MARTINGALES AND HARMONIC ANALYSIS

TUOMAS HYTÖNEN

Lecture notes of a course at the University of Helsinki, Autumn 2012.

1. Conditional expectation

1.1. **Basic notions of measure theory.** A triplet $(\Omega, \mathscr{F}, \mu)$ is called a *measure space* if

- Ω is a set,
- $\mathscr{F}$ is a *σ-algebra* of Ω, i.e., a collection of subsets of Ω which satisfies

$$\varnothing, \Omega \in \mathscr{F}, \qquad E \in \mathscr{F} \Rightarrow E^{\mathsf{C}} := \Omega \setminus E \in \mathscr{F}, \qquad E_i \in \mathscr{F} \Rightarrow \bigcup_{i=0}^{\infty} E_i \in \mathscr{F},$$

- μ is a *measure*, i.e., a mapping $\mathscr{F} \to [0, \infty]$ which satisfies

$$\mu(\varnothing) = 0, \qquad E_i \in \mathscr{F}, E_i \cap E_j = \varnothing \text{ for } i \neq j \Rightarrow \mu\Big(\bigcup_{i=0}^{\infty} E_i\Big) = \sum_{i=0}^{\infty} \mu(E_i).$$

A function $f : \Omega \to \mathbb{R}$ is called *$\mathscr{F}$-measurable* if $f^{-1}(B) := \{f \in B\} := \{\omega \in \Omega : f(\omega) \in B\} \in \mathscr{F}$ for all Borel sets $B \subseteq \mathbb{R}$. We denote by $L^0(\mathscr{F})$ the space of all $\mathscr{F}$-measurable functions $f : \Omega \to \mathbb{R}$. Most of the time, we think of the space Ω and the measure μ as fixed, but we may consider different σ-algebras (and other families of sets) $\mathscr{F}$; that is why we emphasize $\mathscr{F}$ but not Ω and μ in the notation.

1.2. **More definitions.** We say that a collection $\mathscr{A}$ of subsets of Ω *contains a countable cover* if there are at most countably many sets $A_0, A_1, \ldots \in \mathscr{A}$ such that $\Omega = \bigcup_{i=0}^{\infty} A_i$. We say that a subset $\mathscr{A} \subseteq \mathscr{F}$ of a σ-algebra $\mathscr{F}$ is an *$\mathscr{F}$-ideal* if $A \cap F \in \mathscr{A}$ for all $A \in \mathscr{A}$ and all $F \in \mathscr{F}$. Note that if an $\mathscr{F}$-ideal $\mathscr{A}$ contains a countable cover, then this cover can be chosen pairwise disjoint; indeed, if $A_0, A_1, \ldots \in \mathscr{A} \subseteq \mathscr{F}$ form a cover, then $F_k := \bigcup_{j=0}^{k} A_j \in \mathscr{F}$ and

$$A'_k := A_k \setminus F_{k-1} = A_k \cap F_{k-1}^{\mathsf{C}} \in \mathscr{A}$$

by the ideal property, and these form a disjoint countable cover.

We denote by $\mathscr{F}^0$ the collection of sets in $\mathscr{F}$ with finite measure, i.e.,

$$\mathscr{F}^0 := \{E \in \mathscr{F} : \mu(E) < \infty\}.$$

Clearly $\mathscr{F}^0$ is an $\mathscr{F}$-ideal. A σ-algebra $\mathscr{F}$ is called *σ-finite* if $\mathscr{F}^0$ contains a countable cover.

1.3. **Lemma.** *Let $\mathscr{A}, \mathscr{B} \subseteq \mathscr{F}$ be $\mathscr{F}$-ideals. Then*

$$\mathscr{A} \cap \mathscr{B} = \{A \cap B : A \in \mathscr{A}, B \in \mathscr{B}\},$$

and this set is an $\mathscr{F}$-ideal. If both ideals $\mathscr{A}, \mathscr{B}$ contain countable covers, then so does $\mathscr{A} \cap \mathscr{B}$.

Proof. If $A \in \mathscr{A}$ and $B \in \mathscr{B} \subseteq \mathscr{F}$, then $A \cap B \in \mathscr{A}$ by the ideal property of A, and similarly $A \cap B \in \mathscr{B}$ by the ideal property of B. Thus $A \cap B \in \mathscr{A} \cap \mathscr{B}$. On the other hand, if $E \in \mathscr{A} \cap \mathscr{B}$, then $E = A \cap B$, where $A := E \in \mathscr{A}$ and $B := E \in \mathscr{B}$. The ideal property of $\mathscr{A} \cap \mathscr{B}$ is immediate to check. If $\{A_i : i \in \mathbb{N}\} \subseteq \mathscr{A}$ and $\{B_j : j \in \mathbb{N}\} \subseteq \mathscr{B}$ are countable covers, then $\{A_i \cap B_j : i, j \in \mathbb{N}\} \subseteq \mathscr{A} \cap \mathscr{B}$ is also a countable cover. □

1.4. **Lemma.** *Let $\mathscr{A} \subseteq \mathscr{F}$ be an $\mathscr{F}$-ideal that contains a countable cover. If $f \in L^1_{\mathscr{A}}(\mathscr{F}, \mu)$ satisfies $\int_E f \, \mathrm{d}\mu \geq 0$ for all $E \in \mathscr{A}$, then $f \geq 0$ a.e. (almost everywhere). The same is true if "$\geq$" is replaced by "$\leq$" or "$=$".*

Version: January 17, 2013.

Proof. Let $F_i := \{f < -1/i\} \in \mathscr{F}$ and $A_j \in \mathscr{A}$ be one of the sets from the definition of countable cover. Since $\mathscr{A}$ is an ideal, we have $F_i \cap A_j \in \mathscr{A}$ and then

$$0 \leq \int_{F_i \cap A_j} f \, \mathrm{d}\mu \leq \int_{F_i \cap A_j} \big(-\frac{1}{i}\big) \, \mathrm{d}\mu = -\frac{1}{i}\mu(F_i \cap A_j) \leq 0.$$

Hence $\mu(F_i \cap A_j) = 0$, and summing up over $j \in \mathbb{N}$ it follows that $\mu(F_i) \leq \sum_{j=0}^{\infty} \mu(F_i \cap A_j) = 0$. Since $\{f < 0\} = \bigcup_{i=1}^{\infty} F_i$, we see that $\mu(\{f < 0\}) = 0$, which is the same as $f \geq 0$ a.e.

The case $\leq$ is obtained from the one already treated by considering the function $-f$. The case $=$ follows from the other two upon observing that $x = 0$ if and only if $x \geq 0$ and $x \leq 0$. □

1.5. **Sub-σ-algebra.** If $\mathscr{G} \subseteq \mathscr{F}$ is another σ-algebra, it is called a sub-σ-algebra of $\mathscr{F}$. In this situation the $\mathscr{G}$-measurability of a function is a stronger requirement than its $\mathscr{F}$-measurability, since there are fewer choices for the preimages $\{f \in B\}$. Similarly the σ-finiteness of $(\Omega, \mathscr{G}, \mu)$ is a stronger requirement than that of $(\Omega, \mathscr{F}, \mu)$. In the sequel, however, all measure spaces are assumed to be σ-finite unless otherwise mentioned.

For $f \in L^0(\mathscr{F})$, we denote

$$\mathscr{G}_f := \{G \in \mathscr{G} : 1_G f \in L^1(\mathscr{F})\}.$$

Clearly this is a $\mathscr{G}$-ideal.

1.6. **Lemma.** *For every $f \in L^0(\mathscr{F})$, the collection $\mathscr{F}_f$ contains a countable cover.*

It is important to observe that, in this lemma, we have the same σ-algebra $\mathscr{F}$ is "$f \in L^0(\mathscr{F})$" and in "$\mathscr{F}_f$" here. It is not always true that $\mathscr{G}_f$ contains a countable cover if $\mathscr{G} \subsetneq \mathscr{F}$ is a smaller σ-algebra than the one with respect to which f is measurable.

Proof. Let $E_i \in \mathscr{F}^0$ be the sets from the definition of σ-finiteness of $\mathscr{F}$, and let $F_j := \{|f| \leq j\} \in \mathscr{F}$. Since f is real-values, we have $\Omega = \{|f| < \infty\} = \bigcup_{j=0}^{\infty} F_j$. Then each $1_{E_i \cap F_j} f \in L^1(\mathscr{F})$; indeed $\|1_{E_i \cap F_j} f\|_1 \leq j \cdot \mu(E_i) < \infty$. Thus $\{E_i \cap F_j : i, j \in \mathbb{N}\} \subseteq \mathscr{F}_f$ is a countable cover of Ω. □

1.7. **The conditional expectation.** Let $f \in L^0(\mathscr{F})$ and $g \in L^0(\mathscr{G})$, where $\mathscr{G} \subseteq \mathscr{F}$ is a σ-finite sub-σ-algebra. We say that g is a *conditional expectation* of f with respect to $\mathscr{G}$ if there exists a $\mathscr{G}$-ideal $\mathscr{A} \subseteq \mathscr{G}_f \cap \mathscr{G}_g$ that contains a countable cover and satisfies

$$\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu \qquad \forall\, A \in \mathscr{A}.$$

1.8. **Lemma.** *If $f \in L^0(\mathscr{F})$ has a conditional expectation $g \in L^0(\mathscr{G})$, then it is unique (a.e.). Moreover, $\mathscr{G}_f$ contains a countable cover, we have $\mathscr{G}_f \subseteq \mathscr{G}_g$, and*

$$\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu \qquad \forall A \in \mathscr{G}_f.$$

Proof. We prove the last assertion first. So let g be a conditional expectation, and $\mathscr{A} \subseteq \mathscr{G}_f \cap \mathscr{G}_g$ an associated $\mathscr{G}$-ideal with a countable cover $\{A_i : i \in \mathbb{N}\}$ which we now choose disjoint. Let $G \in \mathscr{G}_f$ be arbitrary, and note that $\{g \geq 0\}, \{g < 0\} \in \mathscr{G}$. By the ideal property, we find that $A_i \cap G \cap \{g \geq 0\} \in \mathscr{A}$, and thus

$$\int_{A_i \cap G \cap \{g \geq 0\}} g \, \mathrm{d}\mu = \int_{A_i \cap G \cap \{g \geq 0\}} f \, \mathrm{d}\mu.$$

Summing over $i \in \mathbb{N}$, we get

$$\int_{G \cap \{g \geq 0\}} g \, \mathrm{d}\mu = \int_{G \cap \{g \geq 0\}} f \, \mathrm{d}\mu.$$

Similarly, we get $\int_{G \cap \{g < 0\}} g \, \mathrm{d}\mu = \int_{G \cap \{g < 0\}} f \, \mathrm{d}\mu$, and hence

$$\int_G |g| \, \mathrm{d}\mu = \int_{G \cap \{g \geq 0\}} g \, \mathrm{d}\mu - \int_{g \cap \{g < 0\}} g \, \mathrm{d}\mu \leq \int_G |f| \, \mathrm{d}\mu < \infty \qquad G \in \mathscr{G}_g,$$
$$\int_G g \, \mathrm{d}\mu = \int_{G \cap \{g \geq 0\}} g \, \mathrm{d}\mu + \int_{g \cap \{g < 0\}} g \, \mathrm{d}\mu = \int_G f \, \mathrm{d}\mu.$$

So indeed $\mathscr{G}_f \subseteq \mathscr{G}_g$, and we have $\int_G g\,\mathrm{d}\mu = \int_G f\,\mathrm{d}\mu$ for all $G \in \mathscr{G}_f$. Clearly $\mathscr{G}_f$ contains a countable cover, since $\mathscr{G}_f \supseteq \mathscr{A}$, and $\mathscr{A}$ contains a countable cover.

Suppose then that $g_1, g_2 \in L^0(\mathscr{G})$ are two conditional expectations of f. By the part that we already proved, we find that

$$\int_G g_1\,\mathrm{d}\mu = \int_G f\,\mathrm{d}\mu = \int_G g_2\,\mathrm{d}\mu \qquad \forall G \in \mathscr{G}_f,$$

where $\mathscr{G}_f$ is a $\mathscr{G}$-ideal that contains a countable cover. Thus $g = g_1 - g_2 \in L^0(\mathscr{G})$ satisfies $\int_G g\,\mathrm{d}\mu = 0$ for all $G \in \mathscr{G}_f$, and Lemma 1.4 shows that $g_1 = g_2$ almost everywhere. □

The conditional expectation of f with respect to $\mathscr{G}$, now that it has been proven unique, will be denoted by $\mathbb{E}[f|\mathscr{G}]$. Next it will be shown that it always exists, provided only that $\mathscr{G}_f$ contains a countable cover. Clearly this condition is necessary by Lemma 1.8.

1.9. **Existence for L^2-functions.** The spaces $L^2(\mathscr{F},\mu)$ and $L^2(\mathscr{G},\mu)$ are both Hilbert spaces, and the latter is a closed subspace of the former. If $f \in L^2(\mathscr{F},\mu)$, let $g \in L^2(\mathscr{G},\mu)$ be its orthogonal projection onto the space $L^2(\mathscr{G},\mu)$. Hence $f - g \perp L^2(\mathscr{G},\mu)$. If $G \in \mathscr{G}^0$, then $1_G \in L^2(\mathscr{G},\mu)$. Hence

$$0 = (f-g, 1_G) = \int_G (f-g)\,\mathrm{d}\mu,$$

and thus g satisfies the definition of conditional expectation for the countably-covering $\mathscr{G}$-ideal $\mathscr{G}^0$.

1.10. **Simple observations.** If $g \in L^0(\mathscr{G},\mu)$, it is its own conditional expectation, $g = \mathbb{E}[g|\mathscr{G}]$. In particular, the conditional expectation of a constant function is the same constant.

Linearity. Suppose that $f_1, f_2 \in L^0(\mathscr{F})$ have conditional expectations $g_i = \mathbb{E}[f_i|\mathscr{G}] \in L^0(\mathscr{G})$, so in particular $\int_G g_i\,\mathrm{d}\mu = \int_G f_i\,\mathrm{d}\mu$ for all $G \in \mathscr{G}_{f_i} \subseteq \mathscr{G}_{g_i}$, and $\mathscr{G}_{f_i}$ contains a countable cover. By Lemma 1.3, $\mathscr{A} := \mathscr{G}_{f_1} \cap \mathscr{G}_{f_2} \subseteq \mathscr{G}_{g_1} \cap \mathscr{G}_{g_2}$ is a $\mathscr{G}$-ideal containing a countable cover, and clearly

$$\mathscr{A} \subseteq \mathscr{G}_{\alpha_1 f_1 + \alpha_2 f_2} \cap \mathscr{G}_{\alpha_1 g_1 + \alpha_2 g_2}$$

and for all $A \in \mathscr{A}$, we have

$$\int_A (\alpha_1 f_1 + \alpha_2 f_2)\,\mathrm{d}\mu = \alpha_1 \int_A f_1\,\mathrm{d}\mu + \alpha_2 \int_A f_2\,\mathrm{d}\mu = \alpha_1 \int_A g_1\,\mathrm{d}\mu + \alpha_2 \int_A g_2\,\mathrm{d}\mu = \int_A (\alpha_1 g_1 + \alpha_2 g_2)\,\mathrm{d}\mu.$$

Thus $\mathbb{E}[\alpha_1 f_1 + \alpha_2 f_2|\mathscr{G}]$ exists and equals $\alpha_1 \mathbb{E}[f_1|\mathscr{G}] + \alpha_2 \mathbb{E}[f_2|\mathscr{G}]$.

Comparison. If the pointwise (a.e.) inequality $f_1 \le f_2$ holds, then also $\mathbb{E}[f_1|\mathscr{G}] \le \mathbb{E}[f_2|\mathscr{G}]$. This follows from the fact that for all $G \in \mathscr{G}_{f_1} \cap \mathscr{G}_{f_2}$ (which is a $\mathscr{G}$-ideal containing a countable cover by Lemma 1.3) we have

$$\int_G \mathbb{E}[f_1|\mathscr{G}]\,\mathrm{d}\mu = \int_G f_1\,\mathrm{d}\mu \le \int_G f_2\,\mathrm{d}\mu = \int_G \mathbb{E}[f_2|\mathscr{G}]\,\mathrm{d}\mu$$

and Lemma 1.4.

This implies that if $f \in L^0(\mathscr{F},\mu)$, and both $\mathbb{E}[f|\mathscr{G}]$ and $\mathbb{E}[|f|\,|\mathscr{G}]$ exist, there holds

$$|\mathbb{E}[f|\mathscr{G}]| = \max\big\{\mathbb{E}[f|\mathscr{G}], -\mathbb{E}[f|\mathscr{G}]\big\} = \max\big\{\mathbb{E}[f|\mathscr{G}], \mathbb{E}[-f|\mathscr{G}]\big\} \le \mathbb{E}[|f|\,|\mathscr{G}],$$

where the last estimate was based on the facts that both $f \le |f|$ and $-f \le |f|$.

1.11. **Existence for L^1-functions.** Let then $f \in L^1(\mathscr{F},\mu)$. By basic integration theory there exists a sequence of functions $f_n \in L^1(\mathscr{F},\mu) \cap L^2(\mathscr{F},\mu)$ such that $f_n \to f$ in $L^1(\mathscr{F},\mu)$. By 1.9 the conditional expectations $g_n := \mathbb{E}[f_n|\mathscr{G}]$ and $\mathbb{E}[|f_n|\,|\mathscr{G}]$ exist and belong to $L^2(\mathscr{G},\mu)$.

Note that $\mathscr{G}_{f_n} = \mathscr{G} = \mathscr{G}_{g_n}$ for $f_n \in L^1(\mathscr{F})$, and hence

$$\|g_n\|_1 = \int_\Omega |g_n|\,\mathrm{d}\mu = \int_\Omega |\mathbb{E}[f_n|\mathscr{G}]|\,\mathrm{d}\mu \le \int_\Omega \mathbb{E}[|f_n|\,|\mathscr{G}]\,\mathrm{d}\mu = \int_\Omega |f_n|\,\mathrm{d}\mu = \|f_n\|_1.$$

Repeating the previous computation with g_n replaced by $g_n - g_m$, it similarly follows that $\|g_n - g_m\|_1 \le \|f_n - f_m\|_1$, and this tends to zero as $n, m \to \infty$, since $f_n \to f$. Hence $(g_n)_{n=1}^\infty$ is a

Cauchy sequence in $L^1(\mathscr{G})$ and hence converges to some functions $g \in L^1(\mathscr{G})$. This g satisfies, for all $G \in \mathscr{G}$, the equality

$$\int_G g \,\mathrm{d}\mu = \lim_{n\to\infty} \int_G g_n \,\mathrm{d}\mu = \lim_{n\to\infty} \int_G \mathbb{E}[f_n|\mathscr{G}] \,\mathrm{d}\mu = \lim_{n\to\infty} \int_G f_n \,\mathrm{d}\mu = \int_G f \,\mathrm{d}\mu,$$

and hence $g = \mathbb{E}[f|\mathscr{G}]$.

1.12. **Existence in general.** Let finally $f \in L^0(\mathscr{F})$ such that $\mathscr{G}_f$ contains a countable cover. Let $G_i \in \mathscr{G}_f$ be disjoint sets such that $\bigcup_{i=0}^\infty G_i = \Omega$, which thus can be chosen. Now $f_i := 1_{G_i} f \in L^1(\mathscr{F})$, so there exists $g_i := \mathbb{E}[f_i|\mathscr{G}] \in L^1(\mathscr{G}, \mu)$.

We set $g := \sum_{i=0}^\infty 1_{G_i} g_i$, which converges pointwise trivially, since the G_i are disjoint sets. We check that g is a conditional expectation of f, with an associated $\mathscr{G}$-ideal

$$\mathscr{A} := \{G \in \mathscr{G} : G \subseteq G_i \text{ for some } i \in \mathbb{N}\}.$$

Clearly this is an ideal containing a countable cover, namely, the sets G_i, $i \in \mathbb{N}$. Moreover, if $A \in \mathscr{A}$ with $A \subseteq G_i$, then both $1_A g = 1_A g_i$ and $1_A f = 1_A f_i$ are integrable, and

$$\int_A g \,\mathrm{d}\mu = \int_A g_i \,\mathrm{d}\mu = \int_A f_i \,\mathrm{d}\mu = \int_A f \,\mathrm{d}\mu.$$

Thus $\mathscr{A} \subseteq \mathscr{G}_f \cap \mathscr{G}_g$, and the required identity holds for all $A \in \mathscr{A}$.

Altogether, the following result has been established:

1.13. **Theorem.** *Let $f \in L^0(\mathscr{F})$, and $\mathscr{G} \subseteq \mathscr{F}$ be a sub-σ-algebra. Then $\mathbb{E}[f|\mathscr{G}] \in L^0(\mathscr{G})$ exists if and only if*

$$\mathscr{G}_f := \{G \in \mathscr{G} : 1_G f \in L^1(\mathscr{F})\}$$

contains a countable cover.

Next, some further properties of the conditional expectation will be investigated with the help of the following auxiliary result:

1.14. **Lemma.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function, i.e.,*

$$\phi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\phi(x_1) + (1-\lambda)\phi(x_2) \qquad \forall\ x_1, x_2 \in \mathbb{R},\ \forall\ \lambda \in [0,1],$$

and

$$H_\phi := \{h : \mathbb{R} \to \mathbb{R} | h(x) = ax + b \text{ for some } a, b \in \mathbb{R}, \text{ and } h \leq \phi\}.$$

Then $\phi(x) = \sup\{h(x) : h \in H_\phi\}$.

Proof. For all $h \in H_\phi$ there holds $h \leq \phi$, and hence $\sup_{h \in H_\phi} h \leq \phi$; thus it remains to prove the reverse inequality.

Let $x_0 \in \mathbb{R}$. We claim that

$$\frac{\phi(y) - \phi(x_0)}{y - x_0} \leq \frac{\phi(z) - \phi(x_0)}{z - x_0} \qquad \text{for all } y < x_0 < z.$$

Indeed, the claim is equivalent to

$$\Big(\frac{1}{x_0 - y} + \frac{1}{z - x_0}\Big)\phi(x_0) \leq \frac{1}{x_0 - y}\phi(y) + \frac{1}{z - x_0}\phi(z),$$

which can be written as

$$\phi(x_0) \leq \lambda\phi(y) + (1-\lambda)\phi(z), \qquad \lambda := \frac{1/(x_0 - y)}{1/(x_0 - y) + 1/(z - x_0)} = \frac{z - x_0}{z - y}$$

where

$$\lambda y + (1-\lambda)z = \frac{(z - x_0)y + (x_0 - y)x}{z - y} = x_0.$$

Thus the claim follows by definition of convexity with $x_1 = y$, $x_2 = z$, and $\lambda = (z - x_0)/(z - y)$.

Let us now choose any

$$a \in \Big[\sup_{y < x_0} \frac{\phi(y) - \phi(x_0)}{y - x_0}, \inf_{z > x_0} \frac{\phi(z) - \phi(x_0)}{z - x_0}\Big],$$

where the interval in nonempty by what we just proved. This choice implies that

$$a(x - x_0) \le \phi(x) - \phi(x_0) \qquad \forall x \neq x_0,$$

or in other words $h_0(x) := \phi(x_0) + a(x - x_0) \le \phi(x)$ for all $x \in \mathbb{R}$. and $h_0(x_0) = \phi(x_0)$. Since x_0 was arbitrary, this completes the proof. $\square$

1.15. **Theorem** (Jensen's inequality). *Let $f \in L^0(\mathscr{F})$, $\phi : \mathbb{R} \to \mathbb{R}$ be convex and $\mathscr{G}_f, \mathscr{G}_{\phi\circ f}$ contain countable covers. Then*

$$\phi\big(\mathbb{E}[f|\mathscr{G}]\big) \le \mathbb{E}[\phi(f)|\mathscr{G}].$$

Proof. Let $h \in H_\phi$. Then

$$h\big(\mathbb{E}[f|\mathscr{G}]\big) = a\mathbb{E}[f|\mathscr{G}] + b = \mathbb{E}[af + b|\mathscr{G}] = \mathbb{E}[h(f)|\mathscr{G}] \le \mathbb{E}[\phi(f)|\mathscr{G}].$$

Computing the supremum on the left over all $h \in H_\phi$, the claim follows $\square$

1.16. **Corollary.** *Let $p \in [1, \infty]$ and $f \in L^p(\mathscr{F})$. Then $\mathbb{E}[f|\mathscr{G}] \in L^p(\mathscr{G})$ and*

$$\|\mathbb{E}[f|\mathscr{G}]\|_p \le \|f\|_p.$$

Proof. Let $p < \infty$, the case $p = \infty$ being easier. For $f \in L^p(\mathscr{F})$, we have $\mathscr{G}_f \supseteq \mathscr{G}^0$, which contains a countable cover by σ-finiteness, so the conditional expectation $\mathbb{E}[f|\mathscr{G}]$ exists, and similarly $\mathbb{E}[|f|^p\,|\mathscr{G}]$, since $|f|^p \in L^1(\mathscr{F})$. Since the function $t \mapsto |t|^p$ is convex, Jensen's inequality implies that

$$|\mathbb{E}[f|\mathscr{G}]|^p \le \mathbb{E}[|f|^p\,|\mathscr{G}].$$

Integrating over $\Omega \in \mathscr{G}_{|f|^p}$, we get

$$\|\mathbb{E}[f|\mathscr{G}]\|_p^p = \int_\Omega |\mathbb{E}[f|\mathscr{G}]|^p \,\mathrm{d}\mu \le \int_\Omega \mathbb{E}[|f|^p\,|\mathscr{G}] \,\mathrm{d}\mu = \int_\Omega |f|^p \,\mathrm{d}\mu = \|f\|_p^p. \qquad \square$$

Next, versions of the familiar convergence theorems of integration theory are presented for the conditional expectation.

