastic process $g : \Omega \times \mathbb{R}_{\geq 0} \rightarrow E$ is called càdlàg if every path $g(\omega, \cdot)$ is a càdlàg function.

Theorem 6.1 (Regularization, see e.g. [Kal21, Theorem 9.28]). *Let $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_{\geq 0}}$ be a right-continuous filtration. If g is a martingale with respect to \mathcal{F} , then there exists a càdlàg martingale \tilde{g} with respect to \mathcal{F} such that, for every $t \geq 0$, we have*

$$g(\cdot, t) = \tilde{g}(\cdot, t) \quad a.e.$$

For an adapted partition π , we write

$$[t, \pi] := \max\{s \in \pi \mid s \leq t\}, \quad 0 \leq t < \infty. \quad (6.1)$$

For a càdlàg adapted process $f = (f_t)_{t \geq 0}$, a càdlàg martingale $g = (g_t)_{t \geq 0}$, and an adapted partition π , we consider the following Riemann–Stieltjes sums:

$$\Pi^\pi(f, g)_{t, t'} := \sum_{t < \pi_j < t'} \delta f_{[t, \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t'}, \quad 0 \leq t \leq t' < \infty. \quad (6.2)$$

In each summand, the integrand f is evaluated at the left endpoint of the interval $[\pi_j, \pi_{j+1}]$ on which we consider the increment of the integrator g (and also at $[t, \pi]$, which is even further to the left). This is the distinguishing feature of the Itô integral that makes it a martingale in the t' variable.

Unlike in Riemann(–Stieltjes) integration with a bounded variation integrator, evaluating t at other points in general produces different results (e.g. Stratonovich integral, where f is averaged over π_j and π_{j+1}).

Most classical treatments involve sums

$$\sum_{\pi_j < t'} f_{\pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t'},$$

which correspond to fixing $t = 0$, but this obscures the observation of the natural regularity of the Itô integral that fits nicely with rough path theory.

For an adapted process f and an adapted partition π , we write

$$f_t^{(\pi)} := f_{[t, \pi]}. \quad (6.3)$$

Proposition 6.2. Let $0 < q_1 \leq \infty$, $1 \leq q_0 < \infty$, and $0 < r, p_1 \leq \infty$. Suppose

$$1/r < 1/p_1 + 1/2, \quad 1/q = 1/q_0 + 1/q_1. \quad (6.4)$$

Let (f_t) be a càdlàg adapted process and (g_t) a càdlàg martingale.

Then, for every adapted partition π , we have the estimate

$$\left\| V^r \Pi^\pi(f, g) \right\|_{L^q} \lesssim \|V^{p_1} f^{(\pi)}\|_{L^{q_1}} \|V^\infty g\|_{L^{q_0}}. \quad (6.5)$$

Proof of Proposition 6.2. Since $\Pi^\pi(F, g)_{t,t'}$ is càdlàg in both t and t' , we have

$$V^r \Pi^\pi(F, g) = \lim_{n \rightarrow \infty} \sup_{l_{\max}, u_0 < \dots < u_{l_{\max}}, u_l \in \pi^{(n)}} \left(\sum_{l=1}^{l_{\max}} |\Pi^\pi(F, g)_{u_{l-1}, u_l}|^r \right)^{1/r},$$

where $\pi^{(n)} = \pi \cup 2^{-n}\mathbb{N}$. By the monotone convergence theorem, it suffices to consider a fixed $\pi^{(n)}$, as long as the bound does not depend on n .

For any adapted partitions $\pi \subseteq \tau$, we have

$$\begin{aligned} \Pi^\pi(f, g)_{t,t'} &= \sum_{k: t < \pi_k < t'} \delta f_{[t, \pi], \pi_k} \delta g_{\pi_k, \pi_{k+1} \wedge t'} \\ &= \sum_{k: t < \pi_k < t'} \delta f_{[t, \pi], \pi_k} \sum_{l: \pi_k \leq \tau_l < \pi_{k+1} \wedge t'} \delta g_{\tau_l, \tau_{l+1} \wedge t'} \\ &= \sum_{k: t < \pi_k < t'} \sum_{l: \pi_k \leq \tau_l < \pi_{k+1} \wedge t'} \delta f_{[t, \pi], [\tau_l, \pi]} \delta g_{\tau_l, \tau_{l+1} \wedge t'} \\ &= \sum_{l: t < \tau_l < t'} \delta f_{[t, \tau], \tau_l}^{(\pi)} \delta g_{\tau_l, \tau_{l+1} \wedge t'} \\ &= \Pi^\tau(f^{(\pi)}, g)_{t,t'}, \end{aligned} \quad (6.6)$$

where $f^{(\pi)}$ is given by (6.3). Define discrete time processes $f_\tau^{(\pi)}, g_\tau$ by

