11

Conditional expectations and martingales

Having finished our discussion of the stochastic Cauchy problem with additive noise, we now turn to the more difficult case of equations with multiplicative noise, where the fixed operator $B \in \mathscr{L}(H, E)$ is replaced by an operator-valued function $B : E \to \mathscr{L}(H, E)$:

$$\begin{cases} dU(t) = AU(t)\,dt + B(U(t))\,dt, \quad t \in [0, T], \\ U(0) = u_0. \end{cases}$$

The solutions are then no longer given in closed form by the explicit formula (8.1). Instead, they arise as fixed points of the stochastic integral equation

$$U(t) = S(t)x + \int_0^t S(t-s)B(U(s))\,dW_H(s).$$

The new difficulty here is that integrand is an $\mathscr{L}(H, E)$-valued process depending on U. This requires an extension of the stochastic integration theory of Lecture 6 to this more general situation. As it turns out, in the setting of UMD Banach spaces this can be achieved by a decoupling technique which reduces the construction of the stochastic integral to the one already covered.

In this lecture we introduce the notion of E-valued martingales. They will be used to define UMD Banach spaces as the class of Banach spaces E such that certain a priori estimates hold for E-valued martingales. This may sound rather technical, but the important fact is that Hilbert spaces, L^p-spaces ($1 < p < \infty$), and spaces constructed from these, are UMD spaces. From the point of view of applications, the UMD spaces therefore constitute an important class of spaces.

11.1 Conditional expectations

Throughout this section we fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$. For $1 \leqslant p \leqslant \infty$ we denote by $L^p(\Omega, \mathscr{G})$ the subspace of all $\xi \in L^p(\Omega)$ having a $\mathscr{G}$-measurable representative. With this notation, $L^p(\Omega) = L^p(\Omega, \mathscr{F})$.

Lemma 11.1. *$L^p(\Omega, \mathscr{G})$ is a closed subspace of $L^p(\Omega)$.*

Proof. Suppose that $(\xi_n)_{n=1}^\infty$ is a sequence in $L^p(\Omega, \mathscr{G})$ such that $\lim_{n\to\infty} \xi_n = \xi$ in $L^p(\Omega)$. We may assume that the ξ_n are pointwise defined and $\mathscr{G}$-measurable. After passing to a subsequence (when $1 \leqslant p < \infty$) we may furthermore assume that $\lim_{n\to\infty} \xi_n = \xi$ almost surely. The set C of all $\omega \in \Omega$ where the sequence $(\xi_n(\omega))_{n=1}^\infty$ converges is $\mathscr{G}$-measurable. Put $\widetilde{\xi} := \lim_{n\to\infty} 1_C \xi_n$, where the limit exists pointwise. The random variable $\widetilde{\xi}$ is $\mathscr{G}$-measurable and agrees almost surely with ξ. This shows that ξ defines an element of $L^p(\Omega, \mathscr{G})$. ☐

Our aim is to show that $L^p(\Omega, \mathscr{G})$ is the range of a contractive projection in $L^p(\Omega)$. For $p = 2$ this is clear: we have the orthogonal decomposition

$$L^2(\Omega) = L^2(\Omega, \mathscr{G}) \oplus L^2(\Omega, \mathscr{G})^\perp$$

and the projection we have in mind is the orthogonal projection, denoted by $P_{\mathscr{G}}$, onto $L^2(\Omega, \mathscr{G})$ along this decomposition. Following common usage we write

$$\mathbb{E}(\xi|\mathscr{G}) := P_{\mathscr{G}}\xi, \quad \xi \in L^2(\Omega),$$

and call $\mathbb{E}(\xi|\mathscr{G})$ the *conditional expectation* of ξ with respect to $\mathscr{G}$. Let us emphasise that $\mathbb{E}(\xi|\mathscr{G})$ is defined as an element of $L^2(\Omega, \mathscr{G})$, that is, as an equivalence class of random variables.

Lemma 11.2. *For all $\xi \in L^2(\Omega)$ and $G \in \mathscr{G}$ we have*

$$\int_G \mathbb{E}(\xi|\mathscr{G})\, d\mathbb{P} = \int_G \xi \, d\mathbb{P}.$$

As a consequence, if $\xi \geqslant 0$ almost surely, then $\mathbb{E}(\xi|\mathscr{G}) \geqslant 0$ almost surely.

Proof. By definition we have $\xi - \mathbb{E}(\xi|\mathscr{G}) \perp L^2(\Omega, \mathscr{G})$. If $G \in \mathscr{G}$, then $1_G \in L^2(\Omega, \mathscr{G})$ and therefore

$$\int_\Omega 1_G(\xi - \mathbb{E}(\xi|\mathscr{G}))\, d\mathbb{P} = 0.$$

This gives the desired identity. For the second assertion, choose a $\mathscr{G}$-measurable representative of $g := \mathbb{E}(\xi|\mathscr{G})$ and apply the identity to the $\mathscr{G}$-measurable set $\{g < 0\}$. ☐

Taking $G = \Omega$ we obtain the identity $\mathbb{E}(\mathbb{E}(\xi|\mathscr{G})) = \mathbb{E}\xi$. This will be used in the lemma, which asserts that the mapping $\xi \mapsto \mathbb{E}(\xi|\mathscr{G})$ is L^1-bounded.