1.17. **Monotone convergence theorem.** Recall that the version of integration theory says that if a sequence of measurable functions satisfies $0 \le f_n \nearrow f$ a.e., then $\int f_n \,\mathrm{d}\mu \nearrow \int f \,\mathrm{d}\mu$, where $f_n \nearrow f$ means "converges increasingly", which entails both the convergence $f_n \to f$ and the fact that $f_n \le f_{n+1}$ for all n. The corresponding statement for the conditional expectation is the following:

$$0 \le f_n \nearrow f \in L^0(\mathscr{F}, \mu) \quad \big(\text{where } \mathscr{G}_f \text{ contains a countable cover}\big) \quad \Rightarrow \quad \mathbb{E}[f_n|\mathscr{G}] \nearrow \mathbb{E}[f|\mathscr{G}].$$

Proof. Note that each $\mathscr{G}_{f_n} \supseteq \mathscr{G}_f$ also contains a countable cover. Since the conditional expectation respects pointwise inequalities (part 1.10), it follows that

$$0 \le f_n \le f_{n+1} \le f \quad \Rightarrow \quad 0 \le \mathbb{E}[f_n|\mathscr{G}] \le \mathbb{E}[f_{n+1}|\mathscr{G}] \le \mathbb{E}[f|\mathscr{G}].$$

Hence $\big(\mathbb{E}[f_n|\mathscr{G}]\big)_{n=1}^\infty$ is a bounded increasing sequence, so it has a pointwise $\mathscr{G}$-measurable limit, $\mathbb{E}[f_n|\mathscr{G}] \nearrow g$, and $0 \le g \le \mathbb{E}[f|\mathscr{G}]$. It remains to prove that $g = \mathbb{E}[f|\mathscr{G}]$.

For all $G \in \mathscr{G}_f \subseteq \bigcap_{n=0}^\infty \mathscr{G}_{f_n}$ there holds

$$\begin{aligned}\int_G g \,\mathrm{d}\mu = \int_G \lim_{n\to\infty} \mathbb{E}[f_n|\mathscr{G}] \,\mathrm{d}\mu &\overset{\star}{=} \lim_{n\to\infty} \int_G \mathbb{E}[f_n|\mathscr{G}] \,\mathrm{d}\mu = \lim_{n\to\infty} \int_G f_n \,\mathrm{d}\mu \\ &\overset{\star}{=} \int_G \lim_{n\to\infty} f_n \,\mathrm{d}\mu = \int_G f \,\mathrm{d}\mu,\end{aligned}$$

where the steps marked with $\star$ were based on the usual monotone convergence theorem. Thus it follows that $g = \mathbb{E}[f|\mathscr{G}]$, which completes the proof. $\square$

1.18. **Fatou's lemma.** The version of integration theory says that

$$f_n \geq 0 \quad \Rightarrow \quad \int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

For the conditional expectation we similarly prove

$$\begin{aligned} &0 \leq f_n \in L^0(\mathscr{F}), \; f := \liminf_{n\to\infty} f_n \in L^0(\mathscr{F}) \\ &\qquad \big(\text{where each } \mathscr{G}_f, \mathscr{G}_{f_n} \text{ contains a countable cover}\big) \\ &\quad \Rightarrow \quad \mathbb{E}[f|\mathscr{G}] \leq \liminf_{n\to\infty} \mathbb{E}[f_n|\mathscr{G}]. \end{aligned}$$

Proof. Write out the definition of the limes inferior:

$$f = \liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{m\geq n} f_m =: \lim_{n\to\infty} h_n, \qquad h_n := \inf_{m\geq n} f_m.$$

For every $m \geq n$, we have $h_n \leq f_m$, hence $\mathbb{E}[h_n|\mathscr{G}] \leq \mathbb{E}[f_m|\mathscr{G}]$, and by taking the infimum, $\mathbb{E}[h_n|\mathscr{G}] \leq \inf_{m\geq n} \mathbb{E}[f_m|\mathscr{G}]$. Moreover, we have $0 \leq h_n \nearrow f \in L^0(\mathscr{F})$, so we can use the monotone convergence theorem to the result that

$$\mathbb{E}[f|\mathscr{G}] = \mathbb{E}[\lim_{n\to\infty} h_n|\mathscr{G}] = \lim_{n\to\infty} \mathbb{E}[h_n|\mathscr{G}] \leq \lim_{n\to\infty} \inf_{m\geq n} \mathbb{E}[f_m|\mathscr{G}] = \liminf_{n\to\infty} \mathbb{E}[f_n|\mathscr{G}]. \qquad \square$$

1.19. **Dominated convergence theorem.** In integration theory one proves that

$$f_n \to f, \; |f_n| \leq g \in L^1(\mathscr{F},\mu) \quad \Rightarrow \quad \int |f_n - f| \, \mathrm{d}\mu \to 0 \quad \Rightarrow \quad \int f_n \, \mathrm{d}\mu \to \int f \, \mathrm{d}\mu,$$

while the conditional version reads as follows:

$$\begin{aligned} &f_n \to f, \; |f_n| \leq g \in L^0(\mathscr{F},\mu) \quad \big(\text{where } \mathscr{G}_g \text{ contains a countable cover}\big) \\ &\quad \Rightarrow \quad \mathbb{E}\big[|f_n - f|\big|\mathscr{G}\big] \to 0 \quad \Rightarrow \quad \mathbb{E}[f_n|\mathscr{G}] \to \mathbb{E}[f|\mathscr{G}]. \end{aligned}$$

Proof. This is left as an exercise. □

The following central result concerning the conditional expectation has no obvious analogue in the basic integration theory:

1.20. **Theorem.** *Let $f \in L^0(\mathscr{F})$, where $\mathscr{G}_f$ contains a countable cover. If $g \in L^0(\mathscr{G})$, then also $\mathscr{G}_{g\cdot f}$ contains a countable cover, and*

$$\mathbb{E}[g \cdot f|\mathscr{G}] = g \cdot \mathbb{E}[f|\mathscr{G}].$$

Proof. Let $\{F_i : i \in \mathbb{N}\} \subseteq \mathscr{G}_f$ be a countable cover, and let $G_j := \{|g| \leq j\} \in \mathscr{G}$. Then $\|1_{F_i\cap G_j} g \cdot f\|_1 \leq j \cdot \|1_{F_i} f\|_1 < \infty$, and $\{F_i \cap G_j : i,j \in \mathbb{N}\} \subseteq \mathscr{G}_f \cap \mathscr{G}_{g\cdot f}$ is a countable cover.

To prove the identity, suppose first that g is a simple $\mathscr{G}$-measurable function, $g = \sum_{k=1}^N a_k 1_{G_k}$, where $G_k \in \mathscr{G}$. Then for all $G \in \mathscr{G}_f \cap \mathscr{G}_{g\cdot f}$ there holds

$$\int_G g \cdot \mathbb{E}[f|\mathscr{G}] \, \mathrm{d}\mu = \sum_{k=1}^N a_k \int_{G\cap G_k} \mathbb{E}[f|\mathscr{G}] \, \mathrm{d}\mu = \sum_{k=1}^N a_k \int_{G\cap G_k} f \, \mathrm{d}\mu = \int_G g \cdot f \, \mathrm{d}\mu,$$

hence $g \cdot \mathbb{E}[f|\mathscr{G}] = \mathbb{E}[g \cdot f|\mathscr{G}]$ by the uniqueness of the conditional expectation.

If g is a general $\mathscr{G}$-measurable function, by measure theory there exists a sequence of $\mathscr{G}$-simple functions g_n with $|g_n| \leq |g|$ and $g_n \to g$. Hence also $|g_n \cdot f| \leq |g \cdot f|$ and $g_n \cdot f \to g \cdot f$. An application of the dominated convergence theorem and the first part of the proof gives

$$\mathbb{E}[g \cdot f|\mathscr{G}] = \lim_{n\to\infty} \mathbb{E}[g_n \cdot f|\mathscr{G}] = \lim_{n\to\infty} g_n \cdot \mathbb{E}[f|\mathscr{G}] = g \cdot \mathbb{E}[f|\mathscr{G}],$$

which was to be proven. □

1.21. **Exercises.** These deal with some further important properties of the conditional expectation.

In all exercised it is assumed that Ω is a set, $\mathscr{F}$ and $\mathscr{G}$ are its σ-algebras with $\mathscr{G} \subseteq \mathscr{F}$, and $\mu : \mathscr{F} \to [0, \infty]$ is a measure. Moreover, all functions are assumed to be $\mathscr{F}$-measurable. Except in Exercise 1, it is also assumed that all measure spaces are σ-finite.

1. Give an example of the following situation: $(\Omega, \mathscr{F}, \mu)$ is σ-finite but $(\Omega, \mathscr{G}, \mu)$ is not.
2. Prove the dominated convergence theorem for conditional expectations. (See Section 1.19.)
3. Prove the *tower rule* of conditional expectations: Let $\mathscr{H} \subseteq \mathscr{G}$ be yet another σ-algebra, and $f \in L^0(\mathscr{F})$ be such that $\mathscr{H}_f$ (and hence also $\mathscr{G}_f \supseteq \mathscr{H}_f$) contains a countable cover. Then
$$\mathbb{E}\big(\mathbb{E}[f|\mathscr{G}]\big|\mathscr{H}\big) = \mathbb{E}[f|\mathscr{H}].$$
4. Prove the *conditional Hölder inequality*: If $f \in L^0(\mathscr{F}, \mu)$ and $g \in L^0(\mathscr{F}, \mu)$, where $\mathscr{G}_{|f|^p}$ and $\mathscr{G}_{|g|^{p'}}$ contain countable covers, then also $\mathscr{G}_{f\cdot g}$ contains a countable cover, and
$$\mathbb{E}[f \cdot g|\mathscr{G}] \leq \mathbb{E}[|f|^p|\mathscr{G}]^{1/p} \cdot \mathbb{E}[|f|^{p'}|\mathscr{G}]^{1/p'}.$$
(Hint: prove or recall first that for all $a, b \geq 0$ there holds $ab \leq a^p/p + b^{p'}/p'$.)
5. Let $\mathscr{B} = \mathscr{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$, and
$$\tilde{\mathscr{B}} := \{B \in \mathscr{B} : -B = B\}, \qquad -B := \{-x : x \in B\}.$$
Check that $\tilde{\mathscr{B}}$ is a σ-algebra, and that for every $f \in L^0(\mathbb{R}, \mathscr{B}, \mathrm{d}x)$ (where $\mathrm{d}x$ is the Lebesgue measure), $\tilde{\mathscr{B}}_f$ contains a countable cover, and
$$\mathbb{E}[f|\tilde{\mathscr{B}}](x) = \tfrac{1}{2}(f(x) + f(-x)).$$

1.22. **References.** The material of this chapter, when restricted to the case of a probability space (i.e., a measure space with $\mu(\Omega) = 1$), is standard in modern Probability and can be found in various textbooks, such as the lively presentation of Williams [16]. It is well known "among specialists" that most of the results remain true in more general measure spaces, but this extension is seldom found in a systematic way in the literature. A somewhat different framework from the present one was recently introduced by Tanaka and Terasawa [15].

2. Discrete-time martingales and Doob's inequality

2.1. **Definition.** Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and I an ordered set.

- A family of σ-algebras $(\mathscr{F}_i)_{i\in I}$ is called a *filtration* of $\mathscr{F}$ if $\mathscr{F}_i \subseteq \mathscr{F}_j \subseteq \mathscr{F}$ whenever $i, j \in I$ and $i < j$.
- A family of functions $(f_i)_{i\in I}$ is called *adapted* to the given filtration if $f_i \in L^0(\mathscr{F}_i)$ for all $i \in I$.

Let, in addition, the measure spaces $(\Omega, \mathscr{F}_i, \mu)$ be σ-finite.

- An adapted family of functions is called a *submartingale* if $f_i \in L^0(\mathscr{F}_i)$ for all $i \in I$, and for all $i < j$, the conditional expectations $\mathbb{E}[f_j|\mathscr{F}_i]$ exist and satisfy $f_i \leq \mathbb{E}[f_j|\mathscr{F}_i]$.
- It is called a martingale if the last inequality is strengthened to the equality $f_i = \mathbb{E}[f_j|\mathscr{F}_i]$ whenever $i < j$.

If $f \in L^0(\mathscr{F}, \mu)$ and $(\mathscr{F}_i)_{i\in I}$ is a filtration such that $(\mathscr{F}_i)_f$ contains a countable cover for every $i \in I$, then setting $f_i := \mathbb{E}[f|\mathscr{F}_i]$ for all $i \in I$ one gets a martingale. If $(f_i)_{i\in I}$ is a martingale, then $(|f_i|)_{i\in i}$ is a submartingale. These facts are easy to check.

In applications the index $i \in I$ often admits the interpretation of a time parameter. In these lectures the considerations are restricted to *discrete-time* filtrations and martingales, where $I \subseteq \mathbb{Z}$. If $I \subset \mathbb{Z}$ is a proper subset and $(\mathscr{F}_i)_{i\in I}$ and $(f_i)_{i\in I}$ are a filtration and a martingale with the corresponding index set, then one can always define $\mathscr{F}_i$ and f_i also for $i \in \mathbb{Z} \setminus I$ in such a way that $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ and $(f_i)_{i\in\mathbb{Z}}$ are also a filtration and an adapted martingale (exercise).

2.2. **Questions of density.** Let us denote

$$\mathscr{F}_\infty = \sigma\Big(\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i\Big).$$

Recall that the notation $\sigma(\mathscr{A})$, where $\mathscr{A}$ is any collection of subsets of Ω, designates the smallest σ-algebra of Ω which contains $\mathscr{A}$. It is obtained as the intersection of all σ-algebras containing $\mathscr{A}$: there is at least one such σ-algebra (the one containing all subsets of Ω) and one easily checks that the intersection of (arbitrarily many) σ-algebras is again a σ-algebra.

Although it is not required in the definition of a filtration, it is interesting to consider the situation where the filtration $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ generates the full σ-algebra $\mathscr{F}$, i.e., $\mathscr{F} = \mathscr{F}_\infty$. In the described situation it is natural to ask whether $\mathscr{F}$-measurable sets or functions can be approximated by sets in $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ or functions measurable with respect to these generating σ-algebras. The following results provide positive answers to these questions.

Let us denote by $\tilde{\mathscr{F}}$ the collection of those sets of $\mathscr{F}$ whose finite parts can be approximated by sets of $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$, more precisely

$$\tilde{\mathscr{F}} := \Big\{E \in \mathscr{F}\Big|\forall\, E_0 \in \mathscr{F}^0\, \forall\, \varepsilon > 0\, \exists\, F \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i : \mu(E_0 \cap [E\triangle F]) < \varepsilon\Big\}.$$

Here $E\triangle F$ designates the symmetric difference of sets, $E\triangle F := (E \setminus F) \cup (F \setminus E)$.

2.3. **Lemma.** *Let $(\mathscr{F}_i)_{i\in I}$ be a filtration, and $\mathscr{F}_\infty = \sigma\Big(\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i\Big)$. Then $\mathscr{F}_\infty = \tilde{\mathscr{F}}$.*

Proof. Clearly $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i \subseteq \tilde{\mathscr{F}} \subseteq \mathscr{F}_\infty$. (The first "$\subseteq$" follows from the fact that if $E \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$, then $F = E$ works as the approximating set in the definition of $\tilde{\mathscr{F}}$ for all E_0 and ε.) Thus it suffices to show that $\tilde{\mathscr{F}}$ is a σ-algebra. For then — due to the fact that $\mathscr{F}_\infty$ was the smallest σ-algebra containing $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ — it follows that $\mathscr{F}_\infty \subseteq \tilde{\mathscr{F}}$, and this implies the assertion.

Trivially $\varnothing, \Omega \in \tilde{\mathscr{F}}$, and the implication $E \in \tilde{\mathscr{F}} \Rightarrow E^{\mathsf{C}} \in \tilde{\mathscr{F}}$ follows from the fact that if $\mu(E_0 \cap [E\triangle F]) < \varepsilon$, then also $\mu(E_0 \cap [E^{\mathsf{C}}\triangle F^{\mathsf{C}}]) < \varepsilon$ (since $E^{\mathsf{C}}\triangle F^{\mathsf{C}} = E\triangle F$), and thus $F^{\mathsf{C}} \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ works as an approximating set for E^{C}. It remains to prove that $E_k \in \tilde{\mathscr{F}} \Rightarrow E := \bigcup_{k=1}^\infty E_k \in \tilde{\mathscr{F}}$.

Fix $E_0 \in \mathscr{F}^0$ and $\varepsilon > 0$, and denote $\mu_0(G) := \mu(E_0 \cap G)$; this is a finite measure. Since $\bigcup_{k=1}^N E_k \nearrow E$, i.e., $E \setminus \bigcup_{k=1}^N E_k \searrow \varnothing$, for sufficiently large N there holds the estimate

$$\mu_0\Big(E \setminus \bigcup_{k=1}^N E_k\Big) < \varepsilon.$$

Let $F_k \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ satisfy $\mu_0(E_k\triangle F_k) < \varepsilon \cdot 2^{-k}$. Hence $F_k \in \mathscr{F}_{i(k)}$ for some $i(k) \in \mathbb{Z}$. Let $j := \max\{i(k); k = 1, \ldots, n\}$, so that $F_k \in \mathscr{F}_j$ for all $k = 1, \ldots, n$ (since $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ is a filtration) and hence also

$$F := \bigcup_{k=1}^N F_k \in \mathscr{F}_j \subseteq \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i.$$

Now one can estimate

$$\mu_0(E \setminus F) \le \mu_0\Big(E \setminus \bigcup_{k=1}^N E_k\Big) + \sum_{k=1}^N \mu_0(E_k \setminus F_k) < \varepsilon + \sum_{k=1}^N \varepsilon \cdot 2^{-k} < 2\varepsilon,$$

$$\mu_0(F \setminus E) \le \sum_{k=1}^N \mu_0(F_k \setminus E_k) < \varepsilon,$$

hence $\mu_0(E\triangle F) < 3\varepsilon$, and the proof is complete. □

2.4. **Lemma.** *Let the assumption of Lemma 2.3 be satisfied, and in addition the measure spaces $(\Omega, \mathscr{F}_i, \mu)$ be σ-finite. If $E \in \mathscr{F}^0$, then for all $\varepsilon > 0$ one can find an $F \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$, such that $\mu(E\triangle F) < \varepsilon$.*

The difference compared to the previous lemma is the fact that there is an estimate for the measure of the full difference set $E\triangle F$ and not only its intersection with a given E_0.

Proof. Since (e.g.) $\mathscr{F}_0$ is σ-finite, there are sets $A_k \in \mathscr{F}_0^0$ of finite measure with $A_k \nearrow \Omega$. Then $E \setminus A_k \searrow \varnothing$, so for some k there holds $\mu(E \setminus A_k) < \varepsilon$. Set $E_0 := A_k$ and apply Lemma 2.3. This gives a set $F \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ such that $\mu(E_0 \cap [E\triangle F]) < \varepsilon$. Also $F_0 := E_0 \cap F \in \bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$, and this satisfies

$$\mu(E \setminus F_0) = \mu(E \setminus E_0) + \mu(E_0 \cap E \setminus F) < 2\varepsilon, \qquad \mu(F_0 \setminus E) = \mu(E_0 \cap F \setminus E) < \varepsilon.$$

Hence F_0 is a set of the desired type (with the value 3ε). □

As a consequence we get a density result for functions:

2.5. **Theorem.** *Let $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ be a filtration of the space $(\Omega, \mathscr{F}, \mu)$, where the associated measure spaces are σ-finite and $\mathscr{F} = \sigma\Big(\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i\Big)$. Let $p \in [1,\infty)$. Then*

$$\bigcup_{i\in\mathbb{Z}} L^p(\mathscr{F}_i, \mu)$$

is dense in $L^p(\mathscr{F}, \mu)$.

Proof. Let $f \in L^p(\mathscr{F}, \mu)$. By integration theory there exists a simple function $g = \sum_{k=1}^N a_k 1_{E_k}$, where $E_k \in \mathscr{F}^0$, such that $\|f - g\|_p < \varepsilon$. By Lemma 2.4 there are sets $F_k \in \mathscr{F}_{i(k)}^0 \subseteq \mathscr{F}_j^0$, where $j := \max\{i(k) : k = 1, \ldots, N\}$, such that $\mu_0(E_k \triangle F_k) < \delta$. Letting $h := \sum_{k=1}^N a_k 1_{F_k}$, it follows that

$$\|g - h\|_p \leq \sum_{k=1}^N |a_k| \cdot \|1_{E_k} - 1_{F_k}\|_p = \sum_{k=1}^N |a_k| \cdot \mu(E_k \triangle F_k)^{1/p} < \delta^{1/p} \sum_{k=1}^N |a_k| < \varepsilon,$$

as soon as δ is chosen sufficiently small. Hence $\|f - h\|_p < 2\varepsilon$ and $h \in L^p(\mathscr{F}_j, \mu)$. □

2.6. **Lemma.** *For $f \in L^p(\mathscr{F})$, $p \in [1,\infty]$, and $\mathscr{G} \subseteq \mathscr{F}$, we have*

$$\|f - \mathbb{E}[f|\mathscr{G}]\|_p \leq 2 \inf_{g\in L^p(\mathscr{G})} \|f - g\|_p.$$

Proof. Let $g \in L^p(\mathscr{G})$ be arbitary. Then $g = \mathbb{E}[g|\mathscr{G}]$, so using the linearity of $\mathbb{E}[\ |\mathscr{G}]$ we get

$$\|\mathbb{E}[f|\mathscr{G}] - f\|_p = \|\mathbb{E}[f - g|\mathscr{G}] + g - f\|_p \leq \|\mathbb{E}[g - f|\mathscr{G}]\|_p + \|g - f\|_p \leq 2\|g - f\|_p,$$

and taking the supremum over all $g \in L^p(\mathscr{G})$ completes the proof. □

2.7. **Corollary.** *Under the assumption of Theorem 2.5, for all $f \in L^p(\mathscr{F}, \mu)$ there is convergence*

$$\mathbb{E}[f|\mathscr{F}_i] \to f \quad \text{in the sense of } L^p(\mathscr{F}, \mu)\text{-norm, when } i \to \infty.$$

Theorem 2.5 said that there exist good approximations for f in the spaces $L^p(\mathscr{F}_i, \mu)$; this corollary tells that conditional expectations provide a way of finding them explicitly.

Proof. Let $\varepsilon > 0$. By Theorem 2.5 there are $j \in \mathbb{Z}$ and $g \in L^p(\mathscr{F}_j, \mu) \subseteq L^p(\mathscr{F}_i, \mu)$ for all $i \geq j$, such that $\|f - g\|_p < \varepsilon$. Hence, by Lemma 2.6

$$\|\mathbb{E}[f|\mathscr{F}_i] - f\|_p \leq 2\|g - f\|_p < 2\varepsilon,$$

for $i \geq j$. □

2.8. **A question of pointwise convergence.** According to the general integration theory, a sequence of functions which converges in the L^p norm also has a subsequence converging pointwise a.e. In the situation of the previos corollary, one does not even need to restrict to a subsequence, but proving this fact requires a certain auxiliary device. Let us sketch the proof as far as we can at the present to see which estimate we are still lacking. First of all,

$$\{\mathbb{E}[f|\mathscr{F}_i] \not\to f\} = \big\{\limsup_{i\to\infty} |\mathbb{E}[f|\mathscr{F}_i] - f| > 0\big\} = \bigcup_{n=1}^{\infty} \big\{\limsup_{i\to\infty} |\mathbb{E}[f|\mathscr{F}_i] - f| > \frac{1}{n}\big\},$$

so it suffices to prove that for all $\varepsilon > 0$ there holds

$$\mu\big(\big\{\limsup_{i\to\infty} |\mathbb{E}[f|\mathscr{F}_i] - f| > \varepsilon\big\}\big) = 0.$$

Let $\delta > 0$, and let $j \in \mathbb{Z}$ and $g \in L^p(\mathscr{F}_j, \mu)$ be such that $\|f - g\|_p < \delta$. Then

$$|\mathbb{E}[f|\mathscr{F}_i] - f| \leq |\mathbb{E}[f - g|\mathscr{F}_i]| + |\mathbb{E}[g|\mathscr{F}_i] - g| + |g - f|.$$

Taking $\limsup_{i\to\infty}$ of both sides and observing that the middle term approaches zero (it even equal zero as soon as $i \geq j$), it follows that

$$\mu\big(\big\{\limsup_{i\to\infty} |\mathbb{E}[f|\mathscr{F}_i] - f| > 2\varepsilon\big\}\big) \leq \mu\big(\big\{\limsup_{i\to\infty} |\mathbb{E}[f - g|\mathscr{F}_i]| > \varepsilon\big\}\big) + \mu\big(\big\{|g - f| > \varepsilon\big\}\big).$$

The latter term satisfies the basic estimate

$$\mu\big(\big\{|g - f| > \varepsilon\big\}\big) \leq \varepsilon^{-p}\|g - f\|_p^p < (\delta/\varepsilon)^p,$$

which can be made arbitrarily small, since $\delta > 0$ can be chosen at will.