$$(f_\tau^{(\pi)})_j = f_{\tau_j}^{(\pi)}, \quad (g_\tau)_j = g_{\tau_j}.$$

Then, we have

$$\begin{aligned} \Pi^\pi(f, g)_{\tau_j, \tau_{j'}} &= \Pi^\tau(f^{(\pi)}, g)_{\tau_j, \tau_{j'}} \\ &= \sum_{l: \tau_j < \tau_l < \tau_{j'}} \delta f_{[\tau_j, \tau], \tau_l}^{(\pi)} \delta g_{\tau_l, \tau_{l+1} \wedge \tau_{j'}} \\ &= \sum_{l: j < l < j'} \delta f_{\tau_j, \tau_l}^{(\pi)} \delta g_{\tau_l, \tau_{l+1}} \\ &= \Pi(f_\tau^{(\pi)}, g_\tau)_{j, j'}, \end{aligned}$$

where the last line is the discrete time paraproduct defined in (5.3). By Theorem 5.7 and the BDG inequality (Corollary 1.13) for the discrete time martingale g_τ , we obtain

$$\|V^r \Pi(f_\tau^{(\pi)}, g_\tau)\|_q \lesssim \|V^{r_1} f_\tau^{(\pi)}\|_{q_1} \|Sg_\tau\|_{q_0} \lesssim \|V^{r_1} f_\tau^{(\pi)}\|_{q_1} \|V^\infty g_\tau\|_{q_0} \leq \|V^{r_1} f\|_{q_1} \|V^\infty g\|_{q_0}. \quad \square$$

Lemma 6.3. Let $(f_t)_{t \geq 0}$ be a càdlàg adapted process. Suppose that $V^{p_1} f \in L^{q_1}$ for some $p_1, q_1 \in (0, \infty]$. Then, for every $\tilde{p}_1 \in (p_1, \infty) \cup \{\infty\}$, we have

$$\lim_{\pi} \|V^{\tilde{p}_1}(f - f^{(\pi)})\|_{L^{q_1}} = 0.$$

Proof. We have $V^{p_1} f^{(\pi)} \leq V^{p_1} f$ and, by Hölder's inequality,

$$V^{\tilde{p}_1}(f - f^{(\pi)}) \leq V^{p_1}(f - f^{(\pi)})^{1-\theta} V^\infty(f - f^{(\pi)})^\theta$$

with some $\theta \in (0, 1]$, so it suffices to consider $\bar{p}_1 = \infty$.

Let $\epsilon > 0$ and define a sequence of stopping times recursively, starting with $\pi_0 := 0$, by

$$\pi_{j+1} := \inf\{t > \pi_j \mid |\delta f_{\pi_j, t}| \geq \epsilon\}.$$

Since f is càdlàg, the infimum is either $+\infty$ or a minimum, so that this is indeed a stopping time. Also by the càdlàg assumption, this sequence of stopping times is strictly monotonically increasing in the sense that $\pi_j < \infty \implies \pi_j < \pi_{j+1}$. Moreover, if $T := \sup_j \pi_j < \infty$, then the left limit f_{T-} does not exist, contradicting the càdlàg hypothesis. Therefore, $\pi_j \rightarrow \infty$, so that π is an adapted partition.

Then, by (4.18), for any adapted partition $\pi' \supseteq \pi$ and $s \leq t$, we have

$$\begin{aligned} |\delta f_{s,t} - \delta f_{s,t}^{(\pi')}| &\leq |f_t - f_{[t, \pi']}| + |f_s - f_{[s, \pi']}| \\ &\leq 2\epsilon + 2\epsilon. \end{aligned} \quad \square$$

Theorem 6.4 (Itô integral). *In the situation of Proposition 6.2, suppose that the right-hand side of (6.8) is finite. Then*

$$\Pi(f, g) := \lim_{\pi} \Pi^{\pi}(f, g) \quad (6.7)$$

exists in $L^q(\Omega, V^r)$, satisfies the bound

$$\left\| V^r \Pi(f, g) \right\|_{L^q} \lesssim \|V^{p_1} f\|_{L^{q_1}} \|V^{\infty} g\|_{L^{q_0}}, \quad (6.8)$$

and, for any $0 \leq t \leq t' \leq t'' < \infty$, Chen's relation

$$\Pi(f, g)_{t, t''} - \Pi(f, g)_{t, t'} - \Pi(f, g)_{t', t''} = \delta f_{t, t'} \delta g_{t', t''}. \quad (6.9)$$

The limit (6.7) is called the *Itô integral* (with integrand f and integrator g).