Lemma 11.3. *For all $\xi \in L^2(\Omega)$ we have $\mathbb{E}|\mathbb{E}(\xi|\mathscr{G})| \leqslant \mathbb{E}|\xi|$.*

Proof. It suffices to check that $|\mathbb{E}(\xi|\mathscr{G})| \leqslant \mathbb{E}(|\xi|\,|\mathscr{G})$, since then the lemma follows from $\mathbb{E}|\mathbb{E}(\xi|\mathscr{G})| \leqslant \mathbb{E}\mathbb{E}(|\xi|\,|\mathscr{G}) = \mathbb{E}|\xi|$. Splitting ξ into positive and negative parts, almost surely we have

$$\begin{aligned}|\mathbb{E}(\xi|\mathscr{G})| &= |\mathbb{E}(\xi^+|\mathscr{G}) - \mathbb{E}(\xi^-|\mathscr{G})| \\ &\leqslant |\mathbb{E}(\xi^+|\mathscr{G})| + |\mathbb{E}(\xi^-|\mathscr{G})| = \mathbb{E}(\xi^+|\mathscr{G}) + \mathbb{E}(\xi^-|\mathscr{G}) = \mathbb{E}(|\xi|\,|\mathscr{G}). \quad \square\end{aligned}$$

Since $L^2(\Omega)$ is dense in $L^1(\Omega)$ this lemma shows that the conditional expectation operator has a unique extension to a contractive projection on $L^1(\Omega)$, which we also denote by $\mathbb{E}(\cdot|\mathscr{G})$. This projection is again positive in the sense that it maps positive random variables to positive random variables; this follows from Lemma 11.2 by approximation.

Lemma 11.4 (Conditional Jensen inequality). *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex, then for all $\xi \in L^1(\Omega)$ such that $\phi \circ \xi \in L^1(\Omega)$ we have, almost surely,*

$$\phi \circ \mathbb{E}(\xi|\mathscr{G}) \leqslant \mathbb{E}(\phi \circ \xi|\mathscr{G}).$$

Proof. If $a, b \in \mathbb{R}$ are such that $at + b \leqslant \phi(t)$ for all $t \in \mathbb{R}$, then the positivity of the conditional expectation operator gives

$$a\mathbb{E}(\xi|\mathscr{G}) + b = \mathbb{E}(a\xi + b|\mathscr{G}) \leqslant \mathbb{E}(\phi \circ \xi|\mathscr{G})$$

almost surely. Since ϕ is convex we can find real sequences $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ such that $\phi(t) = \sup_{n\geqslant 1}(a_n t + b_n)$ for all $t \in \mathbb{R}$; we leave the proof of this fact as an exercise. Hence almost surely,

$$\phi \circ \mathbb{E}(\xi|\mathscr{G}) = \sup_{n\geqslant 1} a_n\mathbb{E}(\xi|\mathscr{G}) + b_n \leqslant \mathbb{E}(\phi \circ \xi|\mathscr{G}). \quad \square$$

Theorem 11.5 (L^p-contractivity). *For all $1 \leqslant p \leqslant \infty$ the conditional expectation operator extends to a contractive positive projection on $L^p(\Omega)$ with range $L^p(\Omega, \mathscr{G})$. For $\xi \in L^p(\Omega)$, the random variable $\mathbb{E}(\xi|\mathscr{G})$ is the unique element of $L^p(\Omega, \mathscr{G})$ with the property that for all $G \in \mathscr{G}$,*

$$\int_G \mathbb{E}(\xi|\mathscr{G})\, d\mathbb{P} = \int_G \xi\, d\mathbb{P}. \tag{11.1}$$

Proof. For $1 \leqslant p < \infty$ the L^p-contractivity follows from Lemma 11.4 applied to the convex function $\phi(t) = |t|^p$. For $p = \infty$ we argue as follows. If $\xi \in L^\infty(\Omega)$, then $0 \leqslant |\xi| \leqslant \|\xi\|_\infty 1_\Omega$ and therefore $0 \leqslant \mathbb{E}(|\xi|\,|\mathscr{G}) \leqslant \|\xi\|_\infty 1_\Omega$ almost surely. Hence, $\mathbb{E}(|\xi|\,|\mathscr{G}) \in L^\infty(\Omega)$ and $\|\mathbb{E}(|\xi|\,|\mathscr{G})\|_\infty \leqslant \|\xi\|_\infty$.

For $2 \leqslant p \leqslant \infty$, (11.1) follows from Lemma 11.2. For $\xi \in L^p(\Omega)$ with $1 \leqslant p < 2$ we choose a sequence $(\xi_n)_{n=1}^\infty$ in $L^2(\Omega)$ such that $\lim_{n\to\infty} \xi_n = \xi$ in $L^p(\Omega)$. Then $\lim_{n\to\infty} \mathbb{E}(\xi_n|\mathscr{G}) = \mathbb{E}(\xi|\mathscr{G})$ in $L^p(\Omega)$ and therefore, for any $G \in \mathscr{G}$,

$$\int_G \mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P} = \lim_{n\to\infty}\int_G \mathbb{E}(\xi_n|\mathscr{G})\,d\mathbb{P} = \lim_{n\to\infty}\int_G \xi_n\,d\mathbb{P} = \int_G \xi\,d\mathbb{P}.$$

If $\eta \in L^p(\Omega,\mathscr{G})$ satisfies $\int_G \eta\,d\mathbb{P} = \int_G \xi\,d\mathbb{P}$ for all $G \in \mathscr{G}$, then $\int_G \eta\,d\mathbb{P} = \int_G \mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P}$ for all $G \in \mathscr{G}$. Since both η and $\mathbb{E}(\xi|\mathscr{G})$ are $\mathscr{G}$-measurable, as in the proof of the second part of Lemma 11.2 this implies that $\eta = \mathbb{E}(\xi|\mathscr{G})$ almost surely.

In particular, $\mathbb{E}(\mathbb{E}(\xi|\mathscr{G})|\mathscr{G}) = \mathbb{E}(\xi|\mathscr{G})$ for all $\xi \in L^p(\Omega)$ and $\mathbb{E}(\xi|\mathscr{G}) = \xi$ for all $\xi \in L^p(\Omega,\mathscr{G})$. This shows that $\mathbb{E}(\cdot|\mathscr{G})$ is a projection onto $L^p(\Omega,\mathscr{G})$. □

The next two results develop some properties of conditional expectations.

Proposition 11.6.

