MARTINGALES AND HARMONIC ANALYSIS

TUOMAS HYTÖNEN

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-algebera 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

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

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 $B \subseteq \mathbb{R}$.

Denote by $\mathscr{F}^0$ the collection of sets in $\mathscr{F}$ with finite measure, i.e., $\mathscr{F}^0 := \{E \in \mathscr{F} : \mu(E) < \infty\}$. The measure space $(\Omega, \mathscr{F}, \mu)$ is called σ-finite if there exist sets $E_i \in \mathscr{F}^0$ such that $\bigcup_{i=0}^{\infty} E_i = \Omega$. If needed, these sets may be chosen to additionally satisfy either (a) $E_i \subseteq E_{i+1}$ or (b) $E_i \cap E_j = \varnothing$ whenever $i \neq j$. Part (a) follows by taking $E_i' := \bigcup_{j=0}^{i} E_i$, and part (b) by setting $E_i'' := E_i' \setminus E_{i-1}'$, where $E_{-1}' := \varnothing$. Unless otherwise stated, it is always assumed in the sequel that $(\Omega, \mathscr{F}, \mu)$ is σ-finite.

An $\mathscr{F}$-measurable function $f : \Omega \to \mathbb{R}$ is called σ-integrable if it is integrable on all sets of finite measure, i.e., if $1_E f \in L^1(\mathscr{F}, \mu)$ for all $E \in \mathscr{F}^0$. Denote the collection of all such functions by $L^1_\sigma(\mathscr{F}, \mu)$.

1.2. **Lemma.** *If $f \in L^1_\sigma(\mathscr{F}, \mu)$ satisfies $\int_E f \,\mathrm{d}\mu \geq 0$ for all $E \in \mathscr{F}^0$, then $f \geq 0$ a.e. (almost everywhere). The same is true if $\geq$ is replaced by $\leq$ or $=$.*

Proof. Let $F_i := \{f < -1/i\} \in \mathscr{F}$ and E_j be one of the sets in the definition of σ-finiteness, which are now chosen to be disjoint. Then

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

Hence $\mu(F_i \cap E_j) = 0$, and summing up over $j \in \mathbb{N}$ it follows that $\mu(F_i) = 0$. Since $\{f < 0\} = \bigcup_{i=1}^{\infty} F_i$, one sees 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 $x \leq 0$. $\square$

1.3. **Sub-σ-algebra and the conditional expectation with respect to it.** 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.

Version: November 8, 2012.

A function $g \in L^1_\sigma(\mathscr{G},\mu)$ is called the conditional expectation of $f \in L^1_\sigma(\mathscr{F},\mu)$ with respect to $\mathscr{G}$ if there holds

$$\int_G f \,\mathrm{d}\mu = \int_G g \,\mathrm{d}\mu \qquad \forall G \in \mathscr{G}^0.$$

This means that g is (in a certain sense) the best possible $\mathscr{G}$-measurable approximation for f.

Observe that the conditional expectation of f with respect to $\mathscr{G}$, if it exists, is unique (a.e.). Namely, if $g_1, g_2 \in L^1_\sigma(\mathscr{G},\mu)$ were both conditional expectations of f, the function $g := g_1 - g_2 \in L^1_\sigma(\mathscr{G},\mu)$ would satisfy $\int_G g \,\mathrm{d}\mu = 0$ for all $G \in \mathscr{G}^0$. By Lemma 1.2 this implies $g = 0$ a.e., hence $g_1 = g_2$ a.e.

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.

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

1.5. **Simple observations.** If $g \in L^1_\sigma(\mathscr{G},\mu)$, 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. If $f_1, f_2 \in L^1_\sigma(\mathscr{F},\mu)$, and they have conditional expectations $\mathbb{E}[f_i|\mathscr{G}]$, then for all constants α_1, α_2 also the function $\alpha_1 f_1 + \alpha_2 f_2$ has a conditional expectation, and it is

$$\mathbb{E}[\alpha_1 f_1 + \alpha_2 f_2|\mathscr{G}] = \alpha_1 \mathbb{E}[f_1|\mathscr{G}] + \alpha_2 \mathbb{E}[f_2|\mathscr{G}].$$

These observations follow easily directly from the definition of the conditional expectation.