Proof of Theorem 6.4. By the Cauchy criterion for net convergence, the existence of the limit (6.7) will follow if we can show that

$$\limsup_{\pi} \sup_{\tau \supseteq \pi} \left\| V^r (\Pi^{\pi}(f, g) - \Pi^{\tau}(f, g)) \right\|_{L^q} = 0. \quad (6.10)$$

To this end, we use that, by (6.6), we have

$$\Pi^{\pi}(f, g) - \Pi^{\tau}(f, g) = \Pi^{\tau}(f^{(\pi)} - f^{(\tau)}, g).$$

Let $\bar{p}_1 \in (p_1, \infty] \cup \{\infty\}$ be such that $1/r < 1/\bar{p}_1 + 1/2$. By Proposition 6.2 with f replaced by $f^{(\pi)} - f^{(\tau)}$, we obtain

$$\begin{aligned} &\left\| V^r \Pi^{\tau}(f^{(\pi)} - f^{(\tau)}, g) \right\|_{L^q} \\ &\lesssim \|V^{\bar{p}_1} (f^{(\pi)} - f^{(\tau)})^{(\tau)}\|_{L^{q_1}} \|V^{\infty} g\|_{L^{q_0}} \end{aligned}$$

This converges to 0 by Lemma 6.3.

In order to show the Chen relation (6.9), we first show that the corresponding relation holds pointwise for the discretized paraproducts Π^{π} . Indeed, by definition (6.2), for $t \leq t' \leq t''$, we have

$$\Pi^{\pi}(f, g)_{t, t''} - \Pi^{\pi}(f, g)_{t, t'} - \Pi^{\pi}(f, g)_{t', t''} \quad (6.11)$$

$$\begin{aligned} &= \sum_{t < \pi_j < t''} \delta f_{[t, \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t''} - \sum_{t < \pi_j < t'} \delta f_{[t, \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t'} \\ &\quad - \sum_{t' < \pi_j < t''} \delta f_{[t', \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t''} \end{aligned}$$

$$\begin{aligned}
&= \sum_{t < \pi_j < t'} f_{[t, \pi], \pi_j} (\delta g_{\pi_j, \pi_{j+1} \wedge t''} - \delta g_{\pi_j, \pi_{j+1} \wedge t'}) + \sum_{t < \pi_j = t' < t''} \delta f_{[t, \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t''} \\
&\quad + \sum_{t' < \pi_j < t''} (\delta f_{[t, \pi], \pi_j} - \delta f_{[t', \pi], \pi_j}) \delta g_{\pi_j, \pi_{j+1} \wedge t''} \\
&= \sum_{t < \pi_j < t'} f_{[t, \pi], \pi_j} \delta g_{\pi_{j+1} \wedge t', \pi_{j+1} \wedge t''} + \sum_{t < \pi_j = t' < t''} \delta f_{[t, \pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t''} \\
&\quad + \delta f_{[t, \pi], [t', \pi]} \sum_{t' < \pi_j < t''} \delta g_{\pi_j, \pi_{j+1} \wedge t''}
\end{aligned}$$

All summands except possibly the one with $\pi_j < t' < \pi_{j+1}$ in the first sum vanish, and it follows that

$$\begin{aligned}
\cdots &= \delta f_{[t, \pi], [t', \pi]} \left(\sum_{t < \pi_j < t' < \pi_{j+1}} \delta g_{\pi_{j+1} \wedge t', \pi_{j+1} \wedge t''} + \sum_{t < \pi_j = t' < t''} \delta g_{\pi_j, \pi_{j+1} \wedge t''} + \sum_{t' < \pi_j < t''} \delta g_{\pi_j, \pi_{j+1} \wedge t''} \right) \\
&= \delta f_{[t, \pi], [t', \pi]} \left(\sum_{t < \pi_j \leq t' < \pi_{j+1} \wedge t''} \delta g_{t', \pi_{j+1} \wedge t''} + \sum_{t' < \pi_j < t''} \delta g_{\pi_j, \pi_{j+1} \wedge t''} \right) = \delta f_{t, t'}^{(\pi)} \delta g_{t', t''}.
\end{aligned}$$

By the already known conclusion (6.7), the process (6.11) converges to the left-hand side of (6.9). By Lemma 6.3, the last expression in the above chain of equalities converges to the right-hand side of (6.9). \square

6.1 Quadratic covariation

Proposition 6.5. *Let $q_0, q_1 \in [1, \infty)$, $1/q = 1/q_0 + 1/q_1$, and $r > 1$. For càdlàg martingales f, g with $V^\infty f \in L^{q_0}$, $V^\infty g \in L^{q_1}$, let*

$$[f, g]_{t, t'}^\pi := \sum_{[t, \pi] \leq \pi_j < [t', \pi]} \delta f_{\pi_j, \pi_{j+1}} \delta g_{\pi_j, \pi_{j+1}}.$$