(1) *If $\xi \in L^1(\Omega)$ and $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$, then almost surely*

$$\mathbb{E}(\mathbb{E}(\xi|\mathscr{G})|\mathscr{H}) = \mathbb{E}(\xi|\mathscr{H}).$$

(2) *If $\xi \in L^1(\Omega)$ is independent of $\mathscr{G}$ (that is, ξ is independent of 1_G for all $G \in \mathscr{G}$), then almost surely*

$$\mathbb{E}(\xi|\mathscr{G}) = \mathbb{E}\xi.$$

(3) *If $\xi \in L^p(\Omega)$ and $\eta \in L^q(\Omega,\mathscr{G})$ with $1 \leqslant p,q \leqslant \infty$, $\frac{1}{p}+\frac{1}{q}=1$, then almost surely*

$$\mathbb{E}(\eta\xi|\mathscr{G}) = \eta\mathbb{E}(\xi|\mathscr{G}).$$

Proof. (1): For all $H \in \mathscr{H}$ we have $\int_H \mathbb{E}(\mathbb{E}(\xi|\mathscr{G})|\mathscr{H})\,d\mathbb{P} = \int_H \mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P} = \int_H \xi\,d\mathbb{P}$ by Theorem 11.5, first applied to $\mathscr{H}$ and then to $\mathscr{G}$ (observe that $H \in \mathscr{G}$). Now the result follows from the uniqueness part of the theorem.

(2): For all $G \in \mathscr{G}$ we have $\int_G \mathbb{E}\xi\,d\mathbb{P} = \mathbb{E}1_G\mathbb{E}\xi = \mathbb{E}1_G\xi = \int_G \xi\,d\mathbb{P}$, and the result follows from the uniqueness part of Theorem 11.5.

(3): For all $G, G' \in \mathscr{G}$ we have $\int_G 1_{G'}\mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P} = \int_{G\cap G'} \mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P} = \int_{G\cap G'} \xi\,d\mathbb{P} = \int_G \xi 1_{G'}\,d\mathbb{P}$. Hence $\mathbb{E}(\xi 1_{G'}|\mathscr{G}) = 1_{G'}\mathbb{E}(\xi|\mathscr{G})$ by the uniqueness part of Theorem 11.5. By linearity, this gives the result for simple functions η, and the general case follows by approximation. □

If $\mathscr{C}$ is a collection of subsets of Ω, then $\sigma(\mathscr{C})$ denotes the σ-algebra generated by $\mathscr{C}$, that is, the smallest σ-algebra in Ω which contains all sets of $\mathscr{C}$. In this context we shall use self-explanatory notations such as $\mathbb{E}(\xi|\mathscr{C}) := \mathbb{E}(\xi|\sigma(\mathscr{C}))$ and $\mathbb{E}(\xi|\mathscr{C}_1,\mathscr{C}_2) := \mathbb{E}(\xi|\sigma(\mathscr{C}_1 \cup \mathscr{C}_2))$.

If $\eta : \Omega \to E$ is a random variable, then $\sigma(\eta)$ denotes the σ-algebra $\{\eta^{-1}(B) : B \in \mathscr{B}(E)\}$. This is the smallest σ-algebra in Ω with respect to which η is Borel measurable. Again, notations such as $\mathbb{E}(\xi|\eta) := \mathbb{E}(\xi|\sigma(\eta))$ and $\mathbb{E}(\xi|\eta_1,\eta_2) := \mathbb{E}(\xi|\sigma(\eta_1,\eta_2))$ are self-explanatory.

Proposition 11.7. *Let $\mathscr{G}$ and $\mathscr{H}$ be sub-σ-algebras of $\mathscr{F}$, let $\xi \in L^1(\Omega)$, and suppose that $\mathscr{H}$ is independent of $\sigma(\xi,\mathscr{G})$. Then, almost surely,*

$$\mathbb{E}(\xi|\mathscr{G},\mathscr{H}) = \mathbb{E}(\xi|\mathscr{G}).$$

Proof. First we claim that $\sigma(\mathscr{G},\mathscr{H})$ is generated by the collection $\mathscr{C}$ of all sets of the form $G\cap H$ with $G\in\mathscr{G}$ and $H\in\mathscr{H}$. Indeed, from $G=G\cap\Omega\in\mathscr{C}$ and $H=\Omega\cap H\in\mathscr{C}$ we see that $\mathscr{C}$ contains both $\mathscr{G}$ and $\mathscr{H}$.

Next, from $(G_1\cap H_1)\cap(G_2\cap H_2)=(G_1\cap G_2)\cap(H_1\cap H_2)$ it follows that $\mathscr{C}$ is closed under taking finite intersections. This being said, the strategy is to apply Dynkin's lemma. By considering positive and negative parts separately we may assume that $\xi\geqslant 0$ almost surely. Also we may assume that $\mathbb{E}\xi>0$, since otherwise there is nothing to prove.

For $G\cap H\in\mathscr{C}$ we have

$$\begin{aligned}\int_{G\cap H}\mathbb{E}(\xi|\mathscr{G},\mathscr{H})\,d\mathbb{P}=\int_{G\cap H}\xi\,d\mathbb{P}&=\mathbb{E}(1_G1_H\xi)\\ &\overset{\text{(i)}}{=}\mathbb{E}1_H\mathbb{E}(1_G\xi)=\mathbb{E}1_H\mathbb{E}(1_G\mathbb{E}(\xi|\mathscr{G}))\\ &\overset{\text{(ii)}}{=}\mathbb{E}(1_G1_H\mathbb{E}(\xi|\mathscr{G}))=\int_{G\cap H}\mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P}.\end{aligned}$$

In (i) and(ii) we used the independence of $\mathscr{H}$ and $\sigma(\xi,\mathscr{G})$. By Dynkin's lemma, applied to the probability measures $\mu(C):=\frac{1}{\mathbb{E}\xi}\int_C\mathbb{E}(\xi|\mathscr{G},\mathscr{H})\,d\mathbb{P}$ and $\nu(C):=\frac{1}{\mathbb{E}\xi}\int_C\mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P}$ it follows that $\mu=\nu$ on $\sigma(\mathscr{C})=\sigma(\mathscr{G},\mathscr{H})$. This means that

$$\int_C\mathbb{E}(\xi|\mathscr{G},\mathscr{H})\,d\mathbb{P}=\int_C\mathbb{E}(\xi|\mathscr{G})\,d\mathbb{P}\quad\forall C\in\sigma(\mathscr{G},\mathscr{H})$$

and the result follows. □

11.2 Vector-valued conditional expectations

Our next aim is to extend the conditional expectation operators from $L^p(\Omega)$ to $L^p(\Omega;E)$, where E is an arbitrary Banach space.