The remaining lim sup-term can be estimates by

$$\limsup_{i\to\infty} |\mathbb{E}[f - g|\mathscr{F}_i]| \leq \sup_{i\in\mathbb{Z}} \mathbb{E}[|f - g|\,|\mathscr{F}_i] =: M(|f - g|),$$

where the above defined (nonlinear) operator M is *Doob's maximal operator*. So there holds

$$\mu\big(\big\{\limsup_{i\to\infty} |\mathbb{E}[f - g|\mathscr{F}_i]| > \varepsilon\big\}\big) \leq \mu\big(\big\{M(|f - g|) > \varepsilon\big\}\big),$$

and we would need an inequality of the type $\mu\big(\big\{Mh > \varepsilon\big\}\big) \leq C\varepsilon^{-p}\|h\|_p^p$ to finish the estimate. This follows from *Doob's inequality* for the maximal function.

Let us first define the maximal function in a slightly more general setting:

2.9. **Doob's maximal function.** Let $(f_i)_{i\in\mathbb{Z}}$ be a sequence of functions adapted to a filtration $(\mathscr{F}_i)_{i\in\mathbb{Z}}$. Let us denote the whole sequence simply by f; hence $f = (f_i)_{i\in\mathbb{Z}}$ is not itself a function but a sequence of functions. Then its Doob's maximal function is defined pointwise by

$$Mf := f^* := \sup_{i\in\mathbb{Z}} |f_i|.$$

Observe that this notation is in agreement with the situation considered above, where $f \in L^1_\sigma(\mathscr{F}, \mu)$ is as function and $f_i = \mathbb{E}[f|\mathscr{F}_i]$.

2.10. **Theorem** (Doob's inequality). *Let $f = (f_i)_{i\in\mathbb{Z}}$ be a nonnegative ($f_i \geq 0$) submartingale with $\sup_{i\in\mathbb{Z}} \|f_i\|_p < \infty$, where $p \in [1, \infty]$.*

- *If $p = 1$, then for all $\lambda > 0$, we have*

$$\lambda \cdot \mu(f^* > \lambda) \leq \int_{\{f^*>\lambda\}} |f|\,\mathrm{d}\mu \leq \|f\|_1$$

- *If $p \in (1, \infty]$, then $f^* \in L^p(\mathscr{F}, \mu)$ and more precisely*

$$\|f^*\|_p \leq p' \cdot \sup_{i\in\mathbb{Z}} \|f_i\|_p.$$

The analogous results for martingales (even without the requirement that $f_i \geq 0$) follows at once, for if $(f_i)_{i\in\mathbb{Z}}$ is a martingale, then $(|f_i|)_{i\in\mathbb{Z}}$ fulfills the assumptions of the theorem. The constant p' in the inequality is the best possible (in the sense that the result does not hold in general if p' is replaced by any number $c < p'$) – this fact will be proven in the exercises.

Doob's inequality also has a so-called weak-type version for $p = 1$, but this will not be dealt with here.

2.11. **Preliminary considerations.** Before the actual proof of Doob's inequality, we make some simplifying considerations. First of all, notice that the case $p = \infty$ is trivial, so in the sequel we will concentrate on $p \in (1, \infty)$.

Observe that it suffices to prove the claim for submartingales $(f_i)_{i\in\mathbb{N}}$ indexed by the natural numbers. Namely, from this it follows (just by making a change of the index variable) that the estimate also holds for martingales with the index set $\{n, n+1, n+2, \dots\}$ with an arbitrary $n \in \mathbb{Z}$, i.e.,

$$\|\sup_{i\geq n} f_i\|_p \leq p' \cdot \sup_{i\geq n} \|f_i\|_p.$$

But clearly $\sup_{i\geq n} f_i \nearrow \sup_{i\in\mathbb{Z}} f_i$ as $n \to -\infty$, so the monotone convergence theorem (the usual form from integration theory) implies that

$$\|\sup_{i\in\mathbb{Z}} f_i\|_p = \lim_{n\to-\infty} \|\sup_{i\geq n} f_i\|_p \leq p' \cdot \lim_{i\to-\infty} \sup_{i\geq n} \|f_i\|_p = p' \cdot \sup_{i\in\mathbb{Z}} \|f_i\|_p.$$

Next observe that one can even restrict to finite submartingales $(f_i)_{i=0}^n$ with the index set $\{0, 1, \dots, n\}$. Passing from here to the case of all N can be realized by a similar monotone convergence argument as above. So it remains to prove that

$$\Big\| \max_{0\leq k\leq n} f_k \Big\|_p \leq p' \cdot \max_{0\leq k\leq n} \|f_k\|_p = p' \cdot \|f_n\|_p;$$

the equality above follows from the fact that $0 \leq f_k \leq \mathbb{E}[f_n|\mathscr{F}_k]$, and hence $\|f_k\|_p \leq \|f_n\|_p$ for all $k = 0, 1, \dots, n$.

The L^1 estimate can be reduced to the finite case by using the continuity of the measure in place of monotone convergence:

$$\{f^* > \lambda\} = \bigcup_{n\in\mathbb{Z}_-} \{\sup_{i\geq n} f_i > \lambda\} \qquad \Rightarrow \qquad \mu(f^* > \lambda) = \lim_{n\to-\infty} \mu(\sup_{i\geq n} f_i > \lambda)$$

and

$$\{\sup_{i\in\mathbb{N}} f_i > \lambda\} = \bigcup_{n\in\mathbb{N}} \{\sup_{0\leq i\leq n} f_i > \lambda\} \qquad \Rightarrow \qquad \mu(\sup_{i\in\mathbb{N}} f_i > \lambda) = \lim_{n\to\infty} \mu(\sup_{0\leq i\leq n} f_i > \lambda).$$

2.12. **Stopping times.** The proof of Doob's inequality makes use of a simple version of a powerful *stopping time argument.* A stopping time is defined as any function

$$\tau : \Omega \to \mathbb{Z} \cup \{-\infty, \infty\},$$

with the property that

$$\{\tau \leq k\} := \{\omega \in \Omega : \tau(\omega) \leq k\} \in \mathscr{F}_k \qquad \forall k \in \mathbb{Z}.$$

We will usually consider stopping times that do not take the value $-\infty$.

Note also that for any stopping time, we have

$$\{\tau = k\} = \{\tau \leq k\} \setminus \{\tau \leq k-1\} \in \mathscr{F}_k,$$

and

$$\{\tau \geq k\} = \{\tau \leq k-1\}^{\mathsf{C}} \in \mathscr{F}_{k-1}.$$

The idea behind a stopping time is the following: Think of $k \in \mathbb{Z}$ as the parameter, and $\mathscr{F}_k$ as all the information that we have at time k and all previous times. Let τ be the first time when something that we can observe happens. Thus, at time k, we know whether the event has happened by the time k. In other words, the event $\{\tau \leq k\}$ ("something that we were looking for has happened at the time k or earlier") belongs to $\mathscr{F}_k$ ("our complete knowledge at time k").

A typical example of a stopping time is obtained by

$$\tau := \inf\{k : f_k > \lambda\} \qquad (\inf\varnothing := \infty)$$

where $(f_k)_{k\in\mathbb{Z}}$ is an adapted sequence (i.e., f_k is $\mathscr{F}_k$-measurable). To see that this is a stopping time, note that for all $k \in \mathbb{Z}$, we have

$$\{\tau \le k\} = \bigcup_{j\le k}\{f_j > \lambda\},$$

where $\{f_j > \lambda\} \in \mathscr{F}_j \subseteq \mathscr{F}_k$, and thus also the union belongs to $\mathscr{F}_k$. Note also that

$$\{\tau = \infty\} = \{f^* \le \lambda\}.$$

In the proof of the finite version of Doob's inequality obtained above, we will consider the variant

$$\tau := \inf\{k \in \{0,1,\dots,n\} : f_k > \lambda\} \in \{0,1,\dots,n\} \cup \{\infty\}.$$

So, starting from 0, the time τ is the first index k such that $f_k > \lambda$, or $\tau = \infty$, if no such index exists.

2.13. **Proof of Doob's L^1 inequality.** Let us write $f_n^* := \max_{0\le k\le n} f_k$. We use the stopping time just defined:

$$\mu(f_n^* > \lambda) = \mu(\tau < \infty) = \sum_{k=0}^n \mu(\tau = k) = \sum_{k=0}^n \int_{\{\tau=k\}} \mathrm{d}\mu.$$

If $\tau = k$, then, by definition, $f_k > \lambda$, and thus

$$\int_{\{\tau=k\}} \mathrm{d}\mu \le \frac{1}{\lambda}\int_{\{\tau=k\}} f_k\,\mathrm{d}\mu = \frac{1}{\lambda}\int_{\{\tau=k\}} \mathbb{E}(f_n|\mathscr{F}_k)\,\mathrm{d}\mu = \frac{1}{\lambda}\int_{\{\tau=k\}} f_n\,\mathrm{d}\mu,$$

where we used the submartingale property and the definition of the conditional expectation, observing that $\{\tau = k\} \in \mathscr{F}_k$. Combining everything and summing up, we have

$$\begin{aligned}\mu(f_n^* > \lambda) &= \sum_{k=0}^n \int_{\{\tau=k\}} \mathrm{d}\mu \\ &\le \frac{1}{\lambda}\sum_{k=0}^n \int_{\{\tau=k\}} f_n\,\mathrm{d}\mu = \frac{1}{\lambda}\int_{\{\tau<\infty\}} f_n\,\mathrm{d}\mu = \frac{1}{\lambda}\int_{\{f_n^*>\lambda\}} f_n\,\mathrm{d}\mu \le \frac{1}{\lambda}\|f_n\|_1.\end{aligned}$$

which is precisely Doob's L^1 inequality in the finite case.

2.14. **Proof of Doob's L^p inequality.** We make use of the formula

$$\|f_n^*\|_p^p = \int_\Omega f_n^*(\omega)^p\,\mathrm{d}\mu(\omega) = \int_0^\infty p\lambda^{p-1}\mu(f_n^* > \lambda)\,\mathrm{d}\lambda$$

and the L^1 inequality that we already proved:

$$\begin{aligned}\int_0^\infty p\lambda^{p-1}\mu(f_n^* > \lambda)\,\mathrm{d}\lambda &\le \int_0^\infty p\lambda^{p-1}\frac{1}{\lambda}\int_{\{f_n^*>\lambda\}} f_n(\omega)\,\mathrm{d}\mu(\omega)\,\mathrm{d}\lambda \\ &= \int_0^\infty\int_\Omega p\lambda^{p-2}1_{\{(\omega,\lambda):f_n^*(\omega)>\lambda\}} f_n(\omega)\,\mathrm{d}\mu(\omega)\,\mathrm{d}\lambda \\ &= \int_\Omega\int_0^{f_n^*(\omega)} p\lambda^{p-2}\,\mathrm{d}\lambda f_n(\omega)\,\mathrm{d}\mu(\omega) \\ &= \int_\Omega \frac{p}{p-1} f_n^*(\omega)^{p-1} f_n(\omega)\,\mathrm{d}\mu(\omega) = p'\int_\Omega f_n(f_n^*)^{p-1}\,\mathrm{d}\mu.\end{aligned}$$

We apply Hölder's inequality:

$$\int_\Omega f_n(f_n^*)^{p-1}\,\mathrm{d}\mu \le \Big(\int_\Omega f_n^p\,\mathrm{d}\mu\Big)^{1/p}\Big(\int_\Omega (f_n^*)^{(p-1)p'}\,\mathrm{d}\mu\Big)^{1/p'} = \|f_n\|_p\|f_n^*\|_p^{p-1},$$

since $(p-1)p' = p$ and $1/p' = (p-1)/p$. So altogether, we have

$$\|f_n^*\|_p^p \leq p'\|f_n\|_p\|f_n^*\|_p^{p-1}.$$

If $\|f_n^*\|_p < \infty$, we just divide both sides by $\|f_n^*\|_p^{p-1}$, and get what we wanted.

To check the finiteness, observe that

$$f_n^* = \max_{0\leq k\leq n} f_k \leq \sum_{k=0}^{n} f_k,$$

thus

$$\|f_n^*\|_p \leq \sum_{k=0}^{n} \|f_k\|_p \leq \sum_{k=0}^{n} \|f_n\|_p = (1+n)\|f_n\|_p < \infty.$$

Note that the last, easy estimate is of the same form as the bound $\|f_n^*\|_p \leq p'\|f_n\|_p$ that we wanted, but with a bad constant $1+n$, the length of the martingale, instead of p'. It would be impossible to get anything useful for infinite martingales by passing to a limit in this inequality.

2.15. **Convergence of martingales to the reverse direction.** We have seen that if $(\mathscr{F}_j)_{j\in\mathbb{Z}}$ is a filtration with $\sigma(\bigcup_{j\in\mathbb{Z}}\mathscr{F}_j) = \mathscr{F}_\infty = \mathscr{F}$, then for all $f \in L^p(\mathscr{F},\mu)$, $p \in (1,\infty)$, there holds $\mathbb{E}[f|\mathscr{F}_j] \to f$ when $j \to \infty$, both in the L^p norm and pointwise a.e. What about $j \to -\infty$?

Let make the following additional assumption:

$$\forall\, F \in \mathscr{F}_{-\infty} := \bigcap_{j\in\mathbb{Z}} \mathscr{F}_j : \mu(F) \in \{0,\infty\}.$$

Then for all $f \in L^p(\mathscr{F},\mu)$, $p \in (1,\infty)$, there holds $\mathbb{E}[f|\mathscr{F}_j] \to 0$ when $j \to -\infty$, both in the L^p norm and pointwise a.e.

Proof. At a.e. point $\omega \in \Omega$, the sequence $(\mathbb{E}[f|\mathscr{F}_j])_{j\in\mathbb{Z}}$ is bounded from above and from below by the numbers Mf and $-Mf$. In particular, it has finite pointwise lim sup and lim inf as $j \to -\infty$; let the first one be denoted by g. A basic observation is that

$$g = \limsup_{j\to-\infty} \mathbb{E}[f|\mathscr{F}_j] = \limsup_{i\geq j\to-\infty} \mathbb{E}[f|\mathscr{F}_j]$$

can be computed by restricting to the tail $j \leq i$ for any fixed $i \in \mathbb{Z}$. In particular, as the upper limit of $\mathscr{F}_i$-measurable functions, g itself is $\mathscr{F}_i$-measurable. Since this is true for all $i \in \mathbb{Z}$, the function g is in fact $(\bigcap_{j\in\mathbb{Z}}\mathscr{F}_j)$-measurable.

In particular, for all $\varepsilon > 0$ there holds $\mu(\{|g| > \varepsilon\}) \in \{0,\infty\}$. The latter possibility cannot hold, since $|g| \leq Mf \in L^p(\mathscr{F},\mu)$, and hence $\mu(\{|g| > \varepsilon\}) \leq \varepsilon^{-p}\|Mf\|_p^p < \infty$. Thus $\mu(\{g \neq 0\}) = \mu\Big(\bigcup_{n=1}^\infty \{|g| > n^{-1}\}\Big) = 0$, and therefore $g = 0$ a.e.

A similar argument shows that also $\liminf_{j\to-\infty} \mathbb{E}[f|\mathscr{F}_j] = 0$, so in fact there exists the pointwise limit $\lim_{j\to-\infty} \mathbb{E}[f|\mathscr{F}_j] = 0$. By the dominated convergence theorem (the dominating function being Mf) it follows that the convergence also takes places in the L^p norm. □

By combining the convergence results of this section, the following representation of a function in terms of its *martingale differences* is obtained:

2.16. **Theorem.** *Let $(\Omega,\mathscr{F},\mu)$ be a measure space and $(\mathscr{F}_j)_{j\in\mathbb{Z}}$ its filtration, such that the spaces $(\Omega,\mathscr{F}_j,\mu)$ are σ-finite. Let, in addition,*

$$\sigma\Big(\bigcup_{j\in\mathbb{Z}} \mathscr{F}_j\Big) = \mathscr{F}, \qquad \forall\, F \in \bigcap_{j\in\mathbb{Z}} \mathscr{F}_j : \mu(F) \in \{0,\infty\}.$$

Then for all $f \in L^p(\mathscr{F},\mu)$, $p \in (1,\infty)$, there holds

$$f = \sum_{j=-\infty}^{\infty} \big(\mathbb{E}[f|\mathscr{F}_j] - \mathbb{E}[f|\mathscr{F}_{j-1}]\big),$$

where the convergence takes place both in the L^p-norm and pointwise a.e.

Proof. By using the obtained convergence results and writing out the difference as a telescopic sum, it follows that

$$f = f - 0 = \lim_{n\to+\infty} \mathbb{E}[f|\mathscr{F}_n] - \lim_{m\to-\infty} \mathbb{E}[f|\mathscr{F}_m] = \lim_{\substack{n\to+\infty\\ m\to-\infty}} \sum_{j=m+1}^{n} \big(\mathbb{E}[f|\mathscr{F}_j] - \mathbb{E}[f|\mathscr{F}_{j-1}]\big),$$

and the existence of the limit on the right is, by definition, the same as the convergence of the series in the assertion. □

2.17. **Exercises.** Many of the exercises deal with applications of martingale theory and especially Doob's maximal inequality to classical analysis.

1. Prove that a filtration indexed by a subset $I \subset \mathbb{Z}$ of the integers $\mathbb{Z}$ and a martingale adapted to it can be extended so as to be indexed by all of $\mathbb{Z}$. More precisely: Let $(\mathscr{F}_i)_{i\in I}$ be a filtration and $(f_i)_{i\in I}$ a martingale adapted to it, where $I \subset \mathbb{Z}$. Define $\mathscr{F}_i$ and f_i for $i \in \mathbb{Z}\setminus I$ in such a way that also $(\mathscr{F}_i)_{i\in\mathbb{Z}}$ is a filtration and $(f_i)_{i\in\mathbb{Z}}$ a martingale adapted to it.
2. With the help of Doob's inequality, derive *Hardy's inequality*: for all $0 \leq f \in L^p(\mathbb{R}_+)$ (where $\mathbb{R}_+ = (0,\infty)$ is equipped with the Borel σ-algebra and the Lebesgue measure)
$$\Big[\int_0^\infty \Big(\frac{1}{x}\int_0^x f(y)\,\mathrm{d}y\Big)^p \mathrm{d}x\Big]^{1/p} \leq p'\Big[\int_0^\infty f(x)^p\,\mathrm{d}x\Big]^{1/p}.$$
(Hint: For a fixed $\delta > 0$, consider the filtration $(\mathscr{F}_n)_{n\in\mathbb{Z}_-}$, where
$$\mathscr{F}_n := \sigma\big(\{(0,|n|\delta],(k\delta,(k+1)\delta] : |n| \leq k \in \mathbb{Z}\}\big).$$
Take the limit $\delta \searrow 0$ in the end.)
Notice that it is possible (and not particularly hard) to prove Hardy's inequality also by other methods, but the point of the exercise is nevertheless to derive it as a corollary of Doob's inequality.
3. Show that the constant p' is optimal in Hardy's inequality, and hence also in Doob's inequality. (Hint: investigate e.g. the functions $f(x) = 1_I(x)\cdot x^\alpha$, where $I \subset \mathbb{R}_+$ is an appropriate subinterval and $\alpha \in \mathbb{R}$.)
4. Denote the collections of the usual *dyadic intervals* of $\mathbb{R}$ by $\mathscr{D}_k := \{2^{-k}[j,j+1) : j \in \mathbb{Z}\}$, where $k \in \mathbb{Z}$. For all $\beta = (\beta_k)_{k\in\mathbb{Z}} \in \{0,1\}^{\mathbb{Z}}$, define the collection of shifted dyadic intervals by
$$\mathscr{D}_k^\beta := \mathscr{D}_k + \sum_{j>k}\beta_j 2^{-j} := \Big\{I + \sum_{j>k}\beta_j 2^{-j} : I \in \mathscr{D}_k\Big\},$$
where $I + c := [a+c, b+c)$ if $I = [a,b)$. Note that $\mathscr{D}_k^0 = \mathscr{D}_k$, where 0 stands for the zero sequence. Denote the corresponding σ-algebras by $\mathscr{F}_k^\beta := \sigma(\mathscr{D}_k^\beta)$. Show that $(\mathscr{F}_k^\beta)_{k\in\mathbb{Z}}$ is a filtration for all $\beta \in \{0,1\}^{\mathbb{Z}}$.
5. Keeping the notations of the previous exercise, define the collection of all (shifted) dyadic intervals $\mathscr{D}^\beta := \bigcup_{k\in\mathbb{Z}}\mathscr{D}_k^\beta$. Consider the particular sequence $\beta \in \{0,1\}^{\mathbb{Z}}$, where $\beta_k = 0$ if k is even and $\beta_k = 1$ if k is odd. Prove that, with some constant $C \in (0,\infty)$, the following assertion holds: If $J \subset \mathbb{R}$ is any finite subinterval, there exists either $I \in \mathscr{D}^0$ or $I \in \mathscr{D}^\beta$, such that $J \subseteq I$ and $|I| \leq C|J|$. (Hint: it might help to sketch a picture.)
6. For $f \in L^1_{\mathrm{loc}}(\mathbb{R})$, its *Hardy–Littlewood maximal function* is defined by
$$M_{HL}f(x) := \sup_{I\ni x}\frac{1}{|I|}\int_I |f(y)|\,\mathrm{d}y,$$
where the supremum is over all finite subintervals $I \subset \mathbb{R}$ which contain x. (Usually this is denoted simply by M but now the subscript HL is used to distinguish this from Doob's maximal function.) Use Doob's inequality to derive the Hardy–Littlewood maximal inequality
$$\|M_{HL}f\|_p \leq C_p\|f\|_p, \qquad p \in (1,\infty].$$
(Hint: Use the result of the previous exercise to show that $M_{HL}f$ is pointwise dominated by the sum of two Doob's maximal functions related to different filtrations.)

7. For a sequence of functions $\vec{f} = (f_k)_{k\in\mathbb{Z}}$ (not assumed to be adapted), consider the norm
$$\|\vec{f}\|_{L^p(\ell^q)} := \Big\| \Big(\sum_{k\in\mathbb{Z}} |f_k|^q \Big)^{1/q} \Big\|_{L^p}.$$
Recall how to prove that $\|f\|_{L^p} = \sup\{\int fg \,\mathrm{d}\mu : \|g\|_{L^{p'}} \leq 1\}$, and use the same ideas to show that
$$\|\vec{f}\|_{L^p(\ell^q)} = \sup\Big\{ \int \sum_{k\in\mathbb{Z}} f_k g_k \,\mathrm{d}\mu : \|\vec{g}\|_{L^{p'}(\ell^{q'})} \leq 1 \Big\}.$$
8. Prove that Doob's inequality (in $L^{p'}$) is equivalent to the following estimate:
$$(*) \qquad \Big\| \sum_{k\in\mathbb{Z}} \mathbb{E}(f_k|\mathscr{F}_k) \Big\|_p \leq p \Big\| \sum_{k\in\mathbb{Z}} f_k \Big\|_p, \qquad p \in [1,\infty).$$
for all nonnegative sequences of functions $f_k \in L^p(\mathscr{F})$. That is, derive $(*)$ from Doob's inequality, and conversely, give a new proof of Doob's inequality by assuming $(*)$. (Hint: use Exercise 7. You can use the result 'as a black box' even if you did not do that Exercise.)
9. Prove the estimate $(*)$ directly (without using Doob's inequality) for $p = 2$. (Hint: multiply out the square in the L^2 norm.)

2.18. **References.** Doob's inequality appeared for the first time in Doob's classic book [4]. We have followed the ideas of the original proof, which is also commonly done in most more recent books on the topic (e.g. [16]). A completely different proof is found in Burkholder's summer school lectures [3] and in the previous edition of these lecture notes from 2008. The idea for Exercises 2 and 3 is also from Burkholder [3], and Exercises 5 and 6 from Mei [10].

3. Martingale transforms and Burkholder's inequality

3.1. **Martingale differences.** Let us for simplicity consider a finite martingale $(f_k)_{k=0}^n$ adapted to a filtration $(\mathscr{F}_k)_{k=0}^n$. We also define $f_{-1} := 0$. The difference sequence of f is given by
$$d_k := f_k - f_{k-1},$$
and it has the two following properties:

- $d_k \in L^0(\mathscr{F}_k)$; indeed, $f_k \in L^0(\mathscr{F}_k)$ and $f_{k-1} \in L^0(\mathscr{F}_{k-1}) \subseteq L^0(\mathscr{F}_k)$.
- $\mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$ for $k = 1,\ldots,n$; indeed,
$$\mathbb{E}[d_k|\mathscr{F}_{k-1}] = \mathbb{E}[f_k|\mathscr{F}_{k-1}] - \mathbb{E}[f_{k-1}|\mathscr{F}_{k-1}] = f_{k-1} - f_{k-1} = 0.$$

Let us call any sequence $(d_k)_{k=0}^n$ with these properties a *martingale difference sequence.*

3.2. **Lemma.** *There is a one-to-one correspondence between martingales $(f_k)_{k=0}^n$ and martingale difference sequences $(d_k)_{k=0}^n$, given by*
$$d_k = f_k - f_{k-1}, \qquad f_k = \sum_{j=0}^n d_j.$$

Proof. We already saw that a martingale defined a martingale difference sequence. Let then a difference sequence d_k be given, and define f_k by the above formula. Since $d_j \in L^0(\mathscr{F}_j) \subseteq L^0(\mathscr{F}_k)$ for $j \leq k$, we see that $f_k \in L^0(\mathscr{F}_k)$. To check the martingale property, observe that
$$\mathbb{E}[f_k|\mathscr{F}_{k-1}] = \sum_{j=0}^{k-1} \mathbb{E}[d_j|\mathscr{F}_{k-1}] + \mathbb{E}[d_k|\mathscr{F}_{k-1}] = \sum_{j=0}^{k-1} d_j + 0 = f_{k-1},$$
using again that $d_j \in L^0(\mathscr{F}_j) \subseteq L^0(\mathscr{F}_{k-1})$ and the martingale difference property to the last term. □

3.3. **Martingale transform.** A sequence $(v_k)_{k=0}^n$ is called *predictable* (or previsible) if $v_0 \in L^0(\mathscr{F}_0)$ and $v_k \in L^0(\mathscr{F}_{k-1})$ for $k \geq 1$.