Then,

$$[f, g]_{t, t'} := \lim_{\pi} [f, g]_{t, t'}^\pi \quad (6.12)$$

exists in $L^q(V^r)$. Moreover, we have the integration by parts formula

$$\delta f_{t, t'} \delta g_{t, t'} = \Pi(f, g)_{t, t'} + \Pi(g, f)_{t, t'} + [f, g]_{t, t'}. \quad (6.13)$$

The process (6.12) is called the *quadratic covariation* of f and g . Here are a few facts about it that easily follow from the definition:

$$\forall t \leq t' \leq t'', \quad [f, g]_{t, t''} = [f, g]_{t, t'} + [f, g]_{t', t''} \quad \text{a.s.}$$

To see this, note that the same identity holds for $[\cdot, \cdot]^\pi$ for any adapted partition π that contains the times t, t', t'' .

$$\forall t \leq t', \quad \mathbb{E}_t[f, f]_{t, t'} = \mathbb{E}_t |\delta f_{t, t'}|^2.$$

To see this, note that the same identity holds for $[\cdot, \cdot]^\pi$ for any adapted partition π that contains the times t, t' , because, by the optimal sampling theorem, $\delta f_{\pi_j, \pi_{j+1}}$ are martingale increments, and therefore orthogonal in L^2 .

Proof of Proposition 6.5. First, we compute

$$\begin{aligned}
& \delta f_{[t,\pi],t'} \delta g_{[t,\pi],t'} - \Pi^\pi(f, g)_{t,t'} - \Pi^\pi(g, f)_{t,t'} \\
&= \left(\sum_{[t,\pi] \leq \pi_k < t'} \delta f_{\pi_k, \pi_{k+1} \wedge t'} \right) \left(\sum_{[t,\pi] \leq \pi_j < t'} \delta g_{\pi_j, \pi_{j+1} \wedge t'} \right) \\
&- \sum_{[t,\pi] < \pi_j < t'} \delta f_{[t,\pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t'} - \sum_{[t,\pi] < \pi_k < t'} \delta g_{[t,\pi], \pi_k} \delta f_{\pi_k, \pi_{k+1} \wedge t'} \\
&= \sum_{[t,\pi] \leq \pi_j < t'} \delta f_{\pi_j, \pi_{j+1} \wedge t'} \delta g_{\pi_j, \pi_{j+1} \wedge t'} \\
&+ \sum_{[t,\pi] \leq \pi_k < \pi_j < t'} \delta f_{\pi_k, \pi_{k+1} \wedge t'} \delta g_{\pi_j, \pi_{j+1} \wedge t'} - \sum_{[t,\pi] < \pi_j < t'} \delta f_{[t,\pi], \pi_j} \delta g_{\pi_j, \pi_{j+1} \wedge t'} \\
&+ \sum_{[t,\pi] \leq \pi_j < \pi_k < t'} \delta f_{\pi_k, \pi_{k+1} \wedge t'} \delta g_{\pi_j, \pi_{j+1} \wedge t'} - \sum_{[t,\pi] < \pi_k < t'} \delta g_{[t,\pi], \pi_k} \delta f_{\pi_k, \pi_{k+1} \wedge t'}.
\end{aligned}$$

Each of the last two lines vanishes identically.

In particular, replacing t' by $[t', \pi]$, we obtain

$$\delta f_{[t,\pi], [t', \pi]} \delta g_{[t,\pi], [t', \pi]} - \Pi^\pi(f, g)_{t, [t', \pi]} - \Pi^\pi(g, f)_{t, [t', \pi]} = \sum_{[t,\pi] \leq \pi_j < [t', \pi]} \delta f_{\pi_j, \pi_{j+1}} \delta g_{\pi_j, \pi_{j+1}}. \quad (6.14)$$

Note that

$$\Pi^\pi(f, g)_{t, [t', \pi]} = \Pi^\pi(f, g)_{t, t'} - \sum_{\pi_j < t' < \pi_{j+1}} \delta f_{[t,\pi], \pi_j} \delta g_{\pi_j, t'}, \quad (6.15)$$

where the sum consists of either 0 or 1 summands. For any increasing sequence $u_0 \leq \dots \leq u_K$, we have In particular, with any $\alpha \in (0, r - 1)$, we have

$$\begin{aligned}
& \sum_{k=0}^{K-1} \left| \sum_{\pi_j < u_{k+1} < \pi_{j+1}} \delta f_{[u_k, \pi], \pi_j} \delta g_{\pi_j, u_{k+1}} \right|^r \\
&= \sum_{k: [u_k, \pi] < [u_{k+1}, \pi]} |\delta f_{[u_k, \pi], [u_{k+1}, \pi]} \delta g_{[u_{k+1}, \pi], u_{k+1}}|^r \\
&\leq \left(\sum_{k: [u_k, \pi] < [u_{k+1}, \pi]} |\delta f_{[u_k, \pi], [u_{k+1}, \pi]}|^{2r} \right)^{1/2} \left(\sum_{k: [u_k, \pi] < [u_{k+1}, \pi]} |\delta g_{[u_{k+1}, \pi], u_{k+1}}|^{2r} \right)^{1/2} \\
&\leq (V^{2r} f)^r \sup_t |\delta g_{[t, \pi], t}|^\alpha (V^{2(r-\alpha)} g)^{r-\alpha}.
\end{aligned}$$