Let us fix $1\leqslant p<\infty$ for the moment and let $(A,\mathscr{A},\mu)$ be an arbitrary σ-finite measure space. For a Banach space E we denote by $L^p(A)\otimes E$ the linear span of all functions of the form $f\otimes x$ with $f\in L^p(A)$ and $x\in E$.

Lemma 11.8. *$L^p(A)\otimes E$ is dense in $L^p(A;E)$.*

Proof. It has been observed in Lecture 1 that the μ-simple functions are dense in $L^p(A;E)$. Clearly these belong to $L^p(A)\otimes E$. □

Suppose next that a bounded linear operator T on $L^p(A)$ is given. We may define a linear operator $T\otimes I$ on $L^p(A)\otimes E$ by the formula

$$(T\otimes I)(f\otimes x):=Tf\otimes x.$$

We leave it to the reader to check that the resulting linear operator on $L^p(A)\otimes E$ is well-defined. In view of Lemma 11.8 one may now ask whether $T\otimes I$ extends to a bounded operator on $L^p(A;E)$. Unfortunately, without additional assumptions this is generally not the case. For *positive* operators T on $L^p(A)$ we have the following result.

Proposition 11.9. *If T is a positive operator on $L^p(A)$, then $T \otimes I$ extends uniquely to a bounded operator on $L^p(A;E)$ and we have*

$$\|T \otimes I\| = \|T\|.$$

Proof. Let $g \in L^p(A) \otimes E$ be a μ-simple function, say $g = \sum_{n=1}^N 1_{A_n} \otimes x_n$ with the sets $A_n \in \mathscr{A}$ mutually disjoint. Then from the positivity of T we have $|T1_A| = T1_{A_n}$ and we obtain the estimates

$$\begin{aligned}
\Big\|(T \otimes I) \sum_{n=1}^N 1_{A_n} \otimes x_n\Big\|_{L^p(A;E)}^p &= \int_A \Big\| \sum_{n=1}^N T1_{A_n} \otimes x_n \Big\|^p \, d\mu \\
&\leqslant \int_A \Big| \sum_{n=1}^N |T1_{A_n}| \|x_n\| \Big|^p \, d\mu \\
&= \int_A \Big| T \sum_{n=1}^N 1_{A_n} \|x_n\| \Big|^p \, d\mu \\
&\leqslant \|T\|^p \int_A \Big| \sum_{n=1}^N 1_{A_n} \|x_n\| \Big|^p \, d\mu \\
&= \|T\|^p \Big\| \sum_{n=1}^N 1_{A_n} \otimes x_n \Big\|_{L^p(A;E)}^p.
\end{aligned}$$

Since the μ-simple functions are dense in $L^p(A;E)$, this proves that $T \otimes I$ has a unique extension to a bounded operator on $L^p(A;E)$ of norm $\|T \otimes I\| \leqslant \|T\|$. Equality $\|T \otimes I\| = \|T\|$ is obtained by considering functions of the form $f \otimes x$ with $f \in L^p(A)$ and $x \in E$ of norm one. □

Returning to conditional expectations we obtain the following result:

Theorem 11.10. *For $1 \leqslant p \leqslant \infty$ the operator $\mathbb{E}(\cdot|\mathscr{G}) \otimes I$ extends uniquely to a contractive projection on $L^p(\Omega;E)$ with range $L^p(\Omega,\mathscr{G};E)$. For all $X \in L^p(\Omega;E)$, the random variable*

$$\mathbb{E}(X|\mathscr{G}) := (\mathbb{E}(\cdot|\mathscr{G}) \otimes I)X$$

is the unique element of $L^p(\Omega,\mathscr{G};E)$ with the property that for all $G \in \mathscr{G}$,

$$\int_G \mathbb{E}(X|\mathscr{G}) \, d\mathbb{P} = \int_G X \, d\mathbb{P}.$$

Proof. For $1 \leqslant p < \infty$ the L^p-contractivity follows from Proposition 11.9.

Before continuing with the case $p = \infty$, let us note that for a simple random variable of the form $X = \sum_{n=1}^N 1_{A_n} \otimes x_n$ with disjoint sets $A_n \in \mathscr{F}$ we have

$$\|\mathbb{E}(X|\mathscr{G})\| = \Big\| \sum_{n=1}^{N} \mathbb{E}(1_{A_n}|\mathscr{G}) \otimes x_n \Big\| \leqslant \sum_{n=1}^{N} \mathbb{E}(1_{A_n}|\mathscr{G}) \|x_n\| = \mathbb{E}(\|X\|\,|\mathscr{G}).$$

By a density argument, this inequality extends to arbitrary random variables $X \in L^1(\Omega; E)$. By inclusion, this implies the corresponding inequality for random variables $X \in L^p(\Omega; E)$, $1 \leqslant p \leqslant \infty$.

Next let $X \in L^\infty(\Omega; E)$. Then $\|X\| \in L^\infty(\Omega)$, and by the inequality which has just been proved together with the contractivity of the conditional expectation in $L^\infty(\Omega)$ we obtain

$$\|\mathbb{E}(X|\mathscr{G})\|_{L^\infty(\Omega;E)} \leqslant \|\mathbb{E}(\|X\|\,|\mathscr{G})\|_{L^\infty(\Omega)} \leqslant \big\|\|X\|\big\|_{L^\infty(\Omega)} = \|X\|_{L^\infty(\Omega;E)}.$$

This proves that the conditional expectation is a contraction in $L^\infty(\Omega; E)$.