If the pointwise (a.e.) inequality $f_1 \leq f_2$ holds, then also $\mathbb{E}[f_1|\mathscr{G}] \leq \mathbb{E}[f_2|\mathscr{G}]$. This follows from the fact that for all $G \in \mathscr{G}^0$ one has

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

and Lemma 1.2.

This implies that if $f \in L^1_\sigma(\mathscr{F},\mu)$, and both $\mathbb{E}[f|\mathscr{G}]$ and $\mathbb{E}[|f|\,|\mathscr{G}]$ exist, there holds

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

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

1.6. **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 1.4 the conditional expectations $g_n := \mathbb{E}[f_n|\mathscr{G}]$ and $\mathbb{E}[|f_n|\,|\mathscr{G}]$ exist and belong to $L^2(\mathscr{G},\mu)$.

If $G \in \mathscr{G}^0$, then

$$\int_G |g_n| \,\mathrm{d}\mu = \int_G |\mathbb{E}[f_n|\mathscr{G}]| \,\mathrm{d}\mu \leq \int_G \mathbb{E}[|f_n|\,|\mathscr{G}] \,\mathrm{d}\mu = \int_G |f_n| \,\mathrm{d}\mu.$$

By choosing disjoint $G = E_k$ from the definition of σ-finiteness of the measure space $(\Omega, \mathscr{G}, \mu)$ and summing over $k \in \mathbb{N}$, it follows that $\|g_n\|_1 \leq \|f_n\|_1$, and hence $g_n \in L^1(\mathscr{G},\mu)$.

Repeating the previous computation with g_n replaced by $g_n - g_m$, it similarly follows that $\|g_n - g_m\|_1 \leq \|f_n - f_m\|_1$, 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 $L^1(\mathscr{G},\mu)$ and hence converges to some functions $g \in L^1(\mathscr{G},\mu)$. This g satisfies, for all $G \in \mathscr{G}^0$, 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 $g = \mathbb{E}[f|\mathscr{G}]$.

1.7. **Existence in general.** Let finally $f \in L^1_\sigma(\mathscr{F},\mu)$. Let $G_i \in \mathscr{G}^0$ be disjoint sets such that $\bigcup_{i=0}^\infty G_i = \Omega$, which can be chosen by the σ-finiteness of $(\Omega,\mathscr{G},\mu)$. Now $f_i := 1_{G_i} f \in L^1(\mathscr{F},\mu)$, so there exists $g_i := \mathbb{E}[f_i|\mathscr{G}] \in L^1(\mathscr{G},\mu)$.

Observe that then $g_i = 1_{G_i} g_i$. Indeed, for all $G \in \mathscr{G}^0$ there holds

$$\int_G 1_{G_i} g_i \,\mathrm{d}\mu = \int_{G\cap G_i} g_i \,\mathrm{d}\mu = \int_{G\cap G_i} f_i \,\mathrm{d}\mu = \int_G f_i \,\mathrm{d}\mu = \int_G g_i \,\mathrm{d}\mu,$$

where use was made of the definition of the conditional expectation $g_i = \mathbb{E}[f_i|\mathscr{G}]$ and the fact that $f_i = 1_{G_i} f_i$. Now the claim $g_i = 1_{G_i} g_i$ follows from Lemma 1.2.

Then one can set $g := \sum_{i=0}^\infty g_i$, which converges pointwise trivially, since the g_i are supported on the disjoint sets G_i. It is easy to check directly from the definition that $g = \mathbb{E}[f|\mathscr{G}]$.

Altogether, the following result has been established:

1.8. **Theorem.** *Let $(\Omega,\mathscr{F},\mu)$ and $(\Omega,\mathscr{G},\mu)$ be σ-finite measure spaces with $\mathscr{G} \subseteq \mathscr{F}$. Then for all $f \in L^1_\sigma(\mathscr{F},\mu)$ there exists an a.e. unique conditional expectation $\mathbb{E}[f|\mathscr{G}] \in L^1_\sigma(\mathscr{G},\mu)$.*

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

1.9. **Lemma.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function, 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 $x_0 \in \mathbb{R}$. Then the limit

$$a := \lim_{y \searrow x_0} \frac{\phi(y) - \phi(x_0)}{y - x_0}$$

exists. Indeed, from the definition of convexity it follows that the difference quotient inside the limit is an increasing function of y on the set $y \in \mathbb{R} \setminus \{x_0\}$. Choosing some $x_1 < x_0$, the quotient $(\phi(y) - \phi(x_0))/(y - x_0)$ is hence bounded from below by $(\phi(x_1) - \phi(x_0))/(x_1 - x_0)$, when $y > x_0$, and it decreases as $y \searrow x_0$. From this the existence of the limit follows by well-known properties of real numbers.