This expression no longer depends on the sequence (u_k) . By Lépingle's and Hölder's inequalities, this is bounded in L^q , and, taking into account Lemma 6.3, this converges to 0 in L^q . Hence, the second term on the right-hand side of (6.15) converges to 0 in $L^q V^r$. Theorem 6.4 and Lépingle's inequality for f now imply

$$\lim_{\pi} \Pi^\pi(f, g)_{t, [t', \pi]} = \Pi(f, g)_{t, t'} \quad \text{in } L^q V^r.$$

Writing (6.14) as

$$[f, g]_{t, t'}^\pi = \delta f_{t, t'}^{(\pi)} \delta g_{t, t'}^{(\pi)} - \Pi^\pi(f, g)_{t, [t', \pi]} - \Pi^\pi(g, f)_{t, [t', \pi]}, \quad (6.16)$$

we see that the first summand on the right-hand side converges to $\delta f \delta g$ in $L^q V^r$ by Lemma 6.3, and the remaining summands to Itô integrals by the above discussion. Hence, the left hand side converges, as was claimed in (6.12). The identity (6.13) is the limit of the above equality. \square

7 A sharp inequality for the square function

The proof of the BDG inequality that we have seen in this notes was quite indirect: we started with some L^2 identities and used the Davis decomposition to lower the L^p exponent. In this section, we will take a look at a more direct method for proving martingale inequalities. In this special case, it will yield an inequality with an optimal constant.

This section follows [Bur02].

Theorem 7.1. *Let $(f_n)_{n \in \mathbb{N}}$ be a real-valued martingale. Then, for every $N \in \mathbb{N}$, we have*

$$\mathbb{E}\left(3|f_0| + \sum_{n=1}^N \frac{|df_n|^2}{f_n^*}\right) \leq \mathbb{E}\left(2f_N^* + \frac{|f_N|^2}{f_N^*}\right). \quad (7.1)$$

We will see later that the inequality (7.1) implies the Davis inequality for the martingale square function with the sharp constant. In applications, one replaces (7.1) by the slightly weaker inequality

$$\mathbb{E}\left(3|f_0| + \sum_{n=1}^N \frac{|df_n|^2}{f_n^*}\right) \leq 3\mathbb{E}f_N^*.$$

However, the peculiar form of the right-hand side of (7.1) permits to show this result by induction on N .

7.1 A Bellman function

The inductive step in the proof of Theorem 7.1 is usually stated as a concavity property of a special function. Functions used in such arguments are called ‘‘Bellman functions’’; many more examples can be found in the books [Osę12; VV20].

Thoroughout this section,

$$\mathcal{D} := \{(x, t, z) \in \mathbb{R} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \mid |x| \leq z\}.$$

We define $U : \mathcal{D} \rightarrow \mathbb{R}$ by

$$U(x, y, m) := y - \frac{|x|^2 + (\gamma - 1)m^2}{m},$$

where $\gamma = 3$. The main feature of this function is the following concavity property.

Proposition 7.2. *For any $x, h \in \mathbb{R}$ and $y, m \in \mathbb{R}_{\geq 0}$ with $|x| \leq m$, we have*

$$U(x + h, y + \frac{|h|^2}{(|x + h| \vee m)}, |x + h| \vee m) \leq U(x, y, m) - \frac{2(x)h}{m}. \quad (7.2)$$

Proof of Theorem 7.1 assuming Proposition 7.2. Using (7.2) with

$$x = f_n, \quad y = \tilde{S}_n := \gamma|f_0| + \sum_{j=1}^n \frac{|df_j|^2}{f_j^*}, \quad m = f_n^*, \quad h = df_{n+1},$$

we obtain

$$U(f_{n+1}, \tilde{S}_{n+1}, f_{n+1}^*) \leq U(f_n, \tilde{S}_n, f_n^*) - \frac{2f_n df_{n+1}}{f_n^*}. \quad (7.3)$$

By conditional independence, we have

$$\mathbb{E}\left(\frac{2f_n df_{n+1}}{f_n^*} \middle| \mathcal{F}_n\right) = \frac{2f_n}{f_n^*} \mathbb{E}(df_{n+1} | \mathcal{F}_n) = 0.$$

Taking expectations, we obtain

$$\mathbb{E}U(f_{n+1}, \tilde{S}_{n+1}, f_{n+1}^*) \leq \mathbb{E}U(f_n, \tilde{S}_n, f_n^*).$$