For simple random variables X, the identity $\int_G \mathbb{E}(X|\mathscr{G})\, d\mathbb{P} = \int_G X\, d\mathbb{P}$ follows from the corresponding assertion in the scalar case. By density, the identity extends to random variables $X \in L^1(\Omega; E)$, and hence for $X \in L^p(\Omega; E)$, $1 \leqslant p \leqslant \infty$. The uniqueness assertion follows from the scalar case via Corollary 1.14. $\square$

We leave it to the reader to check that Propostions 11.6 and 11.7 extend to the vector-valued setting and finish this section with two important examples. We have already encountered the first example in the proof of Theorem 6.17.

Example 11.11 (Averaging). Consider a decomposition $(0,1) = \bigcup_{n=1}^N I_n$, where the I_n are disjoint intervals with Lebesgue measure $|I_n| > 0$. Let $\mathscr{F}$ be the Borel σ-algebra of $(0,1)$ and let $\mathscr{G}$ be the σ-algebra generated by the intervals I_n. Let E be a Banach space. Then for all $f \in L^1(0,1;E)$ we have

$$\mathbb{E}(f|\mathscr{G}) = \sum_{n=1}^{N} c_n 1_{I_n} \quad \text{with} \quad c_n = \frac{1}{|I_n|} \int_{I_n} f(t)\, dt.$$

This is verified by checking the condition of Theorem 11.5.

Example 11.12 (Sums of independent random variables). Let $(\xi_n)_{n=1}^\infty$ be a sequence of independent integrable E-valued random variables. For each $n \geqslant 1$ let $\mathscr{F}_n := \sigma(\xi_1, \dots, \xi_n)$ and put $S_n = \xi_1 + \dots + \xi_n$. Then for all $N \geqslant n \geqslant 1$ we have

$$\mathbb{E}(S_N|\mathscr{F}_n) = S_n + \mathbb{E}(\xi_{n+1}) + \dots + \mathbb{E}(\xi_N).$$

This follows from Proposition 11.6 (1), (2). In particular, if the ξ_n are centred,

$$\mathbb{E}(S_N|\mathscr{F}_n) = S_n.$$

11.3 Martingales

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(I, \leqslant)$ a partially ordered set, that is, a set I with a relation $\leqslant$ which satisfies the following properties:

(1) $i \leqslant i$ for all $i \in I$;
(2) $i \leqslant j$ and $j \leqslant i$ imply $i = j$;
(3) $i \leqslant j$ and $j \leqslant k$ imply $i \leqslant k$.

Definition 11.13. *Let I be a partially ordered set. A* filtration *with index set I is a family $(\mathscr{F}_i)_{i\in I}$ of sub-σ-algebras of $\mathscr{F}$ such that $\mathscr{F}_i \subseteq \mathscr{F}_j$ whenever $i \leqslant j$ in I. A family $(X_i)_{i\in I}$ of E-valued random variables is* adapted *to the filtration $(\mathscr{F}_i)_{i\in I}$ if each X_i is strongly $\mathscr{F}_i$-measurable.*

In this definition the random variables are defined pointwise. The definitions carry over to equivalence classes modulo null sets, provided one replaces 'is strongly $\mathscr{F}_i$-measurable' with 'has a strongly $\mathscr{F}_i$-measurable representative' in the definition of adaptedness.

Every family $X = (X_i)_{i\in I}$ is adapted to the filtration $(\mathscr{F}_i^X)_{i\in I}$, where $\mathscr{F}_i^X := \sigma(X_j : \ j \leqslant i)$. We call this filtration the *filtration generated by X*.

Definition 11.14. *A family $(M_i)_{i\in I}$ of integrable E-valued random variables is an E-valued* martingale *with respect to a filtration $(\mathscr{F}_i)_{i\in I}$ if it is adapted with respect to $(\mathscr{F}_i)_{i\in I}$ and*

$$\mathbb{E}(M_j|\mathscr{F}_i) = M_i$$

*almost surely whenever $i \leqslant j$ in I. If in addition $\mathbb{E}\|M_i\|^p < \infty$ for all $i \in I$, then we call $(M_i)_{i\in I}$ an E-valued L^p-*martingale.

Example 11.15. Let $X \in L^1(\Omega; E)$ be given. For any filtration $(\mathscr{F}_i)_{i\in I}$, the family $(X_i)_{i\in I}$ defined by

$$X_i := \mathbb{E}(X|\mathscr{F}_i)$$

is a martingale with respect to $(\mathscr{F}_i)_{i\in I}$; this follows from Proposition 11.6 (1).

In most examples, I is a finite or infinite interval in $\mathbb{Z}$ or $\mathbb{R}$. In these cases one speaks of *discrete time martingales* and *continuous time martingales*. Here are two examples.

Example 11.16 (Sums of independent random variables). If $X = (X_n)_{n=1}^\infty$ is a sequence of independent integrable E-valued random variables satisfying $\mathbb{E}X_n = 0$ for all $n \geqslant 1$, then the partial sum sequence $(S_n)_{n=1}^\infty$ is a martingale with respect to the filtration $(\mathscr{F}_n^X)_{n=1}^\infty$, where

$$\mathscr{F}_n^X := \sigma(X_1, \ldots, X_n).$$

This is immediate from Example 11.12.

Example 11.17 (Brownian motion). Every Brownian motion $(W(t))_{t\in[0,T]}$ is a martingale with respect to the filtration $(\mathscr{F}_t^W)_{t\in[0,T]}$ defined by

$$\mathscr{F}_t^W := \sigma(W(s) : \ 0 \leqslant s \leqslant t).$$

Adaptedness and integrability being clear, it remains to show that

$$\mathbb{E}(W(t)|\mathscr{F}_s^W) = W(s)$$

almost surely for all $0 \leqslant s \leqslant t \leqslant T$. Writing $W(t) = W(s) + (W(t) - W(s))$, and noting that $W(s)$ is $\mathscr{F}_s^W$-measurable and $W(t) - W(s)$ is independent of $\mathscr{F}_s^W$ (this will be proved in a moment), with Proposition 11.6 (2), we obtain

$$\mathbb{E}(W(t)|\mathscr{F}_s^W) = W(s) + \mathbb{E}(W(t) - W(s)) = W(s)$$

almost surely.