Let then $h_0(x) := \phi(x_0) + a(x - x_0)$. By what was said about the behaviour of the difference quotient defining a, the inequality $a(x-x_0) \leq \phi(x) - \phi(x_0)$ holds for all $x \in \mathbb{R}$, and hence $h_0 \in H_\phi$. On the other hand, clearly $h_0(x_0) = \phi(x_0)$, so $\sup_{h\in H_\phi} h(x_0) \geq \phi(x_0)$. Since $x_0 \in \mathbb{R}$ was arbitrary, the claim follows. ☐

1.10. **Theorem** (Jensen's inequality)**.** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex and $f, \phi(f) \in L^1_\sigma(\mathscr{F},\mu)$. Then*

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

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

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

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

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

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

Proof. Let $p < \infty$, the case $p = \infty$ being easier. Since $L^p(\mathscr{F},\mu) \subseteq L^1_\sigma(\mathscr{F},\mu)$, the conditional expectation $\mathbb{E}[f|\mathscr{G}]$ exists, and similarly $\mathbb{E}[|f|^p|\mathscr{G}]$. Since the function $t \mapsto |t|^p$ is convex, Jensen's inequality implies that

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

Hence for all $G \in \mathscr{G}^0$ there holds

$$\int_G |\mathbb{E}[f|\mathscr{G}]|^p \,\mathrm{d}\mu \leq \int_G \mathbb{E}[|f|^p|\mathscr{G}] \,\mathrm{d}\mu = \int_G |f|^p \,\mathrm{d}\mu.$$

By choosing disjoint $G_k \in \mathscr{G}^0$ from the definition of the σ-finiteness of $(\Omega, \mathscr{G}, \mu)$ and adding up the above estimates over all $G = G_k$, the claim follows. □

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

1.12. **Monotone convergence theorem.** Recall that the version of integration theory says that if a sequence of measurable functions satisfies $0 \leq f_n \nearrow f$ 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 $f_n \leq f_{n+1}$ for all n. The corresponding statement for the conditional expectation is the following:

$$0 \leq f_n \nearrow f \in L^1_\sigma(\mathscr{F}, \mu) \quad \Rightarrow \quad \mathbb{E}[f_n|\mathscr{G}] \nearrow \mathbb{E}[f|\mathscr{G}].$$

Proof. Since the conditional expectation respects pointwise inequalities (part 1.5), it follows that

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

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 $0 \leq g \leq \mathbb{E}[f|\mathscr{G}]$, so that $g \in L^1_\sigma(\mathscr{G}, \mu)$. It remains to prove that $g = \mathbb{E}[f|\mathscr{G}]$.

Since for all $G \in \mathscr{G}^0$ there holds

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

where the first and last steps were based on the usual monotone convergence theorem, it follows that $g = \mathbb{E}[f|\mathscr{G}]$, mikä which completes the proof. □

1.13. **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 one similarly proves

$$0 \leq f_n \in L^1_\sigma(\mathscr{F}, \mu),\ f := \liminf_{n\to\infty} f_n \in L^1_\sigma(\mathscr{F}, \mu) \quad \Rightarrow \quad \mathbb{E}[f|\mathscr{G}] \leq \liminf_{n\to\infty} \mathbb{E}[f_n|\mathscr{G}].$$

Proof. Write out the definition of the limes inferior:

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

Now $0 \leq h_n \nearrow f$, so one can use the monotone convergence theorem to the result that

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

The estimate $\leq$ above was based on the fact that $h_n \leq f_m$ for all $m \geq n$ and hence $\mathbb{E}[h_n|\mathscr{G}] \leq \mathbb{E}[f_m|\mathscr{G}]$, from which the mentioned step follows upon taking the infimum. The claim is proven. □