Iterating this inequality, we obtain

$$\mathbb{E}\left(3|f_0| + \sum_{n=1}^N \frac{|df_n|^2}{(f_n^*)} - \frac{|f_N|^2}{f_N^*} - 2f_N^*\right) = \mathbb{E}U(f_N, \tilde{S}_N, f_N^*) \leq \mathbb{E}U(f_0, \tilde{S}_0, f_0^*) = 0. \quad \square$$

Remark. The above proof in fact shows the pathwise inequality

$$3|f_0| + \sum_{n=1}^N \frac{|df_n|^2}{(f_n^*)} \leq 2f_N^* + \frac{|f_N|^2}{f_N^*} - \sum_{n=1}^N \frac{2(f_n)df_{n+1}}{(f_n^*)}.$$

Proof of Proposition 7.2. If $|x + h| \leq m$, then

$$\begin{aligned} & U\left(x + h, y + \frac{|h|^2}{(|x + h| \vee m)}, |x + h| \vee m\right) \\ &= \left(y + \frac{|h|^2}{m}\right) - \frac{|x + h|^2 + (\gamma - 1)m^2}{m} \\ &= y + \frac{|h|^2}{m} - \frac{|x|^2 + 2xh + |h|^2 + (\gamma - 1)m^2}{m} \\ &= y - \frac{|x|^2 + (\gamma - 1)m^2}{m} - \frac{2xh}{m} \\ &= U(x, y, m) - \frac{2(x)h}{m}. \end{aligned}$$

If $|x + h| > m$, then we need to show

$$\left(y + \frac{|h|^2}{|x + h|}\right) - \frac{|x + h|^2 + (\gamma - 1)|x + h|^2}{|x + h|} \leq y - \frac{|x|^2 + (\gamma - 1)m^2}{m} - \frac{2xh}{m}. \quad (7.4)$$

This is equivalent to

$$\frac{|h|^2 - \gamma|x + h|^2}{|x + h|} \leq \frac{-|x|^2 - (\gamma - 1)m^2}{m} - \frac{2xh}{m}. \quad (7.5)$$

The inequality (7.5) is equivalent to

$$\frac{|h|^2 m}{m^2|x + h|} - \gamma \frac{|x + h|}{m} \leq \frac{-|x + h|^2 + |h|^2}{m^2} - (\gamma - 1).$$

Let $t := |x + h|/m > 1$ and $\tilde{t} := |h|/m$. Note that $|t - \tilde{t}| = ||x + h| - |h||/m \leq |x|/m \leq 1$. With this notation, (7.5) is equivalent to

$$\tilde{t}^2/t - \gamma t \leq -t^2 + \tilde{t}^2 - (\gamma - 1),$$

or

$$\gamma \geq \frac{1}{t-1}(t^2 - 1 - \tilde{t}^2(1 - 1/t)) = (t + 1) - \tilde{t}^2/t.$$

Hence, it suffices to ensure

$$\gamma \geq \sup_{t>1, |t-\tilde{t}|\leq 1} (t + 1) - \tilde{t}^2/t.$$

The supremum in \tilde{t} is assumed for $\tilde{t} = (t - 1)$, so this condition becomes

$$\gamma \geq \sup_{t>1} t + 1 - (t - 1)^2/t = \sup_{t>1} 3 - 1/t = 3. \quad \square$$

7.2 Sharp constant in the Davis inequality for the square function

Proposition 7.3. *Let f be a real-valued martingale. Then*

$$\mathbb{E}Sf \leq \sqrt{3}\mathbb{E}f^*.$$

Proof. By Hölder's inequality and (7.1), we obtain

$$\begin{aligned} \mathbb{E}Sf &\leq \mathbb{E}((f^*)^{1/2}(\sum_{n=1}^N \frac{|df_n|^2}{f_n^*})^{1/2}) \\ &\leq (\mathbb{E}f^*)^{1/2}(\mathbb{E}(\sum_{n=1}^N \frac{|df_n|^2}{f_n^*}))^{1/2} \\ &\leq \sqrt{3}\mathbb{E}f^*. \end{aligned} \quad \square$$

The next result shows that $\sqrt{3}$ is the best possible constant in Proposition 7.3.

Proposition 7.4. *Let $\gamma \geq 0$. Suppose that the inequality*

$$\mathbb{E}Sf \leq \gamma\mathbb{E}f^* \tag{7.6}$$

holds for all real-valued simple martingales f with $f_0 = 0$. Then $\gamma \geq \sqrt{3}$.

Proof. Let $V : \mathcal{D} \rightarrow \mathbb{R}$ be given by

$$V(x, t, z) := \sqrt{t} - \gamma z.$$

Define $U : \mathcal{D} \rightarrow (-\infty, \infty]$ by

$$U(x, t, z) := \sup\{\mathbb{E}V(f_\infty, t + S^2(f), f^* \vee z) \mid f_0 = x\}, \tag{7.7}$$

where the supremum is taken over all simple martingales (that is, martingales f that take only finitely many values). This function will play the role of the “best” Bellman function for the inequality (7.6).