That $W(t) - W(s)$ is independent of $\mathscr{F}_s^W$ is a consequence of the following general observation:

Lemma 11.18. *A random variable X is independent of the family $(Y_j)_{j\in J}$ if and only if X is independent of the σ-algebra generated by $(Y_j)_{j\in J}$.*

Proof. Let X take its values in E and each Y_j in E_j.

We begin with the 'only if' part. Suppose that X is independent of $(Y_j)_{j\in J}$. By definition this means that X is independent of $(Y_{j_1}, \dots, Y_{j_N})$ for all $j_1, \dots, j_N \in J$. In particular,

$$\mathbb{P}(\{X \in B\} \cap C) = \mathbb{P}\{X \in B\}\mathbb{P}(C) \tag{11.2}$$

for all sets $C = \{Y_{j_1} \in B_1, \dots, Y_{j_N} \in B_N\}$. The collection of all such sets C, which we shall denote by $\mathscr{C}$, is closed under taking finite intersections and generates $\sigma((Y_j)_{j\in J})$. We must show that (11.2) holds for all $C \in \sigma((Y_j)_{j\in J})$. Fix $B \in \mathscr{B}(E)$ and assume without loss of generality that $\mathbb{P}\{X \in B\} > 0$. Consider the probability measure $\mathbb{P}_B$ on $(\Omega, \mathscr{F})$ defined by

$$\mathbb{P}_B(F) := \frac{\mathbb{P}(\{X \in B\} \cap F)}{\mathbb{P}\{X \in B\}}.$$

The measures $\mathbb{P}_B$ and $\mathbb{P}$ coincide on $\mathscr{C}$, and therefore they coincide on $\sigma(\mathscr{C}) = \sigma((Y_j)_{j\in J})$ by Dynkin's lemma.

To prove the 'if' part it suffices to observe that the sets $\{(Y_{j_1}, \dots, Y_{j_N}) \in B\})$ with $B \in \mathscr{B}(E_{j_1} \times \dots \times E_{j_N})$ belong to $\sigma((Y_j)_{j\in J})$. □

More generally, this argument can be used to show that two families $(X_i)_{i\in I}$ and $(Y_j)_{j\in J}$ are independent of each other if and only if $\sigma((X_i)_{i\in I})$ and $\sigma((Y_j)_{j\in J})$ are independent.

Our final example will be used in the next lecture.

Example 11.19 (Martingale transforms). A real-valued sequence $v = (v_n)_{n=1}^N$ is said to be *predictable* with respect to a filtration $(\mathscr{F}_n)_{n=1}^N$ if v_n is $\mathscr{F}_{n-1}$-measurable for $n = 1, \dots, N$ (with the understanding that $\mathscr{F}_0 = \{\varnothing, \Omega\}$, so v_1 is constant almost surely). If $M = (M_n)_{n=1}^N$ is an E-valued martingale with respect to $(\mathscr{F}_n)_{n=1}^N$, the sequence $v * M = ((v * M)_n)_{n=1}^N$ defined by

$$(v * M)_n := \sum_{j=1}^{n} v_j (M_j - M_{j-1}), \quad n = 1, \dots, N$$

(with the understanding that $M_0 = 0$) is called the *martingale transform* of M by v.

The intuitive meaning is as follows. Suppose the increment $M_j - M_{j-1}$ represents the outcome of the j-th gambling game. The assumption that M is a martingale means that the game is fair. Let v_j be the stake a player puts on this game. The requirement that v_j be $\mathscr{F}_{j-1}$-measurable means that the stake has to be decided knowing the outcomes of the first $j-1$ games only. The random variable $(v * M)_n$ then represents the total winnings after game n. An obvious question is whether the player can devise a favourable strategy. Under a mild additional assumption the answer is 'no': if the v_n are bounded, then $v * M$ is a martingale with respect to $(\mathscr{F}_n)_{n=1}^N$. Let us prove this. Clearly, $v * M$ is adapted with respect to $(\mathscr{F}_n)_{n=1}^N$ and the random variables $v_n(M_n - M_{n-1})$ are integrable. By Proposition 11.6 (3) and the $\mathscr{F}_{n-1}$-measurability of $(v * M)_{n-1}$ and v_n,

$$\mathbb{E}((v * M)_n)|\mathscr{F}_{n-1}) = (v * M)_{n-1} + v_n \mathbb{E}(M_n - M_{n-1}|\mathscr{F}_{n-1}) = (v * M)_{n-1}.$$

11.4 L^p-martingales

An important inequality for L^p-martingales, due to Doob, states that for $1 < p < \infty$ the maximum of an L^p-martingale is in L^p again.

Let $M = (M_n)_{n=1}^N$ be an E-valued martingale with respect to $\mathbb{F} = (\mathscr{F}_n)_{n=1}^N$ and define $M_N^* : \Omega \to \mathbb{R}_+$ by

$$M_N^* := \max_{1 \leqslant n \leqslant N} \|M_n\|.$$

Theorem 11.20 (Doob). *For all $r > 0$ we have*

$$\mathbb{P}\{M_N^* > r\} \leqslant \frac{1}{r} \mathbb{E}\|M_N\|.$$

If $1 < p < \infty$ and $M_N \in L^p(\Omega; E)$, then $M_N^* \in L^p(\Omega)$ and

$$\|M_N^*\|_p \leqslant \frac{p}{p-1} \|M_N\|_p.$$

Proof. The proof proceeds in two steps.

Step 1 – We claim that for all $r > 0$,

$$r\mathbb{P}\{M_N^* > r\} \leqslant \mathbb{E}(1_{\{M_N^* > r\}} \|M_N\|). \tag{11.3}$$

This implies the first inequality.