The transform of a martingale f by a predictable sequence v is the sequence $v*f = ((v*f)_k)_{k=0}^n$ defined by

$$(v * f)_k := \sum_{j=0}^{k} v_j d_j,$$

where d_j is the difference sequence of f. We immediately check that

$$\mathbb{E}[v_j d_j | \mathscr{F}_{j-1}] = v_j \mathbb{E}[d_j | \mathscr{F}_{j-1}] = v_j \cdot 0 = 0,$$

so that $v_j d_j$ is again a martingale difference sequence, and therefore $v * f$ is a martingale. This justifies the name 'martingale transform'.

Possible interpretation in applications: A martingale is a model of a 'fair gamble' between two players: f_k represents your accumulated winnings at time k, so that $d_k = f_k - f_{k-1}$ represents your winning (and thus your opponents loss) on the kth round of the game. Since the game is fair, the expectation of your kth-rounf winning, conditioned on the history $\mathscr{F}_{k-1}$ up till time $k-1$, is $\mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$.

Now suppose you can increase or decrease your stakes at the game, if you believe that you are more or less likely to win on the next round. So your winning or loss will now be $v_k d_k$, where v_k is your chosen stake, and d_k is the winning with a unit stake. But again, by fairness, you need to decide about your stakes for the kth round before the kth round is played, i.e., only based on the information $\mathscr{F}_{k-1}$ that you have at time $k-1$. This means that v_k must be predictable, and thus the transformed game $v * f$ is still a martingale, i.e., a fair game.

3.4. **Burkholder's inequality for martingale transforms.** On a mathematical level, the following is a basic estimate about the behaviour of martingale transforms in L^p spaces:

3.5. **Theorem** (Burkholder). *Let $(f_k)_{k=0}^n$ be an L^p-martingale, where $p \in (1, \infty)$, and $(v_k)_{k=0}^n$ a bounded predictable sequence with $\|v_k\|_\infty \leq 1$. Then*

$$\|(v * f)_n\|_p \leq \beta_p \|f_n\|_p$$

for some finite constant β_p that depends only on p.

It is known that the best (smallest) constant in this inequality is $\beta_p = \max\{p-1, 1/(p-1)\}$. It is quite remarkable that this constant is known, since it is usually very difficult to find the exact constants in more complicated inequalities of Analysis. In fact, Burkholder's inequality is one of the most important tools for finding the optimal constants also for many other inequalities, by showing that other operations can be interpreted as martingale transforms.

We will now give a proof of this inequality, which does not provide the optimal constant.

3.6. **General observations.** Burkholder's inequality is actually a bound for the norm of some linear operators acting on $f_n \in L^p(\mathscr{F}_n)$. Namely, observe that

$$\begin{aligned}(v * f)_n = \sum_{k=0}^{n} v_k d_k &= v_0 f_0 + \sum_{k=1}^{n} v_k (f_k - f_{k-1}) \\ &= v_0 \mathbb{E}[f_n|\mathscr{F}_0] + \sum_{k=1}^{n} v_k (\mathbb{E}[f_n|\mathscr{F}_k] - \mathbb{E}[f_n|\mathscr{F}_{k-1}]) = T_v(f_n),\end{aligned}$$

where it is easy to check that T_v is linear. This places at our disposal some tools from the general theory of linear operators on L^p spaces:

Interpolation. The *Marcinkiewicz interpolation theorem* says (in particular, and even slightly more) that if a linear operator T is bounded on L^{p_0} and on L^{p_1} for two exponents $p_0 < p_1$, then it is also bounded on L^p for all intermediate exponents $p \in (p_0, p_1)$.

Duality. It is not difficult to check that T_v is seld-adjoint in the sense that

$$\int T_v f \cdot g \,\mathrm{d}\mu = \int f \cdot T_v g \,\mathrm{d}\mu.$$

Thus, if Burkholder's inequality holds in L^p, then we also find for the dual exponent p' that

$$\begin{aligned}\|T_v f\|_{p'} &= \sup\Big\{\int T_v f \cdot g \,\mathrm{d}\mu : \|g\|_p \le 1\Big\} = \sup\Big\{\int f \cdot T_v g \,\mathrm{d}\mu : \|g\|_p \le 1\Big\} \\ &\le \sup\Big\{\|f\|_{p'}\|T_v g\|_p \,\mathrm{d}\mu : \|g\|_p \le 1\Big\} \le \sup\Big\{\|f\|_{p'}\beta_p\|g\|_p \,\mathrm{d}\mu : \|g\|_p \le 1\Big\} = \beta_p\|f\|_{p'},\end{aligned}$$

and thus $\beta_p \le \beta_{p'}$. By reversing the roles of p and p' we see that also $\beta_{p'} \le \beta_p$, so in fact $\beta_p = \beta_{p'}$.

3.7. **Strategy for the proof.** A combination of interpolation and duality shows that it is enough to prove Burkholder's inequality, for example, just for exponents of the form $p = 2^k$, $k = 1, 2, \ldots$. Indeed, interpolation shows that we then have the estimate for all $p \in [2, \infty)$, and duality shows that we have it for all $p \in (1, 2]$.

The case $p = 2$ (with $\beta_2 = 1$) follows from simple orthogonality considerations, and is left as an exercise. The other powers of 2 will be obtained by induction from the following:

3.8. **Proposition.** *Suppose that Burkholder's inequality holds for some $p \in (1, \infty)$. Then it also holds for the exponent $2p$.*

We start the proof by observing that

$$\|g\|_{2p}^2 = \|g^2\|_p.$$

We apply this to $g = (v * f)_n$, and expand

$$(v*f)_n^2 = \Big(\sum_{k=0}^n v_k d_k\Big)^2 = \sum_{k=0}^n v_k^2 d_k^2 + 2\sum_{k=0}^n\sum_{j=0}^{k-1} v_j d_j v_k d_k = \sum_{k=0}^n v_k^2 d_k^2 + 2\sum_{k=0}^n (v*f)_{k-1} v_k d_k,$$

thus

$$\|(v*f)_n^2\|_p \le \Big\|\sum_{k=0}^n v_k^2 d_k^2\Big\|_p + 2\Big\|\sum_{k=0}^n (v*f)_{k-1} v_k d_k\Big\|_p.$$

We will next show that:

3.9. **Lemma.**

$$\Big\|\sum_{k=0}^n (v*f)_{k-1} v_k d_k\Big\|_p \le 8\beta_p \|(v*f)_n\|_{2p}\|f_n\|_{2p}.$$

Proof. The idea is to apply the induction hypothesis, observing that $r_k := (v*f)_{k-1}v_k$ is again predictable. The problem is that it is not necessarily bounded. To overcome this, we first write

$$(v*f)_{k-1}v_k = \frac{(v*f)_{k-1}v_k}{(v*f)_{k-1}^*}\cdot (v*f)_{k-1}^* =: u_k \cdot (v*f)_{k-1}^*, \qquad (v*f)_{k-1} := \max_{j\le k-1}|(v*f)_j|.$$

where u_k is both predictable and bounded by one, and $(v*f)_{k-1}^*$ is increasing in k. We observe that $((v^*f)_{k-1}^* d_k)_{k=1}^n$ is also a martingale difference sequence, so that we can apply the assumed Burkholder inequality to the result that

$$\Big\|\sum_{k=0}^n (v*f)_{k-1} v_k d_k\Big\|_p = \Big\|\sum_{k=0}^n u_k (v*f)_{k-1}^* d_k\Big\|_p \le \beta_p \Big\|\sum_{k=0}^n (v*f)_{k-1}^* d_k\Big\|_p.$$

We then investigate the function on the right pointwise, using a partial summation argument:

$$\begin{aligned}\sum_{k=0}^n (v*f)_{k-1}^* d_k &= \sum_{k=0}^n (v*f)_{k-1}^*(f_k - f_{k-1}) = \sum_{k=0}^n (v*f)_{k-1}^* f_k - \sum_{k=-1}^{n-1} (v*f)_k^* f_k \\ &= (v*f)_{n-1}^* f_n + \sum_{k=0}^{n-1}((v*f)_{k-1}^* - (v*f)_k^*) f_k - (v*f)_{-1}^* f_0,\end{aligned}$$

where the last term is actually zero by our convention about the martingale at -1. We estimate the absolute values, recalling that $(v^*f)^*_k$ is increasing in k, to find that

$$\begin{aligned}\Big|\sum_{k=0}^{n}(v*f)^*_{k-1}d_k\Big| &\leq (v*f)^*_{n-1}|f_n| + \sum_{k=0}^{n-1}((v*f)^*_k - (v^*f)^*_{k-1})|f_k| \\ &\leq (v*f)^*_{n-1}f^*_n + \sum_{k=0}^{n-1}((v*f)^*_k - (v^*f)^*_{k-1})f^*_n \\ &\leq (v*f)^*_{n-1}f^*_n + (v*f)^*_{n-1}f^*_n \leq 2(v*f)^*_n f^*_n,\end{aligned}$$

where we collapsed the telescopic summation. The L^p norm of the right side is now easy to estimate:

$$\|2(v*f)^*_n f^*_n\|_p \leq 2\|(v^*f)^*_n\|_{2p}\|f^*_n\|_{2p} \leq 2((2p)')^2\|(v*f)_n\|_{2p}\|f_n\|_{2p} \leq 8\|(v*f)_n\|_{2p}\|f_n\|_{2p},$$

where we used Doob's inequality, and the fact that $2p \geq 2$, hence $(2p)' \leq 2$. □

We return to the estimation of $\|(v*f)^2_n\|_p$. So far we showed that

$$\|(v*f)^2_n\|_p \leq \Big\|\sum_{k=0}^{n} v_k^2 d_k^2\Big\|_p + 2\cdot 8\beta_p\|(v*f)_n\|_p\|f_n\|_p.$$

To bound the first term, we first recall that $|v_k| \leq 1$, and then follow some of the earlier steps in the opposite order, writing

$$\sum_{k=0}^{n} v_k^2 d_k^2 \leq \sum_{k=0}^{n} d_k^2 = \Big(\sum_{k=0}^{n} d_k\Big)^2 - 2\sum_{k=0}^{n}\sum_{j=0}^{k-1} d_j d_k = f_n^2 - 2\sum_{k=0}^{n} f_{k-1}d_k.$$

Thus

$$\Big\|\sum_{k=0}^{n} v_k^2 d_k^2\Big\|_p \leq \|f_n^2\|_p + 2\Big\|\sum_{k=0}^{n} f_{k-1}d_k\Big\|_p,$$

where $\|f_n^2\|_p = \|f_n\|^2_{2p}$. The second term is exactly as in Lemma 3.9, but with $v_k \equiv 1$. Thus an application of that Lemma shows that

$$\Big\|\sum_{k=0}^{n} f_{k-1}d_k\Big\|_p \leq 8\beta_p\|f_n\|^2_{2p}.$$

Altogether, we have proven that

$$\|(v*f)_n\|^2_{2p} = \|(v*f)^2_n\|_p \leq \|f_n\|^2_{2p} + 16\beta_p\|f_n\|^2_{2p} + 16\beta_p\|(v*f)_n\|_{2p}\|f_n\|_{2p}.$$

Let us divide both sides by $\|f_n\|^2_{2p}$, and denote

$$X := \frac{\|(v*f)_n\|_{2p}}{\|f_n\|_{2p}}.$$

This gives

$$X^2 \leq 1 + 16\beta_p + 16\beta_p X,$$

or equivalently

$$(X - 8\beta_p)^2 = X^2 - 2\cdot 8\beta_p X + (8\beta_p)^2 \leq 1 + 2\cdot 8\beta_p + (8\beta_p)^2 = (1+8\beta_p)^2.$$

Thus $X \leq 1 + 16\beta_p$, and hence $\beta_{2p} \leq 1 + 16\beta_p$. This completes the proof of Proposition 3.8.

3.10. **Exercises.**

1. Give a short proof of Burkholder's inequality for $p = 2$ (Hint: orthogonality of martingale differences in L^2.)
2. Prove the following *Stein's inequality*: For any sequence of functions $f_k \in L^p(\mathscr{F})$, where $p \in (1, \infty)$, we have
$$\|(\mathbb{E}[f_k|\mathscr{F}_k])_{k\in\mathbb{Z}}\|_{L^p(\ell^2)} \leq C_p \|(f_k)_{k\in\mathbb{Z}}\|_{L^p(\ell^2)}.$$
(Hint: for $p \geq 2$, use duality in $L^{p/2}$ and Doob's inequality. For $p < 2$, use the previous case and duality in $L^p(\ell^2)$.)
3. The square function of a martingale f with difference sequence $(d_k)_{k=0}^n$ is defined by
$$S_n f := \Big(\sum_{k=0}^n d_k^2\Big)^{1/2}.$$
Check that $\|S_n f\|_2 = \|f_n\|_2$, and show that if $\|f_n\|_p \leq c_p\|S_n f\|_p$ for some $p \in (1, \infty)$, then we have both inequalities
$$\frac{1}{c_{2p}}\|f_n\|_{2p} \leq \|S_n f\|_{2p} \leq c_{2p}\|f_n\|_{2p}.$$
(Hint: Similar ideas as in Proposition 3.8; write $f_n^2 - (S_n f)^2$ as a sum of martingale differences to prove that $\|f_n^2 - (S_n f)^2\|_p \leq C_p\|f_n\|_{2p}\|S_n f\|_{2p}$.)
4. Complete the proof that
$$\frac{1}{c_p}\|f_n\|_p \leq \|S_n f\|_p \leq c_p\|f_n\|_p \tag{3.11}$$
for all $p \in (1, \infty)$ by following these hints: From Exercise 3, you get that this double estimate holds for all $p = 2^n$, $n = 1, 2, \ldots$. Then, show that if the left inequality holds for some p, then the right inequality holds for p', and the other way round. (Here you might need Stein's inequality.) Thus you also obtain the double estimate for all p such that $p' = 2^n$, $n = 1, 2, \ldots$. In particular, you now have that $f_n \mapsto S_n f$ is bounded (i.e., $\|S_n f\|_p \leq c_p\|f_n\|_p$) for all p with $\max\{p, p'\} = 2^n$, $n = 1, 2, \ldots$. Observe that $f_n \mapsto S_n f$ is a *quasi-linear* operator: $S_n(f + g) \leq S_n f + S_n g$ for any martingales f, g. Thus the Marcinkiewicz interpolation theorem shows that the L^p-boundedness of S_n (i.e., the right inequality $\|S_n f\|_p \leq c_p\|f_n\|_p$) extends to all intermediate values $p \in (1, \infty)$. (You can take this interpolation argument for granted as a black box.) Finally, from the duality already mentioned, you get that also the left estimate $\|f_n\|_p \leq c_n\|S_n f\|_p$ holds for all $p \in (1, \infty)$.
5. Derive Burkholder's inequality from the double inequality (3.11).

3.12. **References.** Burkholder's inequality was originally proved in [1]. (It is only contained as part of the proof of Theorem 9 in [1], not as a separate result!) The optimal constant was obtained in [2]. A more detailed explanation of how Burkholder actually found this amazing result is given in his summer school lectures [3].

4. Up-crossings and convergence of martingales

We have seen that if $(f_n)_{n\in\mathbb{Z}}$ is the martingale generated by a function $f \in L^p(\mathscr{F}_\infty)$, $p \in [1, \infty)$, then $f_n \to f$ both in L^p and almost everywhere. We now want to look at the following related questions:

- What can be said about the convergence of f_n as $n \to \infty$ if we are only given a martingale $(f_n)_{n\in\mathbb{Z}}$, not necessarily generated by a function as above?
- What about the convergence of f_n as $n \to -\infty$?
- Assuming that we have positive answers to the first two questions, with limits f_∞ and $f_{-\infty}$, what more can be said about the mode of convergence of the series

$$\sum_{k=-\infty}^{\infty} d_k = \lim_{\substack{n\to+\infty\\ m\to-\infty}} \sum_{k=m+1}^{n} d_k = \lim_{\substack{n\to+\infty\\ m\to-\infty}} (f_n - f_m) = f_\infty - f_{-\infty}\ ? \tag{4.1}$$

In the last point, we have in mind the so-called *unconditional convergence* in the space L^p. This is defined as follows in a general Banach space (a complete normed space) X.

4.2. **Unconditional convergence.** A series $\sum_{k=1}^\infty x_k$, with $x_k \in X$, is said to converge to $x \in X$ if $\lim_{n\to\infty}\sum_{k=1}^n x_k = x$. It is said to converge unconditionally if

$$\sum_{k=1}^\infty x_{\sigma(k)}$$

converges for every permutation (bijection) σ of the positive integers $\mathbb{Z}_+$.

We will discuss the unconditional convergence of martingale differences at the end of the chapter.

4.3. **Up-crossings of a martingale.** This is our main tool for the analysis of martingale convergence. Fix some interval (a,b) and a martingale $(f_n)_{n\in\mathbb{Z}}$. By an up-crossing we mean a pair of indices (σ,τ) with such that $f_\sigma < a < b < f_\tau$, and $\tau > \sigma$ is the minimal number with this property.

Let

$$u_{m,n}(\omega) := \text{number of up-crossings of } (f_k(\omega))_{k=m}^n.$$

To be more formal, let $\sigma_0 := \tau_0 := m$, and then

$$\sigma_j := \min\{k \geq \tau_{j-1} : f_k < a\}, \qquad \tau_j := \min\{k \geq \sigma_j : f_k > b\}, \qquad j \geq 1,$$

and

$$u_{m,n} := \max\{J : \tau_J \leq n\}.$$

Note that both $\sigma_j, \tau_j : \Omega \to \{m, m+1, \ldots, \infty\}$ are stopping times. (Recall that $\varrho : \Omega \to \mathbb{Z}\cup\{-\infty,+\infty\}$ is a stopping time if $\{\varrho \leq k\} \in \mathscr{F}_k$ for all $k \in \mathbb{Z}$.) Indeed, if we already know this for τ_{j-1}, we observe that

$$\begin{aligned}\{\sigma_j \leq k\} &= \{\tau_{j-1} \leq k, \min_{i:\tau_{j-1}\leq i\leq k} f_i < a\} \\ &= \bigcup_{h=m}^k \{\tau_{j-1} = h\} \cap \{\min_{i:h\leq i\leq k} f_i < a\},\end{aligned}$$

which belongs to $\mathscr{F}_k$, since $\{\tau_{j-1} = h\} \in \mathscr{F}_h \subseteq \mathscr{F}_k$, and each f_i above belongs to $L^0(\mathscr{F}_i) \subseteq L^0(\mathscr{F}_k)$. Hence also σ_j is a stopping time. From this we similarly obtain that τ_j is a stopping time, and we can proceed by induction.

4.4. **A martingale transform that counts the up-crossings.** We now define a martingale transform of $(f_k)_{k\in\mathbb{Z}}$ in such a way that we only want to consider those differences d_k that take place in an up-crossing, i.e., $\sigma_j < k \leq \tau_j$ for some j. Namely, we set

$$v_k := \sum_{j=1}^\infty 1_{\{\sigma_j < k \leq \tau_j\}},$$

which is predictable, since

$$\{\sigma_j < k \leq \tau_j\} = \{\sigma_j \leq k-1\} \cap \{\tau_j > k-1\} = \{\sigma_j \leq k-1\} \cap \{\tau_j \leq k-1\}^{\complement} \in \mathscr{F}_{k-1}$$

by the fact that σ_j and τ_j are stoppping times, and also $v_k(\omega) \in \{0,1\}$, since at most one of the conditions $\sigma_j(\omega) < k \leq \tau_j(\omega)$ holds at any given point $\omega \in \Omega$.

Now consider the transform of f by v given by

$$g_n := (v * f)_n := \sum_{k=m+1}^n v_k d_k, \qquad d_k := f_k - f_{k-1}.$$

From the definition of v_k, we obtain

$$g_n = \sum_{j=1}^\infty \sum_{k:\sigma_j < k \leq \min(\tau_j, n)} d_k = \sum_{j=1}^J (f_{\tau_j} - f_{\sigma_j}) + \begin{cases} f_n - f_{\sigma_{J+1}}, & \text{if } n > \sigma_{J+1}, \\ 0, & \text{else,}\end{cases}$$

where $J = u_{m,n}$ is the number of up-crossings between m and n.

By definition, we have $f_{\tau_j} > b$ and $f_{\sigma_j} < a$; hence

$$\sum_{j=1}^{J}(f_{\tau_j} - f_{\sigma_j}) \geq \sum_{j=1}^{J}(b-a) = J(b-a) = (b-a)u_{m,n}.$$

Also, it is clear that $f_n - f_{\sigma_{J+1}} \geq -|f_n| - a$, and thus altogether we checked that

$$g_n \geq (b-a)u_{m,n} - |f_n| - |a|. \tag{4.5}$$

4.6. **Integrating the up-crossings, $n \to \infty$.** Next, we want to integrate the pointwise bound (4.5) over suitable sets A. For this purpose, we impose the following integrability condition: The ideal

$$(\mathscr{F}_m^0)_{f_k;k\geq m} := \{A \in \mathscr{F}_m^0 : \sup_{k\geq m} \|1_A f_k\|_1 < \infty\}$$

should contain a countable cover. For example, this holds if $\sup_{k\in\mathbb{Z}} \|f_k\|_p < \infty$ for some $p \in [1,\infty]$, since then $\|1_A f_k\|_1 \leq \mu(A)^{1/p'}\|f_k\|_p$, and thus $(\mathscr{F}_m^0)_{f_k;k\geq m} = \mathscr{F}_m^0$, which contains a countable cover by σ-finiteness.

Note that each term of

$$g_n = \sum_{k=m+1}^{n} v_k(f_k - f_{k-1})$$

is integrable over any $A \in (\mathscr{F}_m^0)_{f_k;k\geq m}$ (since the functions f_k are, and $\|v_k\|_\infty \leq 1$), and we get

$$\begin{aligned}\int_A g_n \,\mathrm{d}\mu &= \sum_{k=m+1}^{n} \int_A v_k d_k \,\mathrm{d}\mu \\ &= \sum_{k=m+1}^{n} \int_A \mathbb{E}[v_k d_k|\mathscr{F}_{k-1}] \,\mathrm{d}\mu = \sum_{k=m+1}^{n} \int_A v_k \mathbb{E}[d_k|\mathscr{F}_{k-1}] \,\mathrm{d}\mu = 0,\end{aligned}$$

where we used the facts that $A \in \mathscr{F}_m \subseteq \mathscr{F}_{k-1}$, $v_k \in L^0(\mathscr{F}_{k-1})$, and $\mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$ by the martingale difference property. So integrating (4.5) over $A \in (\mathscr{F}_m)_{f_k;k\geq m}$, we find that

$$(b-a)\int_A u_{m,n} \,\mathrm{d}\mu \leq \int_A (|f_n| + |a|) \,\mathrm{d}\mu.$$

We now let $n \to \infty$. Since $u_{m,n}$ is increasing in n, the monotone convergence theorem shows that $u_{m,\infty} := \lim_{n\to\infty} u_{m,n}$ satisfies

$$\int_A u_{m,\infty} \,\mathrm{d}\mu \leq \frac{1}{b-a} \sup_{n\to\infty} \int_A (|f_n| + |a|) \,\mathrm{d}\mu < \infty \qquad \forall A \in (\mathscr{F}_m^0)_{f_k;k\geq m}.$$

But this means that $u_{m,\infty} < \infty$ almost everywhere on $A \in (\mathscr{F}_m^0)_{f_k;k\geq m}$, and since we can cover Ω by countably many such sets, we find that $u_{m,\infty} < \infty$ almost everywhere on Ω.

4.7. **Pointwise convergence of martingales as $n \to \infty$.** Let $(f_k)_{k\geq m}$ be a martingale for which $(\mathscr{F}_m)_{f_k;k\geq m}$ defined above contains a countable cover. We claim that in this case

- f_n converges to a limit (a priori, maybe $\pm\infty$) almost everywhere as $n \to \infty$, and
- this limit is actually finite (i.e., a real number) almost everywhere.