1.14. **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:

$$f_n \to f,\ |f_n| \leq g \in L^1_\sigma(\mathscr{F}, \mu) \quad \Rightarrow \quad \mathbb{E}\big[|f_n - f|\big|\mathscr{G}\big] \to 0 \quad \Rightarrow \quad \mathbb{E}[f_n|\mathscr{G}] \to \mathbb{E}[f|\mathscr{G}].$$

Proof. This is left as an exercise. □

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

1.15. **Theorem.** *Let $f \in L^1_\sigma(\mathscr{F}, \mu)$, and g be a $\mathscr{G}$-measurable function with $g \cdot f \in L^1_\sigma(\mathscr{F}, \mu)$. Then*

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

Proof. Let first g be 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 $G \in \mathscr{G}^0$ 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 $|g_n| \le |g|$ and $g_n \to g$. Hence also $|g_n \cdot f| \le |g \cdot f|$ 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. □

1.16. **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 1.14.)
3. Prove the *tower rule* of conditional expectations: If $\mathscr{H} \subseteq \mathscr{G}$ is yet another σ-algebra, then for all $f \in L^1_\sigma(\mathscr{F},\mu)$ there holds $\mathbb{E}\big(\mathbb{E}[f|\mathscr{G}]\big|\mathscr{H}\big) = \mathbb{E}[f|\mathscr{H}]$.
4. Prove the *conditional Hölder inequality*: If $f \in L^p_\sigma(\mathscr{F},\mu)$ and $g \in L^{p'}_\sigma(\mathscr{F},\mu)$, then
$$\mathbb{E}[f\cdot g|\mathscr{G}] \le \mathbb{E}[|f|^p|\mathscr{G}]^{1/p} \cdot \mathbb{E}[|f|^{p'}|\mathscr{G}]^{1/p'}.$$
(Hint: prove first that for all $a, b \ge 0$ there holds $ab \le a^p/p + b^{p'}/p'$.)

1.17. **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 [11]. It is well known “among specialists” that most of the results remain true in more general measure spaces, as it has been shown above, but it is difficult to find a systematic presentation in the literature.

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 is $\mathscr{F}_i$-measurable 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 $f_i \in L^1_\sigma(\mathscr{F}_i,\mu)$ for all $i \in I$ and $f_i \le \mathbb{E}[f_j|\mathscr{F}_i]$ whenever $i < j$.
- 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 $f \in L^1_\sigma(\mathscr{F},\mu)$ and $(\mathscr{F}_i)_{i\in I}$ is a filtration, with the related measure spaces σ-finite, 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.** 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.,

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

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.

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 $\mathscr{F} = \sigma\Big(\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i\Big)$. Then $\mathscr{F} = \tilde{\mathscr{F}}$.*

Proof. Clearly $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i \subseteq \tilde{\mathscr{F}} \subseteq \mathscr{F}$. ("$\subseteq$" 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 $\mathscr{F}$ was the smallest σ-algebra containing $\bigcup_{i\in\mathbb{Z}} \mathscr{F}_i$ – it follows that $\mathscr{F} \subseteq \tilde{\mathscr{F}}$, and this implies the assertion.

Trivially $\varnothing, \Omega \in \mathscr{F}$, and the implication $E \in \mathscr{F} \Rightarrow E^{\mathsf{C}} \in \mathscr{F}$ 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 $F_k \in \mathscr{F}^0_{i(k)} \subseteq \mathscr{F}^0_j$, 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. **Corollary.** *Under the assumption of Theorem 2.5, for all $f \in L^p(\mathscr{F}, \mu)$ there is convergence*

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

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 $g \in L^p(\mathscr{F}_j, \mu)$, such that $\|f - g\|_p < \varepsilon$. Now

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

and the middle term vanishes for $i \geq j$, since then $\mathbb{E}[g|\mathscr{F}_i] = g$. Moreover, $\|\mathbb{E}[f - g|\mathscr{F}_i]\|_p \leq \|f - g\|_p$, and hence

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

for $i \geq j$. □

2.7. **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:

2.8. **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 $f_i=\mathbb{E}[f|\mathscr{F}_i]$.

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

2.10. **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 $\{n,n+1,n+2,\ldots\}$ 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 $\{0,1,\ldots,n\}$. 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 $k = 0, 1, \dots, n$.