Let us fix $r > 0$ and define $\tau : \Omega \to \{1, \dots, N+1\}$ by $\tau := \min\{1 \leqslant n \leqslant N : \|M_n\| > r\}$ with the convention that $\min \varnothing := N+1$. Then $\{M_N^* > r\} = \{\tau \leqslant N\}$. On the set $\{\tau = n\}$ we have $\|M_n\| > r$ and therefore

$$\begin{aligned} r\mathbb{P}\{M_N^* > r\} = r \sum_{n=1}^N \mathbb{P}\{\tau = n\} &\leqslant \sum_{n=1}^N \mathbb{E}(1_{\{\tau=n\}} \|M_n\|) \\ &\overset{(*)}{\leqslant} \sum_{n=1}^N \mathbb{E}(1_{\{\tau=n\}} \|M_N\|) = \mathbb{E}(1_{\{\tau \leqslant N\}} \|M_N\|) \\ &= \mathbb{E}(1_{\{M_N^* > r\}} \|M_N\|) \end{aligned}$$

which gives (11.3). The inequality $(*)$ follows from the martingale property, since almost surely we have

$$\|M_n\| = \|\mathbb{E}(M_N | \mathscr{F}_n)\| \leqslant \mathbb{E}(\|M_N\| | \mathscr{F}_n).$$

Step 2 – Next let $1 < p < \infty$ and assume that $\|M_N\|_p < \infty$. We may assume that $\|M_N^*\|_p > 0$, since otherwise there is nothing to prove. Integrating by parts and using (11.3) and Hölder's inequality,

$$\begin{aligned} \|M_N^*\|_p^p = \int_0^\infty pr^{p-1} \mathbb{P}\{M_N^* > r\}\, dr &\leqslant \int_0^\infty pr^{p-2} \mathbb{E}(1_{\{M_N^* > r\}} \|M_N\|)\, dr \\ &= \mathbb{E}\Big(\|M_N\| \int_0^{M_N^*} pr^{p-2}\, dr \Big) \\ &= \frac{p}{p-1} \mathbb{E}(\|M_N\| (M_N^*)^{p-1}) \\ &\leqslant \frac{p}{p-1} \|M_N\|_p \|M_N^*\|_p^{p-1}. \end{aligned}$$

The result follows upon dividing both sides by $\|M_N^*\|_p^{p-1}$. □

We shall apply the first part of Doob's inequality to prove the following result on convergence of certain L^p-martingales.

Suppose a filtration $(\mathscr{F}_n)_{n=1}^\infty$ is given on $(\Omega, \mathscr{F}, \mathbb{P})$. We denote by $\mathscr{F}_\infty$ the σ-algebra generated by $(\mathscr{F}_n)_{n=1}^\infty$, that is, $\mathscr{F}_\infty$ is the smallest σ-algebra containing each of the $\mathscr{F}_n$.

Theorem 11.21. *Let $1 \leqslant p < \infty$ and assume that $X \in L^p(\Omega; E)$. Then,*

$$\lim_{n\to\infty} \mathbb{E}(X | \mathscr{F}_n) = \mathbb{E}(X | \mathscr{F}_\infty)$$

both in $L^p(\Omega; E)$ and almost surely.

Proof. We claim that $\bigcup_{n=1}^\infty L^p(\Omega, \mathscr{F}_n; E)$ is dense in $L^p(\Omega, \mathscr{F}_\infty; E)$. Assuming this for the moment we first show how the L^p-convergence is obtained from this.

For all $Y \in L^p(\Omega, \mathscr{F}_m; E)$ and $n \geqslant m$ we have $\mathbb{E}(Y|\mathscr{F}_n) = \mathbb{E}(Y|\mathscr{F}_\infty) = Y$, and therefore we trivially have $\lim_{n\to\infty} \mathbb{E}(Y|\mathscr{F}_n) = Y$ in $L^p(\Omega, \mathscr{F}_\infty; E)$. Since the conditional operators are contractive, it follows that $\lim_{n\to\infty} \mathbb{E}(Y|\mathscr{F}_n) = Y$ in $L^p(\Omega, \mathscr{F}_\infty; E)$ for all $Y \in L^p(\Omega, \mathscr{F}_\infty; E)$. In particular this is true for $Y = \mathbb{E}(X|\mathscr{F}_\infty)$.

Let us next prove that $\bigcup_{n=1}^\infty L^p(\Omega, \mathscr{F}_n; E)$ is dense in $L^p(\Omega, \mathscr{F}_\infty; E)$. Let $\mathscr{G}$ be the collection of all sets $G \in \mathscr{F}_\infty$ with the property that for all $\varepsilon > 0$ there exists an $n \geqslant 1$ and a set $F \in \mathscr{F}_n$ such that $\mathbb{P}(F \Delta G) < \varepsilon$. Here, $F \Delta G = (F \setminus G) \cup (G \setminus F)$ is the symmetric difference of F and G. It is easily checked that the collection of all approximable sets is a sub-σ-algebra of $\mathscr{F}_\infty$. Clearly, this σ-algebra contains each $\mathscr{F}_n$, and therefore it contains $\mathscr{F}_\infty$.

By what we have shown so far, $G \in \mathscr{F}_\infty$ implies that $1_G = \lim_{k\to\infty} 1_{G_k}$ in $L^p(\Omega; E)$, where $G_k \in \mathscr{F}_{n_k}$ for some $n_k \geqslant 1$. It follows that every simple function of $L^p(\Omega, \mathscr{F}_\infty; E)$ is contained in the closure of $\bigcup_{n=1}^\infty L^p(\Omega, \mathscr{F}_n; E)$ in $L^p(\Omega, \mathscr{F}_\infty; E)$. As a consequence, all of $L^p(\Omega, \mathscr{F}_\infty; E)$ is contained in the closure of $\bigcup_{n=1}^\infty L^p(\Omega, \mathscr{F}_n; E)$.