The key observation is that the upper and lower limits $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ (allowing the values $\pm\infty$) always exist, and the limit exists if and only if the upper and lower limits are equal. If they are not equal, then we can always squeeze to rational numbers $a < b$ in between. Thus,

$$\{\lim_{n\to\infty} f_n \text{ does not exist}\} = \{\liminf_{n\to\infty} f_n < \limsup_{n\to\infty} f_n\} = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \{\liminf_{n\to\infty} f_n < a < b < \limsup_{n\to\infty} f_n\}.$$

But, if $\limsup_{n\to\infty} f_n > b$, it means that there are infinitely many $k \geq m$ such that $f_k > b$, and similarly $\liminf_{n\to\infty} f_n < a$ means that $f_k < a$ for infinitely many $k \geq m$. So, if both conditions hold, it means that there are infinitely many up-crossings of (a,b) between m and ∞, i.e., $u_{m,\infty}=\infty$. But this can only happen in a set of measure zero, and therefore

$$\mu(\lim_{n\to\infty} f_n \text{ does not exist}) \leq \sum_{\substack{a,b\in\mathbb{Q}\\ a<b}} \mu(\liminf_{n\to\infty} f_n < a < b < \limsup_{n\to\infty} f_n) \leq \sum_{\substack{a,b\in\mathbb{Q}\\ a<b}} 0 = 0.$$

Thus there exists a limit function $f_\infty := \lim_{n\to\infty} f_n$. It remains to check that $|f_\infty|<\infty$ almost everywhere. To this end, let $A\in(\mathscr{F}_m)_{f_k;k\geq m}$ and apply Fatou's lemma to the result that

$$\int_A |f_\infty|\,\mathrm{d}\mu = \int_A \lim_{n\to\infty} |f_n|\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_A |f_n|\,\mathrm{d}\mu \leq \sup_{n\geq m} \int_A |f_n|\,\mathrm{d}\mu < \infty,$$

thus $|f_\infty|<\infty$ almost everywhere on A, and then by the countable covering property, almost everywhere on Ω.

If $\sup_{n\in\mathbb{Z}} \|f_n\|_p < \infty$, where $p\in[1,\infty]$, a similar Fatou-argument shows that

$$\|f_\infty\|_p \leq \sup_{n\in\mathbb{Z}} \|f_n\|_p.$$

For more trivial reasons, the last estimate is also true for $p=\infty$.

4.8. **Recovering f_n from f_∞.** Our next goal is to show that, under some conditions, the original martingale $(f_k)_{k\geq m}$ can be recovered from $f_\infty := \lim_{n\to\infty} f_n$ by

$$f_n = \mathbb{E}[f_\infty|\mathscr{F}_n]. \tag{4.9}$$

For this purpose, we strengthen the earlier integrability condition to the requirement that

$$(\mathscr{F}_m^0)_{f^*} := \{A\in\mathscr{F}_m^0 : \|1_A f^*\|_1 < \infty\}$$

contain a countable cover. Since $|f_k|\leq f^*$ for each k, it is clear that $(\mathscr{F}_m^0)_{f^*}\subseteq(\mathscr{F}_m^0)_{f_k;k\geq m}$, and hence this condition is stronger than the one imposed above to guarantee the existence of the limit f_∞.

Now consider some $n\geq m$, and observe that $(\mathscr{F}_n^0)_{f^*}\supseteq(\mathscr{F}_m^0)_{f^*}$ also contains a countable cover. For any $A\in(\mathscr{F}_n^0)_{f^*}$, we can then apply dominated convergence with the dominating function f^* to deduce that

$$\int_A f_\infty\,\mathrm{d}\mu = \int_A \lim_{k\to\infty} f_k\,\mathrm{d}\mu = \lim_{k\to\infty}\int_A f_k\,\mathrm{d}\mu = \lim_{\substack{k\to\infty\\ k\geq n}}\int_A f_k\,\mathrm{d}\mu,$$

and here

$$\int_A f_k\,\mathrm{d}\mu = \int_A \mathbb{E}[f_k|\mathscr{F}_n]\,\mathrm{d}\mu = \int_A f_n\,\mathrm{d}\mu \qquad \forall k\geq n;$$

thus

$$\int_A f_\infty\,\mathrm{d}\mu = \int_A f_n\,\mathrm{d}\mu \qquad \forall A\in(\mathscr{F}_n^0)_{f^*}.$$

Since $(\mathscr{F}_n^0)_{f^*}$ is an ideal containing a countable cover, we deduce from the definition of the conditional expectation that (4.9).

4.10. **Example.** We show by example that the weaker condition that $(\mathscr{F}_m^0)_{f_k;k\geq m}$ contain a countable cover, which guarantees the existence of f_∞, is not sufficient for (4.9). Indeed, let $\Omega=[0,1)$ with Lebesgue measure, $\mathscr{F}_k := \sigma(\{2^{-k}[j-1,j) : j=1,\ldots,2^k\})$, and $f_k := 2^k\cdot 1_{[0,2^{-k})}$ for $k\geq 0$. It is easy to check that this is a martingale, and $\|f_k\|_1=1$ for all $k\in\mathbb{N}$, so that condition for the existence of f_∞ is clearly satisfied. It is easy to see directly that $f_n(\omega)\to 0$ pointwise almost everywhere (indeed, in all other points except $\omega=0$), so that $f_\infty\equiv 0$. But clearly $f_n\neq\mathbb{E}[f_\infty|\mathscr{F}_n]=0$.

Thus $(\mathscr{F}_0^0)_{f^*}$ cannot contain a countable cover, and this is also easy to see directly, since $\mathscr{F}_0^0=\{\varnothing,[0,1)\}$, and $f^*(\omega)\eqsim 1/\omega$, which is not integrable over $[0,1)$. So in fact $(\mathscr{F}_0^0)_{f^*}=\{\varnothing\}$, which clearly does not contain any cover of $[0,1)$.

4.11. **Convergence of f_n to f in L^p.** Suppose that $\sup_{n\in\mathbb{Z}} \|f_n\|_p < \infty$, where $p \in (1,\infty]$. Then by Doob's inequality, we have $\|f^*\|_p < \infty$, so that $(\mathscr{F}_m^0)_{f^*} = \mathscr{F}_m^0$ contains a countable cover by σ-finiteness for every $m \in \mathbb{Z}$. Thus the pointwise limit

$$f_\infty = \lim_{n\to\infty} f_n \text{ exists, and } f_n = \mathbb{E}[f_\infty|\mathscr{F}_n] \ \forall n \in \mathbb{Z}.$$

Moreover, we have

$$\|f_\infty\|_p \le \sup_n \|f_n\|_p = \sup_n \|\mathbb{E}[f_\infty|\mathscr{F}_n]\|_p \le \|f_\infty\|_p,$$

so in fact

$$\|f_\infty\|_p = \sup_n \|f_n\|_p.$$

This shows that all the information about the martingale f_n is actually captured by the limit function f_∞ and the filtration $(\mathscr{F}_n)_{n\in\mathbb{Z}}$.

If $p \in (1,\infty)$, we can further apply the dominated convergence theorem with dominating function $2f^* \in L^p$ to see that

$$\lim_{n\to\infty} \int_\Omega |f_n - f_\infty|^p \, \mathrm{d}\mu \to 0,$$

i.e., $f_n \to f_\infty$ in the norm of L^p.

4.12. **Pointwise convergence of martingales as $m \to -\infty$.** Our next aim is to investigate the limit behaviour of f_m as $m \to -\infty$. Once again, this is done by integrating the pointwise up-crossing inequality (4.5) over appropriate sets A. Let

$$\mathscr{F}_{-\infty} := \bigcap_{n\in\mathbb{Z}} \mathscr{F}_n.$$

As an intersection of σ-algebras, this is automatically a σ-algebra. We make the assumption that

$$(\mathscr{F}_{-\infty}^0)_{f_n} := \{A \in \mathscr{F}_{-\infty}^0 : \|1_A f_n\|_1 < \infty\}$$

contain a countable cover. Note that this collection is contained in $\mathscr{F}_{-\infty}^0$, so this assumption requires in particular that $\mathscr{F}_{-\infty}^0$ be σ-finite. Let us stress that this is a very non-trivial assumption; we will return to this point later. Also note that for $k \le n$ and $A \in \mathscr{F}_{-\infty}$, we have

$$\|1_A f_k\|_1 = \|1_A \mathbb{E}[f_n|\mathscr{F}_k]\|_1 = \|\mathbb{E}[1_A f_n|\mathscr{F}_k]\|_1 \le \|1_A f_n\|_1,$$

so that in fact $(\mathscr{F}_{-\infty}^0)_{f_n} \subseteq (\mathscr{F}_{-\infty}^0)_{f_k}$ for all $k \le n$.

As before, this guarantee that everything in (4.5) is integrable over every $A \in (\mathscr{F}_{-\infty}^0)_{f_n}$, and we deduce that

$$0 = \int_A g_n \, \mathrm{d}\mu \ge (b-a) \int_A u_{m,n} \, \mathrm{d}\mu - \int_A (|f_n| + |a|) \, \mathrm{d}\mu,$$

and so

$$\int_A u_{-\infty,n} \, \mathrm{d}\mu := \int_A \lim_{m\to\infty} u_{m,n} \, \mathrm{d}\mu = \lim_{m\to\infty} \int_A u_{m,n} \, \mathrm{d}\mu \le \frac{1}{b-a} \int_A (|f_n| + |a|) \, \mathrm{d}\mu < \infty.$$

Thus $u_{-\infty,n} < \infty$ almost everywhere on each A, and hence on Ω.

By a similar argument as for $n \to \infty$, the finiteness of the up-crossings implies the existence of a pointwise limit, i.e.,

$$f_{-\infty} := \lim_{k\to-\infty} f_k \text{ exists pointwise almost everywhere.}$$

Recall that $f_k \in L^0(\mathscr{F}_k) \subseteq L^0(\mathscr{F}_m)$ for $k \le m$, and thus

$$f_{-\infty} := \lim_{k\to-\infty} f_k = \lim_{\substack{k\to-\infty \\ k\le m}} f_k \in L^0(\mathscr{F}_m) \qquad \forall m \le n,$$

thus in fact $f_{-\infty} \in \bigcap_{m\le n} L^0(\mathscr{F}_m) = L^0(\mathscr{F}_{-\infty})$.

4.13. **Formula for $f_{-\infty}$.** We keep working under the same assumptions as in the previous section. Thus the pointwise limit $f_{-\infty}$ exists. We will show that it is in fact given by the formula

$$f_{-\infty} = \mathbb{E}[f_m|\mathscr{F}_{-\infty}] \qquad \forall m \leq n.$$

This formula is an immediate consequence of the following lemma, since then for any $A \in (\mathscr{F}^0_{-\infty})_{f_n}$, we have

$$\int_A f_{-\infty}\,\mathrm{d}\mu = \lim_{\substack{k\to-\infty\\ k\leq m}} \int_A f_k\,\mathrm{d}\mu = \lim_{\substack{k\to-\infty\\ k\leq m}} \int_A \mathbb{E}[f_m|\mathscr{F}_k]\,\mathrm{d}\mu = \lim_{\substack{k\to-\infty\\ k\leq m}} \int_A f_m\,\mathrm{d}\mu = \int_A f_m\,\mathrm{d}\mu,$$

which proves the formula by the definition of conditional expectation.

4.14. **Lemma.** *For all $A \in (\mathscr{F}^0_{-\infty})_{f_n}$, we have*

$$\lim_{k\to-\infty} \|1_A(f_k - f_{-\infty})\|_1 = 0.$$

Proof. We consider the two new martingales $(g_k)_{k\leq n}$ and $(h_k)_{k\leq n}$ given by

$$g_k := \mathbb{E}(f_n \cdot 1_{\{f_n\leq R\}}|\mathscr{F}_k), \qquad h_k := \mathbb{E}(f_n \cdot 1_{\{|f_n|>R\}}|\mathscr{F}_k).$$

Note that $|g_n|, |h_n| \leq |f_n|$, so that $(\mathscr{F}^0_{-\infty})_{f_n} \subseteq (\mathscr{F}^0_{-\infty})_{g_n} \cap (\mathscr{F}^0_{-\infty})_{h_n}$. By the previous section, the pointwise limits

$$g_{-\infty} := \lim_{k\to-\infty} g_k, \qquad h_{-\infty} := \lim_{k\to-\infty} h_k \quad \text{exist almost everywhere.}$$

Since $f_k = g_k + h_k$, we also have $f_{-\infty} = g_{-\infty} + h_{-\infty}$. Also, observe that

$$\|g_k\|_\infty = \|\mathbb{E}(f_n \cdot 1_{\{f_n\leq R\}}|\mathscr{F}_k))\|_\infty \leq \|f_n \cdot 1_{\{f_n\leq R\}}\|_\infty \leq R,$$

so that $g^*_n = \sup_{k\leq n} |g_k| \leq R$, and thus each $1_A g_k$, $k \leq n$, is dominated by the integrable function $1_A g^* \leq 1_A \cdot R$, if $\mu(A) < \infty$. Thus the dominated convergence theorem implies that

$$\lim_{k\to-\infty} \|1_A(g_k - g_{-\infty})\|_=0.$$

On the other hand, we have

$$\|1_A h_k\|_1 = \|\mathbb{E}(1_A \cdot f_n \cdot 1_{\{|f_n|>R\}}|\mathscr{F}_k)\|_1 \leq \|f_n \cdot 1_{A\cap\{|f_n|>R\}}\|_1, \qquad \forall k \leq n,$$

and by Fatou's lemma also

$$\|1_A h\|_1 = \int_A \lim_{k\to\infty} |h_k|\,\mathrm{d}\mu \leq \liminf_{k\to\infty} \int_A |h_k|\,\mathrm{d}\mu \leq \|f_n \cdot 1_{A\cap\{|f_n|>R\}}\|_1.$$

Combining everything, we obtain

$$\begin{aligned}
\limsup_{k\to-\infty} \|1_A(f_k - f_{-\infty})\|_1 &= \limsup_{k\to-\infty} \|1_A(g_k + h_k - g_{-\infty} - h_{-\infty})\|_1 \\
&\leq \limsup_{k\to-\infty} \|1_A(g_k - g_{-\infty})\|_1 + \limsup_{k\to-\infty} \|1_A(h_k - h_{-\infty})\|_1 \\
&\leq 0 + 2 \cdot \|f_n \cdot 1_{A\cap\{|f_n|>R\}}\|_1,
\end{aligned}$$

which holds for any $R > 0$, so also in the limit as $R \to \infty$. Next we observe that $f_n \cdot 1_{A\cap\{|f_n|>R\}}$ is dominated by the integrable function $f_n \cdot 1_A$, and converges pointwise to zero almost everywhere, i.e., in all points where $|f_n| < \infty$. Thus

$$\limsup_{k\to-\infty} \|1_A(f_k - f_{-\infty})\|_1 \leq \lim_{R\to\infty} 2 \cdot \|f_n \cdot 1_{A\cap\{|f_n|>R\}}\|_1 = 0.$$

□

4.15. **Convergence of f_m to $f_{-\infty}$ in L^p.** We now assume that $\mathscr{F}_{-\infty}$ is σ-finite and that $\sup_{n\in\mathbb{Z}}\|f_n\|_p < \infty$ with $p \in [1,\infty)$. Thus $(\mathscr{F}^0_{-\infty})_{f_n} = \mathscr{F}^0_{-\infty}$ contains a countable cover for every n, and the previous results apply, i.e.,

$$\lim_{m\to-\infty} f_m = f_{-\infty} = \mathbb{E}[f_n|\mathscr{F}_{-\infty}] \qquad \forall n \in \mathbb{Z}.$$

If $p \in (1,\infty)$, then all f_m are dominated by $f^* \in L^p$, and we obtain from the dominated convergence theorem that $\lim_{m\to-\infty}\|f_m - f_{-\infty}\|_p = 0$, i.e., $f_m \to f_{-\infty}$ in the norm of L^p.

This norm convergence is also true in L^1, but needs a somewhat more complicated proof, using Lemma 4.14: Let $A_1 \subseteq A_2 \subseteq \ldots$ be sets in $\mathscr{F}^0_{-\infty}$ such that $\bigcup_{j=1}^\infty A_j = \Omega$; they exist by σ-finiteness. By Lemma 4.14, we have

$$\lim_{k\to-\infty}\|1_{A_j}(f_k - f_{-\infty})\|_1 = 0 \qquad \forall j \in \mathbb{Z}_+.$$

Thus

$$\begin{aligned}\limsup_{k\to-\infty}\|f_k - f_{-\infty}\|_1 &\le \limsup_{k\to-\infty}\|1_{A_j}(f_k - f_{-\infty})\|_1 + \limsup_{k\to-\infty}\|1_{A_j^\complement}(f_k - f_{-\infty})\|_1 \\ &= 0 + \limsup_{k\to-\infty}\|\mathbb{E}[1_{A_j^\complement}f_n|\mathscr{F}_k]\|_1 + \|\mathbb{E}[1_{A_j^\complement}f_n|\mathscr{F}_{-\infty}]\|_1 \le 2\cdot\|1_{A_j^\complement}f_n\|_1,\end{aligned}$$

which holds for every $j \in \mathbb{Z}_+$, and hence also in the limit as $j \to \infty$. But each $1_{A_j^\complement}f_n$ is dominated by the integrable function f_n, and these functions converge pointwise to zero; hence by dominated convergence

$$\limsup_{k\to-\infty}\|f_k - f_{-\infty}\|_1 \le 2\cdot\lim_{j\to\infty}\|1_{A_j^\complement}f_n\|_1 = 0.$$

4.16. **The case of a "poor" $\mathscr{F}_{-\infty}$.** We mentioned above that the assumption that $\mathscr{F}_{-\infty}$ be σ-finite is quite non-trivial. Indeed, in our basic example of the dyadic filtration

$$\mathscr{F}_k := \sigma(\{2^{-k}[j,j+1) : j \in \mathbb{Z}\})$$

of $\mathbb{R}$, it is not difficult to check that

$$\mathscr{F}_{-\infty} = \bigcap_{k\in\mathbb{Z}}\mathscr{F}_k = \{\varnothing, \mathbb{R}, (-\infty,0), [0,\infty)\},$$

and hence $\mathscr{F}^0_{-\infty} = \{\varnothing\}$ clearly does not contain any cover of $\mathbb{R}$.

So in fact we have $\mu(A) \in \{0,\infty\}$ for all $A \in \mathscr{F}_{-\infty}$. We already considered this situation in Section 2.15, and found that

$$\lim_{\substack{k\to-\infty\\ k\le m}} f_k = 0$$

both pointwise and in L^p, provided that $(f_k)_{k\le m}$ is a martingale with $\|f_m\|_p < \infty$.

We summarize the results of this section so far:

4.17. **Theorem** (Pointwise convergence of martingales, $n \to \infty$)**.** *Suppose that $(f_n)_{n\ge m}$ is a martingale adapted to $(\mathscr{F}_n)_{n\ge m}$. The following assertions hold:*

- *If $(\mathscr{F}^0_m)_{f_k;k\ge m}$ contains a countable cover, then*

$$f_\infty = \lim_{n\to\infty} f_n \text{ exists pointwise almost everywhere.}$$

- *If $(\mathscr{F}^0_m)_{f^*}$ contains a countable cover, then*

$$f_n = \mathbb{E}[f_\infty|\mathscr{F}_n] \qquad \forall n \ge m.$$

4.18. **Theorem** (Pointwise convergence of martingales, $m \to -\infty$)**.** *Suppose that $(f_m)_{m\le n}$ is a martingale adapted to $(\mathscr{F}_m)_{m\le n}$. If $(\mathscr{F}^0_{-\infty})_{f_n}$ contains a countable cover, then*

$$f_{-\infty} = \lim_{n\to-\infty} f_n \text{ exists pointwise a.e., and } f_{-\infty} = \mathbb{E}[f_m|\mathscr{F}_{-\infty}] \qquad \forall m \le n.$$

4.19. **Theorem** (Convergence of L^p martingales)**.** *Suppose that $(f_n)_{n\in\mathbb{Z}}$ is a martingale with $\sup_{n\in\mathbb{Z}}\|f_n\|_p < \infty$, where $p \in (1,\infty)$. Then the following assertions hold:*

- $f_\infty = \lim_{n\to\infty} f_n$ *exists pointwise a.e. and in* L^p*, and we have*
$$f_n = \mathbb{E}[f_\infty|\mathscr{F}_n] \quad \forall n\in\mathbb{Z}, \qquad \|f_\infty\|_p = \sup_{n\in\mathbb{Z}} \|f_n\|_p.$$
- $f_{-\infty} = \lim_{n\to-\infty} f_n$ *exists pointwise a.e. and in* L^p*, provided that one of the following additional conditions holds:*
  - $\mathscr{F}_{-\infty}$ *is* σ*-finite, and in this case* $f_{-\infty} = \mathbb{E}[f_n|\mathscr{F}_{-\infty}]$ *for all* $n\in\mathbb{Z}$*, or*
  - $\mathscr{F}_{-\infty}$ *contains only sets of measure* 0 *or* ∞*, and in this case* $f_{-\infty}=0$*.*

In fact, it can be shown that Ω always splits as $\Omega = \Omega_\sigma \cup \Omega_\infty$, where $\Omega_\sigma, \Omega_\infty \in \mathscr{F}_{-\infty}$ are disjoint, and $\mathscr{F}_{-\infty}\cap\Omega_\sigma$ is σ-finite on Ω_σ, and $\mathscr{F}_{-\infty}\cap\Omega_\infty$ only contains sets of measure 0 or ∞. In this way, the above mentioned two cases together actually cover all situations.

4.20. **Unconditional convergence.** In the remainder of this chapter, we consider the unconditional convergence of the series
$$\sum_{k=-\infty}^{\infty} d_k, \qquad d_k = f_k - f_{k-1},$$
where $(f_k)_{k\in\mathbb{Z}}$ is a martingale with $\sup_{n\in\mathbb{Z}} \|f_n\|_p < \infty$ for some $p\in(1,\infty)$. So in fact $f_n = \mathbb{E}[f_\infty|\mathscr{F}_n]$ for a limit function $f_\infty \in L^p$. We first prove two abstract results in an arbitrary Banach space X. (In our martingale application, we take $X = L^p$.)

4.21. **Proposition.** *Let* $\sum_{k=1}^\infty x_k$ *be a series in a Banach space. If the series converges unconditionally, then the value is independent of the order of summation, i.e.,*
$$\sum_{k=1}^\infty x_{\sigma(k)} = x$$
for a fixed $x\in X$ *and every permutation* σ *of* $\mathbb{Z}_+$*.*

We first observe:

4.22. **Lemma.** *If* σ *is any permutation of* $\mathbb{Z}_+$*, then* $\sigma(k)\to\infty$ *as* $k\to\infty$*.*

Proof of the Lemma. Given any n, we need to find an m such that $\sigma(k) > n$ for all $k > m$. Since σ is surjective, we can find some numbers $k_1,\ldots,k_n$ with $\sigma(k_i) = i$ for each $i = 1,\ldots,n$. Let $m := \max\{k_1,\ldots,k_n\}$, and consider any $k > m$, so in particular $k \notin \{k_1,\ldots,k_n\}$. Since σ is injective, we have $\sigma(k)\notin\{\sigma(k_1),\ldots,\sigma(k_n)\} = \{1,\ldots,n\}$, and thus $\sigma(k) > n$. □

Proof of Proposition 4.21. We argue by contradiction. Suppose that there are two permuations σ,τ with
$$\sum_{k=1}^\infty x_{\sigma(k)} = x \neq y = \sum_{k=1}^\infty x_{\tau(k)}$$
We construct a third permutation ρ so that $\sum_{k=1}^\infty x_{\rho(k)}$ does not converge.

Pick a sequence $\epsilon_k \downarrow 0$. We will inductively choose natural numbers
$$0 < m_1 < n_1 < m_2 < n_2 < \ldots.$$
Suppose that we have already chosen $m_1, n_1,\ldots,m_j,n_j$. (This is trivial if $j=0$; then we haven't chosen anything yet.) Choose $m_{j+1} > n_j$ so large that $\{\tau(1),\ldots,\tau(n_j)\} \subseteq \{\sigma(1),\ldots,\sigma(m_{j+1})\}$ (this condition is empty if $j=0$; otherwise, it is possible since σ is surjective), and
$$\Big\|\sum_{k=1}^{m_{j+1}} x_{\sigma(k)} - x\Big\|_X < \epsilon_{j+1}.$$
Next we pick n_{j+1} so large that $\{\sigma(1),\ldots,\sigma(m_{j+1})\}\subseteq\{\tau(1),\ldots,\tau(n_{j+1})\}$, and
$$\Big\|\sum_{k=1}^{n_{j+1}} x_{\tau(k)} - x\Big\|_X < \epsilon_{j+1}.$$

So now we have

$$\{\tau(1), \dots, \tau(n_j)\} \subseteq \{\sigma(1), \dots, \sigma(m_{j+1})\} \subseteq \{\tau(1), \dots, \tau(n_{j+1})\} \to \mathbb{Z}_+ \text{ as } j \to \infty,$$

since $n_{j+1} \to \infty$, and then also $\tau(n_{j+1}) \to \infty$, and τ is surjective.

We can inductively choose the values of the permuation ρ so that

$$\{\rho(1), \dots, \rho(m_j)\} = \{\sigma(1), \dots, \sigma(m_j)\} \subseteq \{\rho(1), \dots, \rho(n_j)\} = \{\tau(1), \dots, \tau(n_j)\} \quad \forall j \in \mathbb{Z}_+.$$

But then some of the partial sums of $\sum_{k=1}^\infty x_{\rho(k)}$, namely, $\sum_{k=1}^{m_j} x_{\rho(k)} = \sum_{k=1}^{m_j} x_{\sigma(k)}$, converge to x, and some others, namely, $\sum_{k=1}^{n_j} x_{\rho(k)} = \sum_{k=1}^{n_j} x_{\tau(k)}$, converge to y. By the uniqueness of the limit, the full series cannot converge. □

4.23. **Proposition** (Characterization of unconditionality)**.** *Let $\sum_{k=1}^\infty x_k$ be a series in a Banach space. Then the following assertions are equivalent:*

- *$\sum_{k=1}^\infty x_k$ converges unconditionally.*
- *$\sum_{k=1}^\infty \delta_k x_k$ converges for every choice of $\delta_k \in \{0,1\}$.*

Proof. We only prove "$\Leftarrow$", and we argue by contradiction. Suppose that there is a permutation σ so that $\sum_{k=1}^\infty x_{\sigma(k)}$ does not converge (and hence, is not Cauchy). Then we prove that there is a choice of $\delta_k \in \{0,1\}$ so that $\sum_{k=1}^\infty \delta_k x_k$ does not converge either.