2.11. **Burkholder's proof for Doob's inequality.** Let us write $f_k^* := \max_{0\leq j\leq k} f_j$ and $d_k := f_k - f_{k-1}$. The submartingale assumption tells that

$$\mathbb{E}[d_k|\mathscr{F}_{k-1}] = \mathbb{E}[f_k|\mathscr{F}_{k-1}] - f_{k-1} \geq 0,$$

also recall that $f_k \geq 0$. In this new notation, the claim to be proven reads as

$$\int_\Omega v(f_n, f_n^*)\,\mathrm{d}\mu \leq 0, \qquad v(x,y) := y^p - (p'x)^p, \quad 0 \leq x \leq y.$$

Burkholder's idea is to replace v by a new function u which has "better" properties. Suppose that we had a measurable function u of two variables, defined in the domain where $0 \leq x \leq y$, such that

$$\begin{aligned}&v(x,y) \leq u(x,y), \qquad |u(x,y)| \leq Cy^p, \qquad u(x,x) \leq 0,\\ &u(x+h, \max\{x+h,y\}) \leq u(x,y) - w(x,y)h, \quad 0 \leq x \leq y, \quad 0 \leq x+h,\end{aligned}$$

where w is a function defined on the same domain which satisfies $0 \leq w(x,y) \leq Cy^{p-1}$. The first two properties express the fact that, on the one hand, the original function v is dominated by u, but also u is not too big, so that $u(x,y) \in L^1(\mathscr{F},\mu)$ when one substitues $L^p(\mathscr{F},\mu)$ functions for the variables x and y.

The somewhat mysterious-looking final requirement on u above is precisely what is needed to estimate the expression $u(f_n, f_n^*)$. Taking $x := f_{n-1}$, $y := f_{n-1}^*$ and $h := d_n$, it follows that $f_n = x + h$ and $f_n^* = \max\{f_n, f_{n-1}^*\} = \max\{x+h, y\}$; siis

$$u(f_n, f_n^*) \leq u(f_{n-1}, f_{n-1}^*) - w(f_{n-1}, f_{n-1}^*)d_n.$$

The integral of the second term on the right is

$$\begin{aligned}\int_\Omega w(f_{n-1}, f_{n-1}^*)d_n\,\mathrm{d}\mu &= \int_\Omega \mathbb{E}[w(f_{n-1}, f_{n-1}^*)d_n|\mathscr{F}_{n-1}]\,\mathrm{d}\mu\\ &= \int_\Omega w(f_{n-1}, f_{n-1}^*)\mathbb{E}[d_n|\mathscr{F}_{n-1}]\,\mathrm{d}\mu \geq 0,\end{aligned}$$

since $w \geq 0$ and $\mathbb{E}[d_n|\mathscr{F}_{n-1}] \geq 0$. Hence

$$\int_\Omega v(f_n, f_n^*) \leq \int_\Omega u(f_n, f_n^*)\,\mathrm{d}\mu \leq \int_\Omega u(f_{n-1}, f_{n-1}^*)\,\mathrm{d}\mu.$$

Iterating the last inequality, after n steps it follows that

$$\int_\Omega u(f_n, f_n^*)\,\mathrm{d}\mu = \int_\Omega u(f_0, f_0^*)\,\mathrm{d}\mu = \int_\Omega u(f_0, f_0)\,\mathrm{d}\mu \leq 0,$$

where we observed that $f_0^* = \max_{0\leq j\leq 0} f_j = f_0$, and used the property that $u(x,x) \leq 0$.

The proof is complete except for showing that functions u and w with the desired properties actually exist.

2.12. **Search for the auxiliary functions I: the form of the functions.** How to find appropriate functions u and w? Let us first exploit the homogeneity of the problem to see that if u and w are suitable functions, then for all $\lambda > 0$ also $u_\lambda(x,y) := \lambda^{-p}u(\lambda x, \lambda y)$ and $w_\lambda(x,y) := \lambda^{-p+1}w(\lambda x, \lambda y)$ are such functions. Here it is decisive that the original function v verifies $v(x,y) = \lambda^{-p}v(\lambda x, \lambda y)$. Furthermore, observe that even the functions $\tilde{u} := \inf_{\lambda>0} u_\lambda$ and $\tilde{w} := \inf_{\lambda>0} w_\lambda$ will do. Here one should note that infima are still real-valued (and never $-\infty$), since there holds $u_\lambda \geq v$ and $w_\lambda \geq 0$ for all λ. The functions $\tilde{u}$ and $\tilde{w}$ satisfy the additional homogeneity properties $\tilde{u}(\lambda x, \lambda y) = \lambda^p \tilde{u}(x,y)$ and $\tilde{w}(x,y) = \lambda^{p-1}\tilde{w}(x,y)$, so it suffices to look for such auxiliary functions which verify these additional restriction. In the sequel, thse will be denoted simply by u and w without the tildes.