So far we have proved the L^p-convergence. To prove the almost sure convergence, note that by the first part of Theorem 11.20 and monotone convergence we have
$$\mathbb{P}\Big\{\sup_{n\geqslant 1} \|M_n\| > r\Big\} \leqslant \frac{1}{r} \sup_{n\geqslant 1} \mathbb{E}\|M_n\|,$$
where we put $M_n := \mathbb{E}(X|\mathscr{F}_n)$ for brevity. Applying this with X replaced by $X - M_N$, for all $n \geqslant N$ we obtain
$$\mathbb{P}\Big\{\sup_{n\geqslant N} \|M_n - M_N\| > r\Big\} \leqslant \frac{1}{r} \sup_{n\geqslant N} \mathbb{E}\|M_n - M_N\|.$$
By what we have proved already we find indices $N_1 < N_2 < \dots$ such that
$$\sup_{n\geqslant N_k} \mathbb{E}\|M_n - M_{N_k}\| < \frac{1}{2^{2k}}.$$
With $r = 1/2^k$ this gives
$$\mathbb{P}\Big\{\sup_{n\geqslant N_k} \|M_n - M_{N_k}\| > \frac{1}{2^k}\Big\} \leqslant \frac{1}{2^k}.$$
The Borel-Cantelli lemma now implies that $\lim_{n\to\infty} M_n = M_\infty = \mathbb{E}(X|\mathscr{F}_\infty)$ almost surely. □

11.5 Exercises

1. Let f, g be random variables on Ω. Prove that if f is $\sigma(g)$-measurable, then $f = \phi \circ g$ for some Borel function ϕ.
   Hint: First suppose that $f = 1_A$ with $A \in \sigma(g)$.

2. a) Let $X_1, \dots, X_N$ be independent and identically distributed integrable E-valued random variables and put $S_N = X_1 + \dots + X_N$. Show that $\mathbb{E}(X_1|S_N) = \dots = \mathbb{E}(X_N|S_N)$ and deduce that $\mathbb{E}(X_n|S_N) = S_N/N$ for all $n = 1, \dots, N$.
   b) (!) Let X and Y be independent and identically distributed integrable E-valued random variables on Ω. Prove that $\mathbb{E}(X - Y|X + Y) = 0$.

3. Let $(M_i)_{i\in I}$ be a martingale with respect to the filtration $(\mathscr{F}_i)_{i\in I}$, and let $(\mathscr{G}_i)_{i\in I}$ be a filtration which is independent of $(\mathscr{F}_i)_{i\in I}$. Define the filtration $(\mathscr{H}_i)_{i\in I}$ by $\mathscr{H}_i := \sigma(\mathscr{F}_i, \mathscr{G}_i)$. Show that $(M_i)_{i\in I}$ is a martingale with respect to $(\mathscr{H}_i)_{i\in I}$.

4. Let W_H be an H-cylindrical Brownian motion. The filtration $(\mathscr{F}_t^{W_H})_{t\in[0,T]}$ generated by W_H is defined by $\mathscr{F}_t^{W_H} := \sigma(W_H(s)h : \ s \in [0,t], \ h \in H)$.
   a) Show that for all $h \in H$ and $0 \leqslant s \leqslant t \leqslant T$ the increment $W_H(t)h - W_H(s)h$ is independent of $\mathscr{F}_s^{W_H}$.
   b) Show that for all $h \in H$ the Brownian motion $(W(t)h)_{t\in[0,T]}$ is a martingale with respect to $(\mathscr{F}_t^{W_H})_{t\in[0,T]}$.

5. This exercise is a continuation of Exercise 6.3 on averaging operators. Using the notations introduced there, show that for all $f \in L^p(0,T;E)$ we have $\lim_{n\to\infty} A_n f = f$ almost everywhere.

Notes. An elementary introduction to the theory of martingales is the book by Williams [109]; for more comprehensive treatments we refer to Kallenberg [55] and Rogers and Williams [95]. A systematic account of the vector-valued theory can be found in Diestel and Uhl [36].

The results of Sections 11.1 and 11.2 are standard. The approach taken in Section 11.1 by first defining conditional expectations in $L^2(\Omega)$ by an orthogonal projection is the most elementary one and, in our opinion, the most intuitive. A shorter, but less elementary approach is to *define* conditional expectations in $L^1(\Omega)$ by the identity (11.1) and then to use the Radon-Nikodým theorem to prove their existence and uniqueness.

For further results on vector-valued extensions of positive operators we refer to the nice paper by Haase [46].

The proof of Doob's inequality (Theorem 11.20) is standard and can be found in many textbooks. It only requires the fact that $(\|M_n\|)_{n=1}^N$ is a non-negative submartingale, that is, it satisfies $\|M_n\| \leqslant \mathbb{E}(\|M_n\||\mathscr{F}_m)$ almost surely for all $m \leqslant n$.

The proof of the martingale convergence theorem (Theorem 11.21) is taken from [36]. In the scalar theory it is true that any L^1-bounded martingale converges almost surely, with convergence in L^1 if the martingale is uniformly integrable (which is the case, e.g., if the martingale is L^p-bounded for some $1 < p < \infty$). For Banach space-valued martingales E, the same result holds if E has the so-called Radon-Nikodým property. Examples of spaces with this property are reflexive spaces and separable dual spaces. We refer to [36] for the full story.

It is worth mentioning the following result of Davis, Ghoussoub, Johnson, Kwapień, Maurey [30], which generalises the Itô-Nisio theorem to E-valued martingales:

Theorem 11.22. *Let E be an arbitrary Banach space and suppose that $(M_n)_{n=1}^{\infty}$ is an L^1-bounded E-valued martingale. For an E-valued random variable M the following assertions are equivalent:*

(1) *For all $x^* \in E^*$ we have $\lim_{n\to\infty}\langle M_n, x^*\rangle = \langle M, x^*\rangle$ almost surely;*
(2) *For all $x^* \in E^*$ we have $\lim_{n\to\infty}\langle M_n, x^*\rangle = \langle M, x^*\rangle$ in probability;*
(3) $\lim_{n\to\infty} M_n = M$ *almost surely;*
(4) $\lim_{n\to\infty} M_n = M$ *in probability.*

If $M \in L^p(\Omega; E)$ for some $1 \leqslant p < \infty$, then $M_n \in L^p(\Omega; E)$ for all $n \geqslant 1$ and we have $\lim_{n\to\infty} M_n = M$ in $L^p(\Omega; E)$.

Note that the Itô-Nisio theorem holds without any integrability conditions. It is clear that in the above theorem we need to impose integrability of the random variables M_n in order to define the their conditional expectations. In [30] a simple example is given which shows that even the L^1-boundedness condition on the M_n cannot be omitted.