Fix an $\epsilon > 0$ so that $\sum_{k=1}^\infty x_{\sigma(k)}$ fails the Cauchy criterion with ϵ, namely, there exist arbitrarily large $m < n$ such that $\| \sum_{k=m}^n x_{\sigma(k)} \|_X \geq \epsilon$.

Again, we first find a sequence of numbers $m_1 < n_1 < m_2 < n_2 < \dots$. Suppose that we have already chosen $m_1, n_1, \dots, m_j, n_j$. Now we choose the numbers $m_{j+1} < n_{j+1}$ so large that $\sigma(k) > \max\{\sigma(i) : 1 \leq i \leq n_j\}$ for all $k \geq m_{j+1}$, and

$$\Big\| \sum_{k=m_{j+1}}^{n_{j+1}} x_{\sigma(k)} \Big\|_X \geq \epsilon.$$

Next, we define

$$\delta_i := \begin{cases} 1, & \text{if } i = \sigma(k) \text{ for some } k \in [m_j, n_j] \text{ and } j \in \mathbb{Z}_+, \\ 0, & \text{else.} \end{cases}.$$

Note that since the intervals $[m_j, n_j]$ are pairwise disjoint and σ is injective, the first condition can hold for at most one $j \in \mathbb{Z}_+$. We now show that $\sum_{i=1}^\infty \delta_i x_i$ is not Cauchy, hence not convergent. Indeed, consider the partial sum

$$\sum_{i=\min\{\sigma(k):k\in[m_j,n_j]\}}^{\max\{\sigma(k):k\in[m_j,n_j]\}} \delta_i x_i = \sum_{k=m_j}^{n_j} x_{\sigma(k)};$$

to verify the identity observe the following: $\delta_i = 1$ if and only if $i = \sigma(k)$ for some $k \in [m_\ell, n_\ell]$, and some $\ell \in \mathbb{Z}_+$. If $\ell < j$, and $h \leq n_\ell \leq n_{j-1}$, then $\sigma(h) < \sigma(k)$ for all $k \geq m_j$, so this case does not appear. If $\ell > j$, and $h \geq m_\ell \geq m_{j+1}$, then $\sigma(h) > \sigma(k)$ for all $k \leq n_j$, so this case also does not appear. This only leaves the possibility that $\ell = j$, and then the identity is clear.

But the sums on the right of the above identity are at least ϵ in norm; hence so are the sums on the left. Since these are partial sums of $\sum_{i=1}^\infty \delta_i x_i$, the series cannot converge. □

Now we are ready for:

4.24. **Theorem** (Unconditional convergence of L^p martingales)**.** *Let $d_k = f_k - f_{k-1}$, where $(f_n)_{n\in\mathbb{Z}}$ is a martingale with $\sup_{n\in\mathbb{Z}} \|f_n\|_p < \infty$. Then the series*

$$\sum_{k=-\infty}^{\infty} d_k$$

converges unconditionally in L^p.

Proof. We already know that the series converges, hence it satisfies the Cauchy criterion

$$\Big\| \sum_{k=m}^{n} d_k \Big\|_{L^p} \to 0 \text{ as } m, n \to \pm\infty,$$

where we mean that both $m, n \to \infty$, or both $m, n \to -\infty$.

Let then $\delta_k \in \{0, 1\}$ be some numbers. In particular, we can view them as predictable (constant) functions on Ω. Then Burkholder's inequality shows that

$$\Big\| \sum_{k=m}^{n} \delta_k d_k \Big\|_{L^p} \leq \beta_p \Big\| \sum_{k=m}^{n} d_k \Big\|_{L^p} \to 0 \text{ as } m, n \to \pm\infty,$$

so also the series $\sum_{k=-\infty}^{\infty} \delta_k d_k$ satisfies the Cauchy criterion, and hence is convergent. Since this holds for every choice of $\delta_k \in \{0, 1\}$, we see that $\sum_{k=-\infty}^{\infty} d_k$ is unconditionally convergent. □

Thanks to this theorem, we are justified to write

$$\sum_{k \in \mathbb{Z}} d_k$$

in place of $\sum_{k=-\infty}^{\infty} d_k$; this notation emphasizes the fact that we sum over all values $k \in \mathbb{Z}$, but the order of summation is not important.

4.25. **Exercises.**

1. Prove the other direction of Proposition 4.23.
2. Use orthogonality considerations to give another (easier) proof of the following martingale convergence results in L^2: Suppose that $(f_n)_{n \in \mathbb{Z}}$ is a martingale with $\sup_{n \in \mathbb{Z}} \|f_n\|_2 < \infty$. Then the limits
$$f_{-\infty} = \lim_{n \to -\infty} f_n, \qquad f_\infty = \lim_{n \to \infty} f_n$$
exist in the sense of L^2 convergence. [Hint: Check first that
$$\sum_{k=-\infty}^{\infty} \|d_k\|_2^2 \leq \sup_{n \in \mathbb{Z}} \|f_n\|_2^2$$
and apply the Cauchy criterion. Use orthogonality!]

4.26. **References.** We have partly followed the presentation of Williams [16] in proving the martingale convergence results with the help of the up-crossing technique.

5. Petermichl's dyadic shift and the Hilbert transform

5.1. **Dyadic systems of intervals.** We call $\mathscr{D}$ a *dyadic system* (of intervals) if $\mathscr{D} = \bigcup_{j \in \mathbb{Z}} \mathscr{D}_j$, where each $\mathscr{D}_j$ is a partition of $\mathbb{R}$ consisting of intervals of the form $[x, x + 2^{-j})$, and each interval $I \in \mathscr{D}$ is a union of two intervals I_- and I_+ (its left and right halves) from $\mathscr{D}_{j+1}$. Let us derive a representation for arbitrary dyadic systems in terms of the standard dyadic system $\mathscr{D}^0 = \bigcup_{j \in \mathbb{Z}} \mathscr{D}_j^0$, where $\mathscr{D}_j^0 = \{2^{-j}[k, k+1) : k \in \mathbb{Z}\}$. (This already appeared in Exercise 2.17(4).)

It is easy to see that $\mathscr{D}_j$ has to be of the form $\mathscr{D}_j^0 + x_j$ for some $x_j \in \mathbb{R}$. If one adds an integer multiple of 2^{-j} to x_j, the collection $\mathscr{D}_j^0 + x_j$ does not change, so one can demand that $x_j \in [0, 2^{-j})$. Then x_j is actually the unique end-point of intervals in $\mathscr{D}_j$, which falls on the interval $[0, 2^{-j})$. Since this is also an end-point of the intervals in $\mathscr{D}_{j+1}$, there must hold $x_j - x_{j+1} \in \{0, 2^{-j-1}\}$. Let us write $\beta_{j+1} := 2^{j+1}(x_j - x_{j+1}) \in \{0, 1\}$ so that $x_j = 2^{-j-1}\beta_j + x_{j+1}$, and by iteration

$$x_j = \sum_{i > j} 2^{-i} \beta_i, \qquad \beta = (\beta_i)_{i \in \mathbb{Z}} \in \{0, 1\}^{\mathbb{Z}}.$$

Hence an arbitrary dyadic system is of the form $\mathscr{D}^\beta$, where $\mathscr{D}_j^\beta := \mathscr{D}_j^0 + \sum_{i > j} 2^{-i} \beta_i$.

In the sequel we will also need *dilated* dyadic systems $r\mathscr{D}^\beta := \{rI : I \in \mathscr{D}^\beta\}$, where $rI = [ra, rb)$ if $I = [a, b)$. Note that $2^j \mathscr{D}^\beta = \mathscr{D}^{\beta'}$ for another $\beta' \in \{0, 1\}^{\mathbb{Z}}$, so only the dilation factors $r \in [1, 2)$ will be relevant.

5.2. **Dyadic σ-algebras and conditional expectations.** Let $\mathscr{F}_j^\beta := \sigma(\mathscr{D}_j^\beta)$, and then $r\mathscr{F}_j^\beta = \sigma(r\mathscr{D}_j^\beta)$. Let us consider $\beta \in \{0,1\}^{\mathbb{Z}}$ and $r \in [1,2)$ fixed for the moment, and write simply $\mathscr{F}_j := r\mathscr{F}_j^\beta$ and $\mathscr{D}_j := r\mathscr{D}_j^\beta$. Then $(\mathscr{F}_j)_{j\in\mathbb{Z}}$ is a filtration (Exercise 2.17(4)). Moreover,

$$\sigma\Big(\bigcup_{j\in\mathbb{Z}} \mathscr{F}_j\Big) = \mathscr{B}(\mathbb{R}), \qquad \forall\, F \in \bigcap_{j\in\mathbb{Z}} \mathscr{F}_j : |F| \in \{0,\infty\},$$

where $\mathscr{B}(\mathbb{R})$ stands for the Borel σ-algebra of $\mathbb{R}$, and $|F|$ for the Lebesgue measure of $F \in \mathscr{B}(\mathbb{R})$. For the first property one checks that every open set $\mathscr{O} \subseteq \mathbb{R}$ is a (necessarily countable) union of dyadic intervals. For the second, note that if $F \in \mathscr{F}_j \setminus \{\varnothing\}$, then $|F| \geq r2^{-j}$, and this tends to $+\infty$ as $j \to -\infty$.

5.3. **Haar functions.** Let $L^p(\mathbb{R}) := L^p(\mathscr{B}(\mathbb{R}), \mathrm{d}x)$. By Theorem 2.16, it follows that every $f \in L^p(\mathbb{R})$ has the following series representation which converges both pointwise and in the L^p norm:

$$\begin{aligned}
f &= \sum_{j=-\infty}^{\infty} \big(\mathbb{E}[f|\mathscr{F}_{j+1}] - \mathbb{E}[f|\mathscr{F}_j]\big) \\
&= \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} \Big(\frac{1_{I_-}}{|I_-|}\int_{I_-} f\,\mathrm{d}x + \frac{1_{I_+}}{|I_+|}\int_{I_+} f\,\mathrm{d}x - \frac{1_I}{|I|}\int_I f\,\mathrm{d}x\Big) \\
&= \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} \Big(1_{I_-}\frac{2}{|I|}\int_{I_-} f\,\mathrm{d}x + 1_{I_+}\frac{2}{|I|}\int_{I_+} f\,\mathrm{d}x - \big(1_{I_+} + 1_{I_-}\big)\frac{1}{|I|}\Big\{\int_{I_+} f\,\mathrm{d}x + \int_{I_-} f\,\mathrm{d}x\Big\}\Big) \\
&= \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} \Big(1_{I_-}\Big\{\frac{1}{|I|}\int_{I_-} f\,\mathrm{d}x - \frac{1}{|I|}\int_{I_+} f\,\mathrm{d}x\Big\} + 1_{I_+}\Big\{\frac{1}{|I|}\int_{I_+} f\,\mathrm{d}x - \frac{1}{|I|}\int_{I_-} f\,\mathrm{d}x\Big\}\Big) \\
&= \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} \Big(1_{I_-} - 1_{I_+}\Big)\frac{1}{|I|}\int \big(1_{I_-} - 1_{I_+}\big) f\,\mathrm{d}x = \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} h_I \int h_I f\,\mathrm{d}x,
\end{aligned}$$

where the *Haar function* h_I associated to the interval I is defined by

$$h_I := |I|^{-1/2}\big(1_{I_-} - 1_{I_+}\big).$$

Note that

$$h_I(x) = |I|^{-1/2} h\Big(\frac{x - \inf I}{|I|}\Big), \qquad h := h_{[0,1)} = 1_{[0,1/2)} - 1_{[1/2,1)}.$$

Let us write $\langle h_I, f\rangle := \int h_I f\,\mathrm{d}x$. By Burkholder's inequality with the random signs (Section **??**), it follows that

$$(*) \qquad \beta^{-1}\|f\|_p \leq \Big(\int_{\mathbb{R}} \mathbb{E}\Big|\sum_{j=-\infty}^{\infty} \varepsilon_j \sum_{I\in\mathscr{D}_j} h_I(x)\langle h_I, f\rangle\Big|^p \,\mathrm{d}x\Big)^{1/p} \leq \beta\|f\|_p.$$

5.4. **Petermichl's dyadic shift.** The dyadic shift operator Ш $=$ Ш$^{\beta,r}$ associated to the dyadic system $\mathscr{D} = r\mathscr{D}^\beta$ is defined as a modification of the Haar expansion $f = \sum_{-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} h_I\langle h_I, f\rangle$:

$$Ш f := \sum_{j=-\infty}^{\infty} \sum_{I\in\mathscr{D}_j} H_I\langle h_I, f\rangle, \qquad H_I := 2^{-1/2}(h_{I_-} - h_{I_+}) = |I|^{-1/2}\big(1_{I_{--}\cup I_{++}} - 1_{I_{-+}\cup I_{+-}}\big),$$

where $I_{--} := (I_-)_-$ and so on. (The symbol Ш is the Cyrillic letter 'š' as a reference to the word 'shift', which starts with this sound.)

Now there is the question of convergence of the above series and the boundedness of the shift operator. For $I \in \mathscr{D}$, let I^* be the unique interval $I^* \in \mathscr{D}$ such that $I^* \supset I$ and $|I^*| = 2|I|$. Let $\alpha_I := +1$ if $I = I^*_-$ and $\alpha_I := -1$ if $I = I^*_+$. Then observe that

$$\sum_{j=m}^{n} \sum_{I\in\mathscr{D}_j} 2^{-1/2}(h_{I_-} - h_{I_+})\langle h_I, f\rangle = \sum_{j=m}^{n} \sum_{J\in\mathscr{D}_{j+1}} \alpha_J 2^{-1/2} h_J \langle h_{J^*}, f\rangle.$$

By $(*)$ of Section 5.3 it follows that

$$\Big\|\sum_{j=m}^{n}\sum_{I\in\mathscr{D}_j}H_I\langle h_I,f\rangle\Big\|_p\leq\beta\Big(\int_{\mathbb{R}}\mathbb{E}\Big|\sum_{j=m}^{n}\varepsilon_j\sum_{J\in\mathscr{D}_{j+1}}\alpha_J 2^{-1/2}h_J(x)\langle h_{J^*},f\rangle\Big|^p\,\mathrm{d}x\Big)^{1/p}.$$

Now comes the core trick of the argument! For a fixed x, there is only one non-zero term in the sum $J\in\mathscr{D}_{j+1}$ for each j — indeed, the one with $J\ni x$. When this term $\xi_j:=\alpha_J 2^{-1/2}h_J(x)\langle h_{J^*},f\rangle$ is multiplied by the random sign ε_j, it does not matter if the ξ_j itself is positive or negative; in any case $\varepsilon_j\xi_j$ is a random variable which is equal to $-\xi_j$ with probability $\frac{1}{2}$ and $+\xi_j$ with probability $\frac{1}{2}$, and it is independent of the other $\varepsilon_i\xi_i$ for $i\neq j$. Hence the resulting random variable would have the same distribution if $h_J(x)$ were replaced by $|h_J(x)|=|J|^{-1/2}1_J(x)$. Thus

$$\begin{aligned}
&\Big(\int_{\mathbb{R}}\mathbb{E}\Big|\sum_{j=m}^{n}\varepsilon_j\sum_{J\in\mathscr{D}_{j+1}}\alpha_J 2^{-1/2}h_J(x)\langle h_{J^*},f\rangle\Big|^p\,\mathrm{d}x\Big)^{1/p}\\
&=\Big(\int_{\mathbb{R}}\mathbb{E}\Big|\sum_{j=m}^{n}\varepsilon_j\sum_{J\in\mathscr{D}_{j+1}}\alpha_J (2|J|)^{-1/2}1_J(x)\langle h_{J^*},f\rangle\Big|^p\,\mathrm{d}x\Big)^{1/p}\\
&=\Big(\int_{\mathbb{R}}\mathbb{E}\Big|\sum_{j=m}^{n}\varepsilon_j\sum_{I\in\mathscr{D}_{j}}|I|^{-1/2}\big(1_{I_-}(x)-1_{I_+}(x)\big)\langle h_I,f\rangle\Big|^p\,\mathrm{d}x\Big)^{1/p}\\
&=\Big(\int_{\mathbb{R}}\mathbb{E}\Big|\sum_{j=m}^{n}\varepsilon_j\sum_{I\in\mathscr{D}_{j}}h_I(x)\langle h_I,f\rangle\Big|^p\,\mathrm{d}x\Big)^{1/p}\leq\beta\Big\|\sum_{j=m}^{n}\sum_{I\in\mathscr{D}_j}h_I\langle h_I,f\rangle\Big\|_p.
\end{aligned}$$

Combining everything

$$\Big\|\sum_{j=m}^{n}\sum_{I\in\mathscr{D}_j}H_I\langle h_I,f\rangle\Big\|_p\leq\beta^2\Big\|\sum_{j=m}^{n}\sum_{I\in\mathscr{D}_j}h_I\langle h_I,f\rangle\Big\|_p.$$

The right side tends to zero as $m,n\to\infty$ or $m,n\to-\infty$; hence so does the left side, and thus by Cauchy's criterion the series $\sum_{j=-\infty}^{\infty}\sum_{I\in\mathscr{D}_j}H_I\langle h_I,f\rangle$ converges in $L^p(\mathbb{R})$, and the limit $\text{Ш}f$ satisfies

$$\|\text{Ш}f\|_p=\lim_{\substack{n\to+\infty\\ m\to-\infty}}\Big\|\sum_{j=m}^{n}\sum_{I\in\mathscr{D}_j}H_I\langle h_I,f\rangle\Big\|_p\leq\beta^2\|f\|_p.$$

5.5. **The Hilbert transform.** The Hilbert transform is formally the singular integral

$$\text{“}Hf(x)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1}{y}f(x-y)\,\mathrm{d}y,\text{”}$$

but to make precise sense of the right side one needs to be a bit more careful. Hence one defines the *truncated* Hilbert transforms

$$H_{\varepsilon,R}f(x):=\frac{1}{\pi}\int_{\varepsilon<|y|<R}\frac{1}{y}f(x-y)\,\mathrm{d}y$$

and, for $f\in L^p(\mathbb{R})$,

$$Hf:=\lim_{\substack{\varepsilon\to0\\ R\to\infty}}H_{\varepsilon,R}f$$

if the limits exists in $L^p(\mathbb{R})$.

Simple examples show that, for a general $f\in L^p(\mathbb{R})$, this can only happen in the range $p\in(1,\infty)$. In fact, for $f=1_{(a,b)}$, there holds

$$H_{\varepsilon,R}1_{(a,b)}(x)\to\log\Big|\frac{x-a}{x-b}\Big|$$

pointwise. In the neighbourhood of the points a,b, the logarithmic singularity belongs to L^p for all $p<\infty$, but of course not to L^∞. As $x\to\infty$,

$$\log\Big|\frac{x-a}{x-b}\Big|=\log\Big|\frac{1-a/x}{1-b/x}\Big|=\log(1-\frac{a}{x})-\log(1-\frac{b}{x})=\frac{b-a}{x}+O(\frac{1}{x^2}),$$

which is in L^p for all $p > 1$ but not in L^1.

5.6. **Invariance considerations.** For $r \in (0, \infty)$, let δ_r denote the dilation of a function by r, $\delta_r f(x) := f(rx)$. For $h \in \mathbb{R}$, let $\tau_h f(x) := f(x+h)$ be the translation by h. Both these are clearly bounded operators on all $L^p(\mathbb{R})$ spaces, $p \in [1, \infty]$.

Simple changes of variables in the defining formula show that

$$H_{\varepsilon,R}\delta_r f = \delta_r H_{\varepsilon r,Rr} f, \qquad H_{\varepsilon,R}\tau_h f = \tau_h H_{\varepsilon,R} f,$$

and hence, if Hf exists, so do $H\delta_r f$ and $H\tau_h f$, and

$$H\delta_r f = \delta_r Hf, \qquad H\tau_h f = \tau_h Hf.$$

These properties are referred to as the invariance of H under dilations and translations.

The aim is to prove the existence of Hf for all $f \in L^p(\mathbb{R})$ by relating H to the dyadic shift operators. The basic obstacle is the fact that the Ш operators are neither translation nor dilation invariant: If $f = h_I$ for a $I \in \mathscr{D}$, then $Шf = H_I$, but if $f = h_J$, where $J \notin \mathscr{D}$ is a slightly translated or dilated version of I, then $Шf$ has a much more complicated expression.

The idea to overcome this problem is to average over the shifts $Ш^{\beta,r}$ associated to all translated and dilated dyadic systems $r\mathscr{D}^\beta$

5.7. **The average dyadic shift operator.** Let the space $\{0,1\}^{\mathbb{Z}}$ be equipped with the probability measure μ such that the coordinates β_j are independent and have probability $\mu(\beta = 0) = \mu(\beta = 1) = 1/2$. On $[1, 2)$, the measure $\mathrm{d}r/r$ will be used; this is the restiction on the mentioned interval of the invariant measure of the multiplicative group $(\mathbb{R}_+, \cdot)$.

We would like to define the average dyadic shift as the following integral:

$$(*) \quad \langle Ш \rangle f(x) := \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) Ш^{\beta,r} f(x) = \sum_{j\in\mathbb{Z}} \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) \sum_{I \in r\mathscr{D}_j^\beta} H_I(x) \langle h_I, f \rangle,$$

but this needs first some justification.

Let $M^{\beta,r}$ denote Doob's maximal operator related to the filtration $(\sigma(r\mathscr{D}_j^\beta))_{j\in\mathbb{Z}}$. Then observe that

$$Ш^{\beta,r}_{-m,n} f(x) := \sum_{j=-m}^{n} \sum_{I \in r\mathscr{D}_j^\beta} H_I(x)\langle h_I, f\rangle = \mathbb{E}[Ш^{\beta,r} f | \sigma(r\mathscr{D}_{n+2})](x) - \mathbb{E}[Ш^{\beta,r} f | \sigma(r\mathscr{D}_{-m+1})](x)$$

is pointwise dominated by $2M^{\beta,r}f(x)$ and converges a.e. to $Ш^{\beta,r} f(x)$ as $m, n \to \infty$. It is easy to see that the above finite sums are measurable with respect to the triplet (x, β, r), and hence so is the pointwise limit $Ш^{\beta,r} f(x)$.

To see that $(\beta, r) \mapsto Ш^{\beta,r} f(x)$ is integrable for a.e. $x \in \mathbb{R}$, and to justify the equality in $(*)$ above, note that by Jensen's inequality, Doob's inequality, and the uniform boundedness of the operators $Ш^{\beta,r}$, there holds

$$\begin{aligned}
&\int_{\mathbb{R}} \Big[\int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) M^{\beta,r}\{Ш^{\beta,r} f\}(x) \Big]^p \,\mathrm{d}x \\
&\le \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) \int_{\mathbb{R}} \big[M^{\beta,r}\{Ш^{\beta,r} f\}(x) \big]^p \,\mathrm{d}x \le \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) C\|f\|_p^p \le C\|f\|_p^p.
\end{aligned}$$

In particular, this shows that

$$\int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) M^{\beta,r}\{Ш^{\beta,r} f\}(x) < \infty$$

for a.e. $x \in \mathbb{R}$. So $Ш^{\beta,r}_{-m,n} f(x)$ is dominated by the integrable function $M^{\beta,r}\{Ш^{\beta,r} f\}(x)$ and converges to $Ш^{\beta,r} f(x)$ as $m, n \to \infty$; hence $Ш^{\beta,r} f(x)$ is integrable and dominated convergence proves that

$$\int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) Ш^{\beta,r} f(x) = \lim_{m,n\to\infty} \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) Ш^{\beta,r}_{-m,n} f(x)$$

which, unravelling the definition of $Ш^{\beta,r}_{-m,n}$, is the same as $(*)$. Finally, since the right side above is dominated by $\int \mathrm{d}r/r \int \mathrm{d}\mu(\beta) M^{\beta,r}\{Ш^{\beta,r}f\} \in L^p(\mathbb{R})$, it follows from another application of dominated convergence that the series in $(*)$ also converges in the L^p norm.

From the first form in $(*)$ it follows that

$$\begin{aligned}\|\langle Ш\rangle f\|_p &\leq \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) \|Ш^{\beta,r} f\|_p \\ &\leq \int_1^2 \frac{\mathrm{d}r}{r} \int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) (p^*-1)^2 \|f\|_p = \log 2 \cdot (p^*-1)^2 \|f\|_p.\end{aligned}$$

5.8. **Evaluation of the integral.** Next, we would like to obtain a new expression for $\langle Ш\rangle f$ in order to relate it to the Hilbert transform. Observe that

$$r\mathscr{D}^\beta_j = r2^{-j}(\mathscr{D}^0_0 + \sum_{i=1}^\infty 2^{-i}\beta_{j+i}).$$

When each of the numbers β_j is independently chosen from $\{0,1\}$, both values having equal probability, the binary expansion $\sum_{i=1}^\infty 2^{-i}\beta_{j+i}$ is uniformly distributed over $[0,1)$, and hence

$$\begin{aligned}&\int_{\{0,1\}^{\mathbb{Z}}} \mathrm{d}\mu(\beta) \sum_{I\in r\mathscr{D}^\beta_j} H_I(x)\langle h_I, f\rangle = \int_0^1 \mathrm{d}u \sum_{I\in r2^{-j}(\mathscr{D}^0_0+u)} H_I(x)\langle h_I, f\rangle \\ &= \int_0^1 \mathrm{d}u \sum_{k\in\mathbb{Z}} H_{r2^{-j}([0,1)+k+u)}(x)\langle h_{r2^{-j}([0,1)+k+u)}, f\rangle \\ &= \int_{-\infty}^\infty H_{r2^{-j}([0,1)+v)}(x)\langle h_{r2^{-j}([0,1)+v)}, f\rangle,\end{aligned}$$

where the second step just used the fact that $\mathscr{D}^0_0 = \{[0,1)+k : k\in\mathbb{Z}\}$, and in the last one the order of summation and integration was first exchanged (this is easy to justify thanks to the support properties) and the new variable $v := k+u$ introduced.