Consider the required inequality $u(x+h, \max\{x+h,y\})$ for $x, x+h \in [0,y]$. We want that

$$u(x+h, y) - u(x,y) \leq -w(x,y)h.$$

Let us rewrite this condition with x replaced by $x+h$ and h by $-h$; this gives

$$u(x,y) - u(x+h,y) \leq w(x+h,y)h.$$

For $h>0$, these imply that

$$-w(x+h,y) \leq \frac{u(x+h,y)-u(x,y)}{h} \leq -w(x,y),$$

so $-w$ is a decreasing as a function of x, and hence w is increasing.

Suppose that u is differentiable and w is continuous with respect to x at the point (x,y). The limit $h \searrow 0$ of the previous inequality gives

$$-w(x,y) \leq u_x(x,y) \leq -w(x,y),$$

where $u_x := \partial u/\partial x$, and hence in fact $w=-u_x$. By the earlier observations, u_x is decreasing with respect to x. Conversely, this condition gives

$$u(x+h,y)-u(x,y) = \int_0^1 u_x(x+th,y)h\,\mathrm{d}t \leq u_x(x,y)h = -w(x,y)h,$$

so the original requirement is verified.

The homogeneity condition implies that

$$u(x,y) = u(y\cdot x/y, y\cdot 1) = y^p u(x/y,1) =: y^p\phi(x/y),$$

where ϕ is defined on $[0,1]$. Furthermore,

$$w(x,y) = u_x(x,y) = y^{p-1}\phi'(x/y).$$

so ϕ' should be increasing.

Consider then the inequality between u and w for $x+h>y$. In terms of the new function ϕ, the condition says that

$$(x+h)^p\phi(1) - y^p\phi(x/y) \leq y^{p-1}\phi'(x/y)h,$$

or equivalently, with the new variables $t:=x/y$ and $s:=h/y$, which satisfy $t+s>1$,

$$(t+s)^p\phi(1)-\phi(t) \leq \phi'(t)s.$$

After moving terms, this gets the form

$$\phi(1)[(t+s)^p-1] \leq \phi(t)-\phi(1)+\phi'(t)s. \tag{$*$}$$

When $s\searrow 1-t$, the left side tends to zero, and hence $\phi(t)-\phi(1)+\phi'(t)(1-t)\geq 0$. On the other hand, since the derivative ϕ' is increasing, the mean value theorem shows that, for some $\xi\in(t,1)$,

$$\phi(t)-\phi(1) = (t-1)\phi'(\xi) \geq (t-1)\phi'(t),$$

so in fact the differential equation $(1-t)\phi'(t)+\phi(t)-\phi(1)=0$ has to be satisfied.

Multiplying both sides by the factor $(1-t)^{-2} = [(1-t)^{-1}]'$, it follows that

$$\frac{d}{dt}\Big(\frac{\phi(t)-\phi(1)}{1-t}\Big) = 0,$$

and integration gives

$$\phi(t) = \phi(0)(1-t)+\phi(1)t =: c_0 - c_1 t.$$

Substituting back, the auxiliary functions are seen to have the form

$$u(x,y) = y^p\phi(x/y) = c_0y^p - c_1y^{p-1}x, \qquad w(x,y) = c_1y^{p-1}.$$

2.13. **Search for the auxiliary functions II: determining the constants.** Let us start from the easy observations that the requirements $y^p = v(0,y) \le u(0,y) = c_0 y^p$ and $(c_0 - c_1)y^p = u(y,y) \le 0$ are equivalent to the condition $c_1 \ge c_0 \ge 1$. Hence also $w(x,y) \ge 0$ and the upper bounds required for the absolute values of the functions are clearly satisfied, but still one has to find c_0 and c_1 in such a way that also the other conditions hold.