Substituting the constant martingale into the definition (7.7), we see that, for any $(x, t, z) \in D$, we have

$$V(x, t, z) \leq U(x, t, z), \tag{7.8}$$

We claim that, for any simple measurable function $d : \Omega \rightarrow \mathbb{R}$ with $\mathbb{E}d = 0$, we have

$$\mathbb{E}U(x + d, t + |d|^2, |x + d| \vee z) \leq U(x, t, z). \tag{7.9}$$

Proof of (7.9). Fix $x \in \mathbb{R}$ with $|x| \leq z$. Let $d : \Omega \rightarrow \mathbb{R}$ be a simple function with $\mathbb{E}d = 0$ and $\mathbb{P}(d = s_j) = p_j \in (0, 1]$ for $1 \leq j \leq m$, where $\sum_{j=1}^m p_j = 1$. Choose $b_j \in \mathbb{R}$ so that

$$U(x + s_j, t + |s_j|^2, |x + s_j| \vee z) > b_j.$$

Then, by the definition of U , there exists a martingale f^j with $f_0^j = x + s_j$ satisfying

$$\mathbb{E}V(f_\infty^j, t + |s_j|^2 + S^2(f^j), (f^j)^* \vee z) > b_j.$$

Let f be a martingale with $f_1 = x + d$ which continues with the same distribution as f^j (rescaled by p_j in measure) on the set $\{d = s_j\}$. Because $|x| \leq z$, we have $f^* \vee z = (f^j)^* \vee z$ on $\{d = s_j\}$. By (7.7), we have

$$\begin{aligned} U(x, t, z) &\geq \mathbb{E}V(f_\infty, t + S^2(f), f^* \vee z) \\ &= \sum_{j=1}^m \int_{\{d=s_j\}} V(f_\infty, t + |s_j|^2 + |df_2|^2 + \dots, f^* \vee z) d\mathbb{P} \\ &= \sum_{j=1}^m p_j \mathbb{E}V(f_\infty^j, t + |s_j|^2 + S^2(f^j), (f^j)^* \vee z) \\ &\geq \sum_{j=1}^m p_j b_j. \end{aligned}$$

Using the freedom in the choice of b_j 's, this implies (7.9). □

For every $\lambda > 0$, we have

$$V(x, t, z) = \lambda V(x/\lambda, t/\lambda^2, z/\lambda), \quad U(x, t, z) = \lambda U(x/\lambda, t/\lambda^2, z/\lambda).$$

Define $u, v : [-1, 1] \times \mathbb{R}_{\geq 0} \rightarrow (-\infty, \infty]$ by

$$v(x, t) := V(x, t, 1), \quad u(x, t) = U(x, t, 1).$$

Since $t \mapsto V(x, t, z)$ is nondecreasing, the same holds for $t \mapsto u(x, t)$, so left limits exist. We claim that

$$u(1, 1-) \geq u(0, 2-) + u(1, 1-). \quad (7.10)$$

To see this, let $0 < s < 1 < r$. Let d be a random variable such that $\mathbb{P}(d = -1) = r/(r+1)$ and $\mathbb{P}(d = r) = 1/(r+1)$. Then, (7.9) and scaling imply that

$$\begin{aligned} u(1, s) &= U(1, s, 1) \\ &\geq \frac{r}{r+1} U(0, s+1, 1) + \frac{1}{1+r} U(1+r, s+r^2, 1+r) \\ &= \frac{r}{r+1} u(0, s+1) + u(1, (s+r^2)/(1+r)^2) \\ &\geq \frac{r}{r+1} u(0, s+1) + u(1, r^2/(1+r)^2), \end{aligned}$$

where we used monotonicity of u in the second variable in the last step. Both summands on the RHS are increasing in both variables s, r . Taking $r \rightarrow \infty$ and $s \rightarrow 1$, we obtain (7.10).

If f is a simple martingale with $f_0 = 0$, then, by the hypothesis (7.6), we have

$$\mathbb{E}V(f_\infty, 0 + S^2 f, f^* \vee 1) = \mathbb{E}Sf - \gamma \mathbb{E}(f^* \vee 1) \leq \mathbb{E}Sf - \gamma \mathbb{E}f^* \leq 0.$$

By the definition (7.7), this implies that $u(0, 0) = U(0, 0, 1) \leq 0$.