Making the further change of variables $t := 2^{-j}r$, it follows that

$$\langle Ш\rangle f(x) = \int_0^\infty \frac{\mathrm{d}t}{t} \int_{-\infty}^\infty H_{t([0,1)+v)}(x)\langle h_{t([0,1)+v)}, f\rangle \,\mathrm{d}v,$$

where $\int_0^\infty$ is actually the indefinite integral $\lim_{m,n\to\infty}\int_{2^{-n}}^{2^m}$. Recall that $h_{t([0,1)+v)}(y) = t^{-1/2}h(y/t - v)$ with $h = h_{[0,1)}$, and similarly for $H_{t([0,1)+v)}$. For a fixed t, the integrand above is hence

$$\begin{aligned}&\int_{-\infty}^\infty t^{-1/2} H\Big(\frac{x}{t}-v\Big) \int_{-\infty}^\infty t^{-1/2} h\Big(\frac{y}{t}-v\Big) f(y)\,\mathrm{d}y\,\mathrm{d}v \\ &= \int_{-\infty}^\infty \frac{1}{t}\int_{-\infty}^\infty H\Big(\frac{x}{t}-v\Big) h\Big(\frac{y}{t}-v\Big)\,\mathrm{d}v f(y)\,\mathrm{d}y.\end{aligned}$$

The inner integral is most easily evaluated by recognizing it as the integral of the function $(\xi,\eta)\mapsto H(\xi)h(\eta)$ along the straight line containing the point $(x/t, y/t)$ and having slope 1. The result depends only on $u := x/t - y/t$ and is the piecewise linear function $k(u)$ of this variable, which takes the values $0, -\frac{1}{4}, 0, \frac{3}{4}, 0, -\frac{3}{4}, 0, \frac{1}{4}, 0$ at the points $-1, -\frac{3}{4}, \ldots, \frac{3}{4}, 1$, interpolates linearly between them, and vanishes outside of $(-1,1)$. So

$$\langle Ш\rangle f = \int_0^\infty k_t * f \frac{\mathrm{d}t}{t} = \lim_{\substack{\varepsilon\to 0\\ R\to\infty}} \int_\varepsilon^R k_t * f\frac{\mathrm{d}t}{t},$$

where the limit exists in $L^p(\mathbb{R})$, and the notations $k_t(x) := t^{-1}k(t^{-1}x)$ and

$$k * f(x) := \int_{-\infty}^\infty k(x-y)f(y)\,\mathrm{d}y = \int_{-\infty}^\infty k(y)f(x-y)\,\mathrm{d}y$$

were used. These notation will also be employed in the sequel. $k*f$ is called the *convolution* of k and f.

5.9. **The appearance of the Hilbert transform.** Let us evaluate the integral

$$\int_\varepsilon^R k_t(x)\frac{\mathrm{d}t}{t} = \int_\varepsilon^R k(x/t)\frac{\mathrm{d}t}{t^2} = \frac{1}{x}\int_{x/R}^{x/\varepsilon} k(u)\,\mathrm{d}u = \frac{1}{x}[K(x/\varepsilon) - K(x/R)], \quad K(x) := \int_0^x k(u)\,\mathrm{d}u.$$

From the fact that k is odd ($k(-x) = -k(x)$) it follows that K is even ($K(-x) = K(x)$). Since k is supported on $[-1,1]$, its integral K is a constant on the complement, and in fact $K(x) = -1/8$ for $|x| \geq 1$. Write $\phi(x) := x^{-1}K(x)1_{[-1,1]}(x)$, which is again an odd function. Then

$$\frac{1}{x}K\big(\frac{x}{\varepsilon}\big) = \frac{1}{\varepsilon}\frac{\varepsilon}{x}\Big(K\big(\frac{x}{\varepsilon}\big)1_{[-1,1]}\big(\frac{x}{\varepsilon}\big) - \frac{1}{8}1_{[-1,1]^c}\big(\frac{x}{\varepsilon}\big)\Big) = \phi_\varepsilon(x) - \frac{1}{8x}1_{|x|>\varepsilon},$$

hence

$$\frac{1}{x}[K(x/\varepsilon) - K(x/R)] = \phi_\varepsilon(x) - \phi_R(x) - \frac{1}{8x}1_{\varepsilon<|x|<R},$$

and finally

$$\int_\varepsilon^R k_t * f\frac{\mathrm{d}t}{t} = \phi_\varepsilon * f - \phi_R * f - \frac{\pi}{8}H_{\varepsilon,R}f.$$

As $\varepsilon \to 0, R \to \infty$ this sum converges to a limit in $L^p(\mathbb{R})$, in fact, to $\langle Ш \rangle f$. So to complete the proof of the existence of the Hilbert transform Hf, it remains to prove that $\phi_\varepsilon * f$ and $\phi_R * f$ also converge in $L^p(\mathbb{R})$. In fact, as will be proved below, they converge to zero. Taking this claim for granted for the moment, it follows that

$$Hf = \lim_{\substack{\varepsilon\to 0\\ R\to\infty}} H_{\varepsilon,R}f = -\frac{8}{\pi}\langle Ш \rangle f$$

in $L^p(\mathbb{R})$. In particular, H is a bounded operator as a constant multiple of the average of the bounded operators $Ш^{\beta,r}$. In fact, one gets the estimate

$$\|Hf\|_p = \frac{8}{\pi}\|\langle Ш \rangle f\|_p \leq \frac{8}{\pi}\log 2 \cdot (p^* - 1)^2\|f\|_p,$$

but this is far from being optimal.

But, as said, it still remains to prove

$$\lim_{\varepsilon\to 0}\phi_\varepsilon * f = \lim_{R\to\infty}\phi_R * f = 0.$$

This will follow from the general results below; it is easy to check that ϕ satisfies all the required properties. It is an odd function, which implies $\int \phi(x)\,\mathrm{d}x = 0$, and since $|k(x)|$ is bounded by $3/4$, it follows that $|K(x)| \leq 3/4 \cdot |x|$ and hence $x^{-1}K(x)$ and then $\phi(x)$ is bounded. Finally, recall that ϕ is supported on $[-1,1]$.

5.10. **Lemma.** *Suppose that $|\phi(x)| \leq C(1+|x|)^{-1-\delta}$ for some $\delta > 0$. Then $|\phi_\varepsilon * f(x)| \leq C'Mf(x)$, where M is the Hardy–Littlewood maximal operator.*

Proof. By making simple changes of variables and splitting the integration domain it follows that

$$\begin{aligned}
|\phi_\varepsilon * f(x)| &= \Big|\int \phi(y)f(x-\varepsilon y)\,\mathrm{d}y\Big| \\
&\leq \int_{[-1,1]} C|f(x-\varepsilon y)|\,\mathrm{d}y + \sum_{k=0}^\infty \int_{2^k<|y|\leq 2^{k+1}} C2^{-k(1+\delta)}|f(x-\varepsilon y)|\,\mathrm{d}y \\
&\leq \frac{2C}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon}|f(u)|\,\mathrm{d}u + \sum_{k=0}^\infty 2^{-k\delta}\frac{4C}{2\varepsilon 2^{k+1}}\int_{x-\varepsilon 2^{k+1}}^{x+\varepsilon 2^{k+1}}|f(u)|\,\mathrm{d}u \\
&\leq 2CMf(x) + \sum_{k=0}^\infty 2^{-k\delta}4CMf(x) = 2C\Big(1 + \frac{2}{1-2^{-\delta}}\Big)Mf(x).
\end{aligned}$$

□

5.11. **Lemma.** *If $\phi \in L^{p'}(\mathbb{R})$ and $f \in L^p(\mathbb{R})$ for $p \in [1,\infty)$, then*

$$\lim_{R\to\infty} \phi_R * f = 0$$

pointwise. If, in addition, $|\phi(x)| \leq C(1+|x|)^{-1-\delta}$, then the convergence also takes place in $L^p(\mathbb{R})$ if $p \in (1,\infty)$

Proof. By Hölder's inequality,

$$|\phi_R * f(x)| = \Big| \int \phi_R(y) f(x-y)\,\mathrm{d}y \Big| \leq \|\phi_R\|_{p'} \|f\|_p$$

and a change of variables shows that $\|\phi_R\|_{p'} = R^{-1/p}\|\phi\|_{p'} \to 0$ as $R \to \infty$.

By the additional assumption and Exercise 2.17(6), $|\phi_R * f| \leq C' Mf \in L^p(\mathbb{R})$, and hence the remaining claim follows from dominated convergence. □

5.12. **Lemma.** *Let $\phi \in L^1(\mathbb{R})$, $a := \int \phi(x)\,\mathrm{d}x$ and $f \in L^p(\mathbb{R})$. Then*

$$\lim_{\varepsilon\to 0} \phi_\varepsilon * f = af$$

in $L^p(\mathbb{R})$ for all $p \in [1,\infty)$ If, in addition $|\phi(x)| \leq C(1+|x|)^{-1-\delta}$, then the convergence also takes place pointwise a.e.

Proof. There holds

$$\phi_\varepsilon * f(x) - af(x) = \int \phi(y)[f(x-\varepsilon y) - f(x)]\,\mathrm{d}y,$$

$$\|\phi_\varepsilon * f - af\|_p \leq \int |\phi(y)| \cdot \|f(\cdot - \varepsilon y) - f\|_p\,\mathrm{d}y.$$

It remains to show that $\|f(\cdot - \varepsilon y) - f\|_p \to 0$ as $\varepsilon \to 0$, since the claim then follows from dominated convergence.

Let first $g \in C_c(\mathbb{R})$ (continuous with compact support). Then for all $0 < \varepsilon \leq |y|^{-1}$, $g(\cdot - \varepsilon y) - g$ is supported in a compact set K, bounded pointwise by $2\|g\|_\infty$ (and hence by $2\|g\|_\infty 1_K$) and converges pointwise to zero by the definition of continuity. Thus $\|g(\cdot - \varepsilon y) - g\|_p \to 0$ by dominated convergence. Such functions are dense in $L^p(\mathbb{R})$ for $p \in [1,\infty)$. Hence, given $f \in L^p(\mathbb{R})$ and $\delta > 0$, there is $g \in C_c(\mathbb{R})$ with $\|f - g\|_p < \delta$, and hence

$$\begin{aligned}\limsup_{\varepsilon\to 0} \|f(\cdot - \varepsilon y) - f\|_p &\leq \limsup_{\varepsilon\to 0} \Big(\|(f-g)(\cdot - \varepsilon y)\|_p + \|g(\cdot - \varepsilon y)\|_p + \|g - f\|_p \Big) \\ &= 2\|f - g\|_p < 2\delta.\end{aligned}$$

Since this holds for any $\delta > 0$, the conclusion is $\|f(\cdot - \varepsilon y) - f\|_p \to 0$, and the proof of the norm convergence is complete.

Concerning pointwise convergence, for $g \in C_c(\mathbb{R})$ one has

$$|\phi_\varepsilon * g(x) - ag(x)| \leq \int |\phi(y)| \cdot |g(x - \varepsilon y) - g(x)|\,\mathrm{d}y,$$

where the second factor is dominated by $2\|g\|_\infty$ and tends to zero everywhere by continuity. In general,

$$\limsup_{\varepsilon\to 0} |\phi_\varepsilon * f - af| \leq \limsup_{\varepsilon\to 0} \Big(M(f-g) + |\phi_\varepsilon * g - ag| + |ag - af| \Big) = M(f-g) + |a||f-g|.$$

Hence

$$|\{\limsup_{\varepsilon\to 0} |\phi_\varepsilon * f - af| > 2\delta\}| \leq |\{M(f-g) > \delta\}| + |\{|a||f-g| > \delta\}| \leq C\delta^{-p}\|f-g\|_p^p,$$

which can be made arbitrarily small. □

Now the proof of

$$Hf = -\frac{8}{\pi}\langle \text{Ш} \rangle f, \qquad \|Hf\|_p \leq C\|f\|_p$$

is complete.

5.13. **Exercises.**

1. Fix $x \in \mathbb{R}$ and consider the translated dyadic system $\mathscr{D}^0 + x = \bigcup_{j\in\mathbb{Z}}(\mathscr{D}^0_j + x)$ (note: same x on every level j), where $\mathscr{D}^0$ is the standard system. Find $\beta(x) \in \{0,1\}^{\mathbb{Z}}$ so that $\mathscr{D}^0 + x = \mathscr{D}^{\beta(x)}$. Observe that $\beta(x)$ has a certain special property and conclude that in general $\mathscr{D}^\beta$ cannot be represented in the form $\mathscr{D}^0 + x$.
2. In $\mathbb{R}^2$, consider the *dyadic squares* $\mathscr{D}_j := \{2^{-j}\big([0,1)^2 + (k,\ell)\big) : k,\ell \in \mathbb{Z}\}$, $j \in \mathbb{Z}$. There is one important difference compared to the one-dimensional case: the squares $I \in \mathscr{D}_j$ are now unions of four (rather than two) squares from $\mathscr{D}_{j+1}$.

   Find suitable intermediate partitions $\mathscr{D}_{j+1/2}$ of $\mathbb{R}^2$ so that each $I \in \mathscr{D}_i$ is a union of two sets from $\mathscr{D}_{i+1/2}$ for all $i \in \frac{1}{2}\mathbb{Z} := \{\ldots,-1,-\frac{1}{2},0,\frac{1}{2},1,\ldots\}$. Follow the computation in Section 5.3 to find a similar representation for $f \in L^p(\mathbb{R}^2)$. What do the Haar functions look like in this case? (Note: there are a couple of different ways to do this, but it suffices to provide one. No uniqueness here, your choice!)
3. Let $\xi_1,\ldots,\xi_n \in \mathbb{C}$. Consider the function $F : \mathbb{R}^n \to \mathbb{C}, (t_k)_{k=1}^n \mapsto \sum_{k=1}^n t_k\xi_k$. Prove that, on the unit cube $[-1,1]^n$, $|F(t)|$ attains its greatest value in one of the corners, $t \in \{-1,1\}^n$. (Hint: Write the numbers $\frac{1}{2}(\lambda_k+1) \in [0,1]$ with their binary expansion, $\frac{1}{2}(\lambda_k+1) = \sum_{j=1}^\infty b_{kj}2^{-j}$, where $b_{kj} \in \{0,1\}$. Then notice that $b_{kj} = \frac{1}{2}(\varepsilon_{kj}+1)$ for appropriate $\varepsilon_{kj} \in \{-1,1\}$, and it follows that $\lambda_k = \sum_{j=1}^\infty \varepsilon_{kj}2^{-j}$.)
4. Let $f_1,\ldots,f_n \in L^p(\mathbb{R})$ and $g_1,\ldots,g_n \in L^\infty(\mathbb{R})$. Let $\varepsilon_1,\ldots,\varepsilon_n$ be independent random signs with $\mathbb{P}(\varepsilon_k = -1) = \mathbb{P}(\varepsilon_k = +1) = 1/2$. Prove that

$$\int_{\mathbb{R}} \mathbb{E}\Big|\sum_{k=1}^n \varepsilon_k g_k(x) f_k(x)\Big|^p \,\mathrm{d}x \le \max_{1\le k\le n} \|g_k\|_\infty^p \int_{\mathbb{R}} \mathbb{E}\Big|\sum_{k=1}^n \varepsilon_k f_k(x)\Big|^p \,\mathrm{d}x.$$

   (Hint: use the previous exercise for each $x \in \mathbb{R}$ and a similar trick as in 5.4.)
5. Let $f \in L^p(\mathbb{R})$, $p \in (1,\infty)$, and $H_{ab}f$ be its truncated Hilbert transform. Consider the limit where $a,b \to 0$ in such a way that $a \le b \le 2a$. Prove that $|H_{ab}f| \le CMf$ for all such a,b, where M is the Hardy–Littlewood maximal operator. Show that $H_{ab}f \to 0$ pointwise a.e. in the considered limit. (Hint: prove the pointwise limit for continuous functions first and obtain the general case with the help of density and the pointwise domination by the maximal function.)

5.14. **References.** Petermichl's representation for the Hilbert transform as an average of the dyadic shifts is from [13]. The proof given here is somewhat different from Petermichl's original one, and was first presented in the first edition of this course in 2008 and published in [5]. The proof of the L^p boundedness of the dyadic shift, and the notation 'Ш', are taken from [12]. Although Petermichl's representation was here used just to derive the classical Hilbert transform boundedness on $L^p(\mathbb{R})$, its motivation comes from applications in the estimation of H, or some new operators derived from it, in more complicated situations like weighted spaces [11].

The L^p boundedness of the Hilbert transform is originally a classical result of M. Riesz [14]. Nowadays, there are many different proofs for this important theorem (which is perhaps most often handled in the framework of the Calderón–Zygmund theory of singular integrals), and even several different ways of getting it as a consequence of Burkholder's inequality. However, most of the martingale proofs rely on continuous-time notions like stochastic integrals and Brownian motion and would require more extensive preliminaries.

6. More on dyadic shifts

6.1. **Background.** In the last couple of years, dyadic shifts have played an important role in the theory of weighted norm inequalities, which is the topic of another course. Here, we will take a look into these operators from the point-of-view of martingale theory. A general dyadic shift of type (m,n) on $\mathbb{R}$ (a similar definition could be made on $\mathbb{R}^d$, $d>1$, as well) is defined as an operator of the form

$$S = \sum_{K \in \mathscr{D}} A_K, \qquad A_K f = \sum_{\substack{I,J \in \mathscr{D};\ I,J \subseteq K \\ |I| = 2^{-m}|K| \\ |J| = 2^{-n}|K|}} a_{IJK} \langle f, h_I \rangle h_J,$$

where the constants a_{IJK} satisfy

$$|a_{IJK}| \leq \frac{\sqrt{|I||J|}}{|K|}.$$

We observe that Petermichl's shift III is the special case with

$$A_K f = \langle f, h_K \rangle H_K = \frac{1}{\sqrt{2}} \langle f, h_K \rangle (h_{K_+} - h_{K_-}),$$

so that $(m, n) = (0, 1)$, and

$$a_{K,K_\pm,K} = \pm \frac{1}{\sqrt{2}} = \pm \frac{\sqrt{|K||K_\pm|}}{|K|}.$$

In a similar way as the Hilbert transform can be obtained as an average of Petermichl's shifts on different dyadic systems, quite general singular integral operators of the form

$$Tf(x) = \int_{\mathbb{R}} K(x, y) f(y) \, \mathrm{d}y$$

can be obtained as averages of the general dyadic shifts S; the general form of this result was first discovered in [6].

Now we look at the L^p boundedness of the shifts S.

6.2. **Separation of scales.** Let $N := \max\{m, n\} + 1$. Then we split (formally)

$$S = \sum_{K \in \mathscr{D}} A_K = \sum_{r=0}^{N-1} \sum_{k \in \mathbb{Z}} \sum_{K \in \mathscr{D}_{kN+r}} A_K.$$

The point of the splitting is this: Let r be fixed. If h_I appears in A_K for some $K \in \mathscr{D}_{kN+r} =: \mathscr{A}_k$, then h_I is constant of the dyadic intervals of length $\frac{1}{2}|I| = 2^{-1-m}|K| \geq 2^{-N}|K| = 2^{-(k+1)N-r}$, so in particular it is constant on all $K' \in \mathscr{D}_{(k+1)N+r} = \mathscr{A}_{k+1}$, and then on all $K' \in \mathscr{A}_{k'}$ for any $k' > k$. Similarly each h_J appearing in A_K is constant on all $K' \in \mathscr{A}_{k'}$ with $k' > k$. On the other hand, clearly the average of h_I and h_J on K, and on all equal or larger intervals is zero. Thus we are in a position to consider a martingale difference sequence with respect to the filtration $\mathscr{F}_k := \sigma(\mathscr{A}_k)$, $k \in \mathbb{Z}$. Note that each $K \in \mathscr{A}_k$ is the union of finitely many (in fact, 2^N) intervals $K' \in \mathscr{A}_{k+1}$.

6.3. **Decoupling: preliminary considerations.** 'Decoupling' vaguely refers to the replacement of an object by a new one, which has more independence. A relatively simple form of this was given by the version of Burkholder's inequality with the independent random signs ε_k:

$$\Big\| \sum_k d_k \Big\|_p \eqsim \Big\| \sum_k \varepsilon_k d_k \Big\|_p;$$

this was enough to get an L^p estimate for Petermichl's shift III. We are now looking for a somewhat more elaborate version.

Let us more generally consider a martingale difference sequence d_k adapted to $\mathscr{F}_k$ as above. Thus

$$d_k = \sum_{K \in \mathscr{A}_{k-1}} 1_K d_k = \sum_{K \in \mathscr{A}_{k-1}} 1_K \sum_{\substack{A \in \mathscr{A}_k \\ A \subseteq K}} 1_A d_k, = \sum_{K \in \mathscr{A}_{k-1}} 1_K \sum_{\substack{A \in \mathscr{A}_k \\ A \subseteq K}} 1_A \langle d_k \rangle_A,$$

where the $\mathscr{F}_k$-measurability of d_k means that d_k is a constant $\langle d_k \rangle_A$ on $A \in \mathscr{A}_k$. The martingale difference property $\mathbb{E}[d_k | \mathscr{F}_{k-1}] = 0$ means that

$$\int_K d_k \, \mathrm{d}x = \sum_{\substack{K' \in \mathscr{A}_k \\ K' \subseteq K}} \langle d_k \rangle_{K'} |K'| = 0.$$

The above $d_k = d_k(x)$ is a function of $x \in \mathbb{R}$. Clearly we could also view it as a function $d_k = d_k(x, y)$ of $x \in \mathbb{R}$ and y in some new space Ω, where the dependence on y is trivial. But next we would like to replace d_k by a 'decoupled' version $\tilde{d}_k(x, y)$, where part of the original dependence on x is pushed into the new variable y, in such a way that the dependence on x is simpler than before.

Consider a fixed $K \in \mathscr{D}_K$ first. Let μ_K be the Lebesgue measure on K divided by $|K|$. Note that the functions

$$1_K(x)d_k(x)1_K(y) = \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K}} \langle d_k\rangle_A 1_A(x)1_B(y)$$

and

$$1_K(x)d_k(y)1_K(y) = \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K}} \langle d_k\rangle_B 1_A(x)1_B(y)$$

are equally distributed (i.e., take the same values in sets of equal measure).

Now consider two new functions, first

$$\begin{aligned} u_K(x,y) &:= \frac{1}{2}(1_K(x)d_k(x)1_K(y) + 1_K(x)d_k(y)1_K(y)) \\ &= \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K}} \frac{1}{2}(\langle d_k\rangle_A + \langle d_k\rangle_B)1_A(x)1_B(y) \\ &= \sum_{\substack{A\in\mathscr{A}_k \\ A\subseteq K}} \langle d_k\rangle_A 1_{A\times A}(x,y) \\ &\qquad + \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K; A<B}} \frac{1}{2}(\langle d_k\rangle_A + \langle d_k\rangle_B)1_{A\times B\cup B\times A}(x,y), \end{aligned}$$

where in the last step we introduced some order among the finite family

$$\{A \in \mathscr{A}_k : A \subseteq K\} = \{A_i\}_{i=1}^{I(K)},$$

and defined $A < B$ if and only if $A = A_i$, $B = A_j$, and $i < j$.

The second new function is

$$\begin{aligned} v_K(x,y) &:= \frac{1}{2}(1_K(x)d_k(x)1_K(y) - 1_K(x)d_k(y)1_K(y)) \\ &= \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K}} \frac{1}{2}(\langle d_k\rangle_A - \langle d_k\rangle_B)1_A(x)1_B(y) \\ &= \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K; A<B}} \frac{1}{2}(\langle d_k\rangle_A - \langle d_k\rangle_B)(1_{A\times B}(x,y) - 1_{B\times A}(x,y)). \end{aligned}$$

Now clearly

$$\begin{aligned} 1_K(x)d_k(x)1_K(y) &= u_K(x,y) + v_K(x,y), \\ 1_K(x)d_k(y)1_K(y) &= u_K(x,y) - v_K(x,y), \end{aligned}$$

but the key observation is that this decomposition realizes the second function as a martingale transform of the first one.