Let d be a random variable with $\mathbb{P}(d = 1) = \mathbb{P}(d = -1) = 1/2$. Using (7.9) and the fact that $u(x, t) = u(-x, t)$, we obtain

$$0 \geq u(0, 0) \geq \frac{1}{2}[u(1, 1) + u(-1, 1)] = u(1, 1) \geq u(1, 1-) \geq v(1, 1-) = 1 - \gamma.$$

This implies that $u(1, 1-)$ is finite. So (7.10) yields $u(0, 2-) \leq 0$. Let d be a random variable with $\mathbb{P}(d = 1) = \mathbb{P}(d = -1) = 1/2$. By (7.9), for every $\epsilon > 0$, we obtain

$$0 \geq u(0, 2-) \geq u(0, 2-\epsilon) \geq \frac{1}{2}[u(1, 3-\epsilon) + u(-1, 3-\epsilon)] = u(1, 3-\epsilon) \geq v(1, 3-\epsilon) = \sqrt{3-\epsilon} - \gamma$$

Therefore, $\gamma \geq \sqrt{3-\epsilon}$. Since $\epsilon > 0$ was arbitrary, this implies $\gamma \geq \sqrt{3}$. \square

References

- [Bur02] D. L. Burkholder. “The best constant in the Davis inequality for the expectation of the martingale square function”. In: *Trans. Amer. Math. Soc.* 354.1 (2002), pp. 91–105. MR: 1859027 (cit. on p. 44).
- [Bur73] D. L. Burkholder. “Distribution function inequalities for martingales”. In: *Ann. Probability* 1 (1973), pp. 19–42. MR: 365692 (cit. on p. 12).
- [FH20] P. K. Friz and M. Hairer. *A course on rough paths*. 2nd ed. Universitext. With an introduction to regularity structures. Springer, Cham, 2020, pp. xvi+346. MR: 4174393 (cit. on p. 31).
- [FZ18] P. K. Friz and H. Zhang. “Differential equations driven by rough paths with jumps”. In: *J. Differential Equations* 264.10 (2018), pp. 6226–6301. arXiv: 1709.05241. MR: 3770049 (cit. on p. 31).
- [FZ20] P. Friz and P. Zorin-Kranich. “Rough semimartingales and p -variation estimates for martingale transforms”. Preprint. 2020. arXiv: 2008.08897 (cit. on p. 38).

- [Gar73] A. M. Garsia. *Martingale inequalities: Seminar notes on recent progress*. Mathematics Lecture Note Series. W. A. Benjamin, Inc., Reading, Mass.-London-Amsterdam, 1973, pp. viii+184. MR: 0448538 (cit. on p. 1).
- [Gra14] L. Grafakos. *Classical Fourier analysis*. Third. Vol. 249. Graduate Texts in Mathematics. Springer, New York, 2014, pp. xviii+638. MR: 3243734 (cit. on p. 19).
- [Gub04] M. Gubinelli. “Controlling rough paths”. In: *J. Funct. Anal.* 216.1 (2004), pp. 86–140. arXiv: math/0306433. MR: 2091358 (cit. on p. 31).
- [Hyt+16] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach spaces. Vol. I: Martingales and Littlewood-Paley theory*. Cham: Springer, 2016, pp. xvi+614. MR: 3617205 (cit. on pp. 1, 20).
- [Kal21] O. Kallenberg. *Foundations of modern probability*. 3rd ed. Vol. 99. Probability Theory and Stochastic Modelling. Springer, Cham, 2021, pp. xii+946. MR: 4226142 (cit. on pp. 1, 39).
- [KZ19] V. Kovač and P. Zorin-Kranich. “Variational estimates for martingale paraproducts”. In: *Electron. Commun. Probab.* 24 (2019), Paper No. 48, 14. arXiv: 1812.09763. MR: 4003122.
- [Lon93] R. L. Long. *Martingale spaces and inequalities*. Peking University Press, Beijing; Friedr. Vieweg & Sohn, Braunschweig, 1993, pp. iv+346. MR: 1224450 (cit. on p. 1).
- [Lyo98] T. J. Lyons. “Differential equations driven by rough signals”. In: *Rev. Mat. Iberoamericana* 14.2 (1998), pp. 215–310. MR: 1654527 (cit. on p. 31).
- [Osę12] A. Osękowski. *Sharp martingale and semimartingale inequalities*. Vol. 72. Instytut Matematyczny Polskiej Akademii Nauk. Monografie Matematyczne (New Series) [Mathematics Institute of the Polish Academy of Sciences. Mathematical Monographs (New Series)]. Birkhäuser/Springer Basel AG, Basel, 2012, pp. xii+462. MR: 2964297 (cit. on pp. 1, 44).
- [Pis16] G. Pisier. *Martingales in Banach spaces*. Vol. 155. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2016, pp. xxviii+561. MR: 3617459 (cit. on p. 1).
- [VV20] V. Vasyunin and A. Volberg. *The Bellman Function Technique in Harmonic Analysis*. Vol. 186. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2020 (cit. on p. 44).