6.4. **A new measure space.** For each $K \in \mathscr{A}_{k-1}$ and $k \in \mathbb{Z}$, let Ω_K be the measure space K equipped with the σ-algebra $\mathscr{F}_K := \mathscr{A}_k \cap K$ and the normalized Lebesgue measure $\mu_K = \mathrm{d}x/|K|$. We consider the big product measure space

$$\Omega := \prod_{K\in\mathscr{A}} \Omega_K \qquad \Big(\text{where} \qquad \mathscr{A} := \bigcup_{k\in\mathbb{Z}} \mathscr{A}_k\Big)$$

of sequences $y = (y_K)_{K\in\mathscr{A}}$ indexed by the (countably many) sets $K \in \mathscr{A}$. The product σ-algebra is defined as

$$\mathscr{Y} := \sigma(\tilde{\mathscr{Y}}), \quad \text{where} \quad \tilde{\mathscr{Y}} := \Big\{ \prod_{K\in\mathscr{A}}^{*} A_K : A_K \in \mathscr{F}_K \Big\},$$

and the $*$ in the product means that only finitely many terms A_K are nontrivial, i.e., $A_K \neq \Omega_K$. The product measure on $(\Omega, \mathscr{Y})$ is the unique measure μ on $\mathscr{Y}$ such that

$$\mu\Big(\prod_{K\in\mathscr{A}} A_K \Big) = \prod_{K\in\mathscr{A}} \mu(A_K) \qquad \forall \prod_{K\in\mathscr{A}} A_K \in \tilde{\mathscr{Y}}.$$

(The existence of such a measure, i.e., the extension of μ form $\tilde{\mathscr{Y}}$ to all of $\mathscr{Y}$ is a nontrivial result, but we take it for granted here.)

We now consider the measure space $\mathbb{R} \times \Omega$. Its points will be denoted by (x, y), where further $y = (y_K)_{K\in\mathscr{A}}$. It is also convenient to split y as $y = (y_{<k}, y_k, y_{>k})$, where

$$y_{<k} := (y_K)_{K\in\bigcup_{j<k}\mathscr{A}_{j-1}}, \quad y_k := (y_K)_{K\in\mathscr{A}_{k-1}}, \quad y_{>k} := (y_K)_{K\in\bigcup_{j>k}\mathscr{A}_{j-1}}.$$

We also define the related space $\Omega_{<k} := \prod_{K\in\bigcup_{j<k}\mathscr{A}_{j-1}} K$ (Ω_k and $\Omega_{>k}$ similarly) and the σ-algebras

$$\mathscr{Y}_{<k} := \sigma(\tilde{\mathscr{Y}}_{<k}), \quad \text{where} \quad \tilde{\mathscr{Y}}_{<k} := \Big\{ \prod_{K\in\bigcup_{j<k}\mathscr{A}_j}^{*} A_K : A_K \in \mathscr{F}_K \Big\}$$

($\mathscr{Y}_k$ and $\mathscr{Y}_{>k}$ similarly), where $\prod^*$ has the same meaning as before. It follows that for every k there are the splittings $\Omega = \Omega_{<k} \times \Omega_k \times \Omega_{>k}$ as well as $\mathscr{Y} = \sigma(\mathscr{Y}_{<k} \times \mathscr{Y}_k \times \mathscr{Y}_{>k})$. We also have the partial product measures $\mu_{<k}, \mu_k, \mu_{>k}$.

6.5. **Filtration on the new measure space.** We have

$$d_k(x) = d_k(x, y) = \sum_{K\in\mathscr{A}_{k-1}} 1_K(x) d_k(x) 1_K(y_K) = \sum_{K\in\mathscr{A}_{k-1}} (u_K(x, y_K) + v_K(x, y_K))$$

and

$$\tilde{d}_k(x, y) := \sum_{K\in\mathscr{A}_{k-1}} 1_K(x) d_k(y_K) 1_K(y_K) = \sum_{K\in\mathscr{A}_{k-1}} (u_K(x, y_K) - v_K(x, y_K)); \tag{6.6}$$

thus

$$d_k = u_k + u_{k+1/2}, \qquad \tilde{d}_k = u_k - u_{k+1/2},$$

where

$$\begin{aligned} u_k(x, y) &:= \sum_{K\in\mathscr{A}_{k-1}} u_K(x, y_K) \\ &= \sum_{K\in\mathscr{A}_{k-1}} \Big[\sum_{\substack{A\in\mathscr{A}_k \\ A\subseteq K}} \langle d_k\rangle_A 1_{A\times A}(x, y_K) \\ &\qquad + \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K \\ A<B}} \frac{1}{2}(\langle d_k\rangle_A + \langle d_k\rangle_B) 1_{A\times B\cup B\times A}(x, y_K) \Big] \\ u_{k+1/2}(x, y) &:= \sum_{K\in\mathscr{A}_{k-1}} v_K(x, y_K) \\ &= \sum_{K\in\mathscr{A}_{k-1}} \sum_{\substack{A,B\in\mathscr{A}_k \\ A,B\subseteq K \\ A<B}} \frac{1}{2}(\langle d_k\rangle_A - \langle d_k\rangle_B)(1_{A\times B} - 1_{B\times A})(x, y_K). \end{aligned} \tag{6.7}$$

Now, we would like to view $(u_k)_{k\in\frac{1}{2}\mathbb{Z}}$ as a martingale difference sequence on the product measure space $(\mathbb{R} \times \Omega, \sigma(\mathscr{B} \times \mathscr{Y}), \mathrm{d}x \times \mu)$. Although the idea is intuitively simple, it gets notationally complicated, so let us informally sketch the idea first. The key features are as follows:

(a) both u_k and $u_{k+1/2}$ are $\mathscr{F}_k$-measurable with respect to x, and $\mathscr{Y}_{\leq k}$-measurable (in fact, even $\mathscr{Y}_k$-measurable, but we need an increasing σ-algebra as a function of k for a filtration) with respect to y.
(b) for every $K \in \mathscr{A}_{k-1}$, the restriction $1_K(x)u_k(x,y)$ depends on x and y_K "in a symmetric way";
(c) as x varies over a $\mathscr{F}_{k-1}$-measurable set (a union of some $K \in \mathscr{A}_{k-1}$) and y over a $\mathscr{Y}_{\leq k-1}$-measurable set (so that y_K for $K \in \mathscr{A}_{k-1}$ must vary over its entire domain K), we have

$$\begin{aligned} &\frac{1}{|K|}\int_K\int_K u_k(x,y_K)\,\mathrm{d}x\,\mathrm{d}y_K \\ &= \frac{1}{|K|}\Big[\sum_{\substack{A\in\mathscr{A}_k\\ A\subseteq K}}\langle d_k\rangle_A|A||A| + \sum_{\substack{A,B\in\mathscr{A}_k\\ A,B\subseteq K\\ A<B}}\frac{1}{2}(\langle d_k\rangle_A+\langle d_k\rangle_B)\cdot 2|A||B|\Big] \\ &= \frac{1}{|K|}\sum_{\substack{A\in\mathscr{A}_k\\ A\subseteq K}}\langle d_k\rangle_A|A|\sum_{\substack{B\in\mathscr{A}_k\\ B\subseteq K}}|B| = \sum_{\substack{A\in\mathscr{A}_k\\ A\subseteq K}}\langle d_k\rangle_A|A| = \int_K d_k(x)\,\mathrm{d}x = 0, \end{aligned} \tag{6.8}$$

since d_k itself is a martingale difference;
(d) for (x, y_K) on the symmetric sets $A\times B\cup B\times A$, the function $u_{k+1/2}$ takes equal positive and negative values on two halves of the set, so it averages to zero.

Here is the idea (to be made more precise in a moment): From (a) and (b) it follows that $u_{k+1/2}$ is measurable with respect to the σ-algebra $\mathscr{U}_{k+1/2} = \sigma(\mathscr{F}_k \times \mathscr{Y}_{\leq k}\times\Omega_{>k})$, while u_k is measurable with respect to a smaller "symmetric part" $\mathscr{U}_k \subset \mathscr{U}_{k+1/2}$. By (c), it follows that $\mathbb{E}[u_k|\mathscr{U}_{k-1/2}] = 0$ and (d) implies that $\mathbb{E}[u_{k+1/2}|\mathscr{U}_k] = 0$.

Now, we want to express this in a systematic way, and we make the following definitions:

$$\begin{aligned} \mathscr{U}_k &:= \Big\{\bigcup_{K\in\mathscr{A}_{k-1}}\bigcup_{\substack{A,B\in\mathscr{A}_k\\ A,B\subseteq K}}\Big[A\times Q_{AB}\times B\cup B\times Q_{AB}\times A\Big] \\ &\qquad\qquad\times\prod_{\substack{H\in\mathscr{A}_{k-1}\\ H\neq K}} H\times\Omega_{>k} : Q_{AB}\in\mathscr{Y}_{<k}\Big\}, \\ \mathscr{U}_{k+1/2} &:= \sigma(\mathscr{F}_k\times\mathscr{Y}_{\leq k}\times\Omega_{>k}) = \Big\{\bigcup_{A\in\mathscr{A}_k} A\times Q_A\times\Omega_{>k} : Q_A\in\mathscr{Y}_{\leq k}\Big\}; \end{aligned} \tag{6.9}$$

These are collections of subsets of $\mathbb{R}\times\Omega$, and we implicitly use the splitting $(x,y) = (x, y_{<k}, y_k, y_{>k})$, where moreover $y_k = (y_K; (y_H)_{H\in\mathscr{A}_{k-1}\setminus\{K\}})$. One can check that both $\mathscr{U}_k$ and $\mathscr{U}_{k+1/2}$ are σ-algebras. By substituting $k-1$ in place of k, it follows that

$$\mathscr{U}_{k-1/2} := \Big\{\bigcup_{K\in\mathscr{A}_{k-1}} K\times Q_K\times\Omega_k\times\Omega_{>k} : Q_K\in\mathscr{Y}_{<k}\Big\}.$$

Writing $\Omega_k = K\times\prod_{H\in\mathscr{A}_{k-1}\setminus\{K\}} H$, and taking $Q_{AB} = Q_K$ for all $A, B\in\mathscr{A}_k$ with $A, B\subseteq K$, we see that $\mathscr{U}_{k-1/2}\subseteq\mathscr{U}_k$. Observing that $Q_{AB}\times B\times\prod_{H\in\mathscr{A}_{k-1}\setminus\{K\}} H\in\mathscr{Y}_{\leq k}$ for $Q_{AB}\in\mathscr{Y}_k$ and $B\in\mathscr{F}_K$, $K\in\mathscr{A}_{k-1}$, we see that $\mathscr{U}_k\subseteq\mathscr{U}_{k+1/2}$. Thus $(\mathscr{U}_k)_{k\in\frac{1}{2}\mathbb{Z}}$ is a filtration, and it is easy to see that u_k is $\mathscr{U}_k$-measurable for all $k\in\frac{1}{2}\mathbb{Z} := \{\ldots,-\frac{3}{2},-1,-\frac{1}{2},0,\frac{1}{2},1,\frac{3}{2},\ldots\}$.

6.10. **The martingale difference property.** It remains to check that $\mathbb{E}[u_k|\mathscr{U}_{k-1/2}] = 0$ for all $k\in\frac{1}{2}\mathbb{Z}$.

Let us first show that $\mathbb{E}[u_k|\mathscr{U}_{k-1/2}] = 0$ for all $k\in\mathbb{Z}$: We integrate u_k over any set of the form $K\times Q_K\times\Omega_k\times\Omega_{>k}$, where $K\in\mathscr{A}_{k-1}$ and $Q_K\in\mathscr{Y}_{<k}$ (finite unions of such sets form a $\mathscr{U}_{k-1/2}$-ideal which contains a countable cover).

$$\int_{K\times Q_K\times\Omega_k\times\Omega_{>k}} u_k(x,y)\,\mathrm{d}x\,\mathrm{d}\mu(y) = \mu_{<k}(Q_K)\int_{K\times K} u_k(x,y_K)\,\mathrm{d}x\frac{\mathrm{d}y_K}{|K|},$$

where we used the product-measure structure $\mathrm{d}\mu(y) = \mathrm{d}\mu_{<k}(y_{-k})\,\mathrm{d}\mu_k(y_k)\,\mathrm{d}\mu_{>k}(y_{>k})$, and the fact that $1_K(x)u_k(x,y)$ only depends on the coordinate y_K of y. In (6.8) we already checked that the right side is zero, verifying the martingale difference property of u_k for $k \in \mathbb{Z}$.

It remains to check that $\mathbb{E}[u_{k+1/2}|\mathscr{U}_k] = 0$ for $k \in \mathbb{Z}$, thus evaluate the integral of $u_{k+1/2}$ over any set of the form $[A \times Q_{AB} \times B \cup B \times Q_{AB} \times A] \times \prod_{H \in \mathscr{A}_{k-1} \setminus \{K\}} H \times \Omega_{>k}$ for $A, B \in \mathscr{A}_k$ with $A, B \subseteq K \in \mathscr{A}_{k-1}$ and $Q_{AB} \in \mathscr{Y}_{<k}$. Using again the product-measure structure and the dependence of $1_K(x)u_{k+1/2}(x,y)$ only on y_K, we find that this integral is equal to

$$\mu_{<k}(Q_K) \int_{A\times B \cup B \times A} u_{k+1/2}(x, y_K)\,\mathrm{d}x \frac{\mathrm{d}y_K}{|K|},$$

which is zero, since $u_{k+1/2}(x, y_K)$ takes opposite values on the sets $A \times B$ and $B \times A$ of equal measure.

Altogether, we have confirmed that:

6.11. **Proposition.** *The functions* $(u_k)_{k\in\frac{1}{2}\mathbb{Z}}$ *defined in* (6.7) *form a martingale difference sequence adapted to the filtration* $(\mathscr{U}_k)_{k\in\frac{1}{2}\mathbb{Z}}$ *defined in* (6.9).

From Burkholder's inequality we can now derive the following consequence:

6.12. **Theorem** (Decoupling of martingale differences)**.** *Let* d_k *be a martingale difference sequence adapted to a filtration* $\mathscr{F}_k$*, where each* $\mathscr{F}_k$ *is generated by a countable collection of atoms* $\mathscr{A}_k$*. Then*

$$\frac{1}{\beta_p}\Big\| \sum_k d_k \Big\|_{L^p(\mathbb{R})} \leq \Big\| \sum_k \eta_k \tilde{d}_k \Big\|_{L^p(\mathbb{R}\times\Omega)} \leq \beta_p \Big\| \sum_k d_k \Big\|_{L^p(\mathbb{R})}, \tag{6.13}$$

where the decoupled sequence $\tilde{d}_k$ *is defined in* (6.6)*,* Ω *is the probability space related to the variable* $y = (y_k)_{k\in\mathbb{Z}} = (y_K)_{K\in\mathscr{A}}$*, and the* $\eta_k = \pm 1$ *are arbitrary signs.*

In particular, we also have the estimate

$$\frac{1}{\beta_p}\Big\| \sum_k d_k \Big\|_{L^p(\mathbb{R})} \leq \Big\| \sum_k \varepsilon_k \tilde{d}_k \Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega')} \leq \beta_p \Big\| \sum_k d_k \Big\|_{L^p(\mathbb{R})}, \tag{6.14}$$

where Ω' *is another probability space supporting the independent random signs* ε_k*.*

Proof. Using the functions $u_k, u_{k+1/2}$, we have

$$\sum_{k\in\mathbb{Z}} \eta_k \tilde{d}_k = \frac{1}{2}\sum_{k\in\mathbb{Z}}(\eta_k u_k - \eta_k u_{k+1/2}) = \frac{1}{2}\sum_{k\in\frac{1}{2}\mathbb{Z}} \zeta_k u_k, \qquad \Big(\zeta_k = \eta_k, \quad \zeta_{k+1/2} = -\eta_k \quad \forall k \in \mathbb{Z}\Big),$$

$$\sum_{k\in\mathbb{Z}} d_k = \frac{1}{2}\sum_{k\in\mathbb{Z}}(u_k + u_{k+1/2}) = \frac{1}{2}\sum_{k\in\frac{1}{2}\mathbb{Z}} u_k.$$

Since $(u_k)_{k\in\frac{1}{2}\mathbb{Z}}$ is a martingale difference sequence, we see that $\sum_{k\in\frac{1}{2}\mathbb{Z}} \zeta_k u_k$ is the martingale transform of $\sum_{k\in\frac{1}{2}\mathbb{Z}} u_k$ by the transforming sequence $(\zeta_k)_{k\in\frac{1}{2}\mathbb{Z}}$, and since $\zeta_k^2 = 1$, we see that the converse holds as well. Thus (6.13) is a direct consequence of Burkholder's inequality.

The estimate (6.14) follows from (6.13) by taking $\eta_k := \varepsilon_k(\omega')$ and integrating the pth power of (6.13) over $\omega' \in \Omega'$. □

6.15. **Back to dyadic shifts.** We now consider a dyadic shift with scales separated, i.e.,

$$Sf = \sum_{k\in\mathbb{Z}} \sum_{K\in\mathscr{D}_{kN+r}=:\mathscr{A}_k} A_K f, \qquad A_K f = \sum_{\substack{I,J\in\mathscr{D}; I,J\subseteq K \\ |I|=2^{-m}|K| \\ |J|=2^{-n}|K|}} a_{IJK}\langle f, h_I\rangle h_J,$$

where $m, n, r \in\in \{0, 1, \ldots, N-1\}$ are fixed. We can also write

$$A_K f(x, x') = \frac{1}{|K|}\int_K a_K(x, x') f(x')\,\mathrm{d}x',$$

where

$$a_K(x,x') := |K| \sum_{\substack{I,J\in\mathscr{D};I,J\subseteq K\\ |I|=2^{-m}|K|\\ |J|=2^{-n}|K|}} a_{IJK} h_I(x') h_J(x)$$

satisfies $|a_K(x,x')| \le 1$, thanks to $|K||a_{IJK}| \le \sqrt{|I|\cdot|J|}$ and the scaling of the Haar functions, $|h_I| \le 1_I/\sqrt{|I|}$.

Let us also denote

$$D_K^{(N)} f := \sum_{\substack{I\in\mathscr{D};I\subseteq K\\ |I|>2^{-N}|K|}} \langle f, h_I\rangle h_I,$$

and observe that

$$f = \sum_{K\in\mathscr{A}:=\bigcup_{k\in\mathbb{Z}}\mathscr{A}_k} D_K^{(N)}, \qquad A_K f = A_K D_K^{(N)} f.$$

Finally, notice that both

$$e_k := \sum_{K\in\mathscr{A}_k} A_K f, \qquad d_k := \sum_{K\in\mathscr{A}_k} D_K^{(N)} f,$$

are martingale difference sequences with respect to the filtration $\mathscr{F}_k := \sigma(\mathscr{A}_k)$, which has the form that we have considered above.

A first application of Theorem 6.12 shows that

$$\|Sf\|_{L^p(\mathbb{R})} = \Big\| \sum_k e_k \Big\|_{L^p(\mathbb{R})} \le \beta_p \Big\| \sum_k \varepsilon_k \tilde{e}_k \Big\|_{L^p(\mathbb{R}\times\Omega)},$$

where

$$\begin{aligned}\tilde{e}_k(x,y) = \sum_{K\in\mathscr{A}_k} 1_K(x) A_K D_K^{(N)} f(y_K) &= \sum_{K\in\mathscr{A}_k} 1_K(x) \frac{1}{|K|}\int_K a_K(y_K,x') D_K^{(N)} f(x')\,\mathrm{d}x' \\ &= \sum_{K\in\mathscr{A}_k} 1_K(x)\int_\Omega a_K(y_K,y'_K) D_K^{(N)} f(y'_K)\,\mathrm{d}\mu(y').\end{aligned}$$

Thus

$$\begin{aligned}\|Sf\|_{L^p(\mathbb{R})} &\le \beta_p \Big\| \int_\Omega \Big(\sum_k \varepsilon_k \sum_{K\in\mathscr{A}_k} 1_K(x) a_K(y_K,y'_K) D_K^{(N)} f(y'_K)\Big)\,\mathrm{d}\mu(y')\Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega',\,\mathrm{d}x\,\mathrm{d}\mu(y)\,\mathrm{d}\varepsilon)} \\ &\le \beta_p \int_\Omega \Big\| \sum_k \varepsilon_k \sum_{K\in\mathscr{A}_k} 1_K(x) a_K(y_K,y'_K) D_K^{(N)} f(y'_K)\Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega',\,\mathrm{d}x\,\mathrm{d}\mu(y)\,\mathrm{d}\varepsilon)} \mathrm{d}\mu(y') \\ &\le \beta_p \Big\| \sum_k \varepsilon_k \sum_{K\in\mathscr{A}_k} 1_K(x) a_K(y_K,y'_K) D_K^{(N)} f(y'_K)\Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega'\times\Omega,\,\mathrm{d}x\,\mathrm{d}\mu(y)\,\mathrm{d}\varepsilon\,\mathrm{d}\mu(y'))}.\end{aligned}$$

Next, we use Fubini's theorem and consider the L^p integral norm over the other variables for a fixed $x \in \mathbb{R}$. Then the summation over $K \in \mathscr{A}_k$ disappears, since there is exactly one $K = K(k,x)$ such that $x \in K \in \mathscr{A}_k$ for every $k \in \mathbb{Z}$. Thus the integral is of the form considered in Exercise 4 in Sec. 5.13, and we can use the conclusion of that exercise to remove the functions a_K with $\|a_K\|_\infty \le 1$:

$$\le \beta_p \Big\| \sum_k \varepsilon_k \sum_{K\in\mathscr{A}_k} 1_K(x) D_K^{(N)} f(y'_K)\Big\|_{L^p(\mathbb{R}\times\Omega'\times\Omega,\,\mathrm{d}x\,\mathrm{d}\varepsilon\,\mathrm{d}\mu(y'))} = \beta_p \Big\| \sum_k \varepsilon_k d_k \Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega')}.$$

We also removed the integration over the probability space $(\Omega, \mathrm{d}\mu(y))$, since there is no more y-dependence inside the norm.

Now we are in a position to apply the other side of the decoupling estimate of Theorem 6.12:

$$\beta_p \Big\| \sum_k \varepsilon_k d_k \Big\|_{L^p(\mathbb{R}\times\Omega\times\Omega')} \le \beta_p^2 \Big\| \sum_k d_K \Big\|_{L^p(\mathbb{R})} = \beta_p^2 \|f\|_{L^p(\mathbb{R})}.$$

We have thus shown that:

6.16. **Theorem.** *Let S be a dyadic shift with scales separated. Then*

$$\|Sf\|_p \leq \beta_p^2 \|f\|_p, \qquad p \in (1, \infty).$$

This bound is independent of the parameter N of the shift. If S is a general dyadic shift without the separation of scales, we can use the splitting in Sec. 6.2 and the triangle inequality to deduce that

$$\|Sf\|_p \leq N\beta_p^2 \|f\|_p.$$

6.17. **References.** Dyadic shifts, in the generality considered here, were introduced by Lacey, Petermichl and Reguera [8]. They played an important role in the resolution of the so-called A_2 conjecture [6] on weighted norm inequalities (see the corresponding course for more details). The present martingale point-of-view to the dyadic shifts has not been presented before, although parts of it can be implicitly seen in [7]. There are other approaches to this estimate for real-valued functions that we have considered, but the present decoupling point-of-view is useful in view of some vector-valued generalizations. The Decoupling Theorem 6.12 is essentially due to McConnell [9] in a more abstract formulation; the more concrete version here is from [7]. The present proof, where this estimate is derived directly from Burholder's inequality, has not been presented before.

Appendix A. English–Finnish-vocabulary

adapted – mukautettu
conditional – ehdollinen
dilation – venytys
dyadic – dyadinen
expectation – odotusarvo
filtration – suodatus
independent – riippumaton
martingale – martingaali
maximal function – maksimaalifunktio
predictable – ennustettava
shift – siirto
stopping time – pysäytysaika
transform – muunnos
translation – siirto
truncated – katkaistu

References

[1] D. L. Burkholder. Martingale transforms. *Ann. Math. Statist.*, 37:1494–1504, 1966.
[2] D. L. Burkholder. Boundary value problems and sharp inequalities for martingale transforms. *Ann. Probab.*, 12(3):647–702, 1984.
[3] D. L. Burkholder. Explorations in martingale theory and its applications. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, volume 1464 of *Lecture Notes in Math.*, pages 1–66. Springer, Berlin, 1991.
[4] J. L. Doob. *Stochastic processes.* John Wiley & Sons Inc., New York, 1953.
[5] Tuomas Hytönen. On Petermichl's dyadic shift and the Hilbert transform. *C. R. Math. Acad. Sci. Paris*, 346(21-22):1133–1136, 2008.
[6] Tuomas Hytönen. The sharp weighted bound for general Calderón-Zygmund operators. *Ann. of Math. (2)*, 175(3):1473–1506, 2012.
[7] Tuomas Hytönen. The vector-valued non-homogeneous *Tb* theorem. *Int. Math. Res. Not.*, to appear. Preprint, arXiv:0809.3097.
[8] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera. Sharp A_2 inequality for Haar shift operators. *Math. Ann.*, 348(1):127–141, 2010.
[9] Terry R. McConnell. Decoupling and stochastic integration in UMD Banach spaces. *Probab. Math. Statist.*, 10(2):283–295, 1989.
[10] Tao Mei. BMO is the intersection of two translates of dyadic BMO. *C. R. Math. Acad. Sci. Paris*, 336(12):1003–1006, 2003.
[11] S. Petermichl. The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical A_p characteristic. *Amer. J. Math.*, 129(5):1355–1375, 2007.
[12] S. Petermichl and S. Pott. A version of Burkholder's theorem for operator-weighted spaces. *Proc. Amer. Math. Soc.*, 131(11):3457–3461 (electronic), 2003.
[13] Stefanie Petermichl. Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(6):455–460, 2000.
[14] Marcel Riesz. Sur les fonctions conjuguées. *Math. Z.*, 27(1):218–244, 1928.
[15] Hitoshi Tanaka and Yutaka Terasawa. Positive operators and maximal operators in a filtered measure space. *Preprint*, 2012. arXiv:1204.6600.
[16] David Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.