Let us return to the inequality between u and w for $x + h > y$. This was already seen to be equivalent to $(*)$ in Section 2.12, i.e., to

$$(c_0 - c_1)[(t+s)^p - 1] \le (c_0 - c_1 t) - (c_0 - c_1) - c_1 s = -c_1[(t+s) - 1].$$

Dividing this by $(t+s-1)$ and taking the limit $s \searrow 1-t$ leads to $(c_0 - c_1)p \le -c_1$, i.e., $c_0 p' \le c_1$. Conversely, this condition implies the original inequality by the mean value theorem.

There remains the requirement $u \ge v$ which, after moving terms, attains the form

$$(c_0 - 1)y^p - c_1 y^{p-1} x + (p'x)^p \ge 0, \qquad y \ge x \ge 0.$$

This inequality holds in the limit $y \to \infty$ if and only if $c_0 > 1$. For $y = 0$ it holds automatically (although this is actually not needed except when $x = 0$). The function above attains its smallest value in the interval $y \in [0,\infty)$ either at the left endpoint or where the derivative vanishes. The last condition happens at the unique point where $(c_0 - 1)py^{p-1} - c_1(p-1)y^{p-2}x = 0$, i.e.,

$$y = \frac{c_1}{c_0 - 1}\frac{x}{p'} \ge \frac{c_0 p'}{c_0 - 1}\frac{x}{p'} > x,$$

so this is contained in the critical interval $[x, \infty)$. At this point the function attains the value

$$(c_0 - 1)^{1-p} c_1^p (x/p')^p - c_1^p (c_0 - 1)^{1-p} (x/p')^{p-1} x + (p'x)^p,$$

and the requirement of its nonnegativity reads (observe that $(p'-1)^{-1} = (p-1)$)

$$(c_0 - 1)^{1-p} c_1^p (p' - 1) \le (p')^{2p}, \quad \text{eli} \quad c_1 \le (p')^2 (p-1)^{1/p} (c_0 - 1)^{1/p'}.$$

All in all, it has been shown that the functions $u(x,y) = c_0 y^p - c_1 y^{p-1} x$ and $w(x,y) = c_1 y^{p-1}$ satisfy the required conditions if and only if

$$p' c_0 \le c_1 \le (p')^2 (p-1)^{1/p} (c_0 - 1)^{1/p'}, \qquad c_0 > 1.$$

Let us denote $t := c_0 - 1$ and look for a solution for the inequality

$$F(t) := p'(p-1)^{1/p} t^{1/p'} - t \ge 1, \qquad t \in [0, \infty).$$

The maximal value of the function is reached at the zero of the derivative, i.e., when $(p-1)^{1/p} t^{1/p'-1} - 1 = 0$ which is solved for $t = p - 1$. At this point, $F(p-1) = p'(p-1) - (p-1) = p - (p-1) = 1$. Hence, unique values $c_0 = t + 1 = p$ and $c_1 = p'c_0 = p'p$ have been found, and these give

$$u(x,y) = py^{p-1}(y - p'x), \qquad w(x,y) = pp'y^{p-1}.$$

2.14. **Convergence of martingales to the reverse direction.** We have seen that if $(\mathscr{F}_j)_{j\in\mathbb{Z}}$ is a filtration with $\sigma(\bigcup_{j\in\mathbb{Z}} \mathscr{F}_j) = \mathscr{F}$, 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:

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

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

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

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:

2.15. **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. □

2.16. **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 $|I| \le C|J|$. (Hint: it might help to sketch a picture.)
6. For $f \in L^1_{\text{loc}}(\mathbb{R})$, 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
$$\|M_{HL}f\|_p \le C_p \|f\|_p, \qquad p \in (1,\infty].$$
(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.)

2.17. **References.** The core of this chapter, which is Burkholder's proof for Doob's inequality, is taken from Burkholder's summer school lectures [4]. The idea for Exercises 2 and 3 is also from there. The results of Exercises 5 and 6 are from Mei [7].

Burkholder developed his technique of proving various inequalities of analysis by ingenious use of convex functions during the 1980's. At that time, however, the popularity of this method was left relatively restricted. Since then, Burkholder's idea has been rediscovered and systematically developed in the 2000's especially by Nazarov, Treil and Volberg. They refer to this proof technique as the *Bellman function* method, since the auxiliary functions ("Bellman functions") appearing in the arguments have a certain connection to Bellman differential equations from Stochastic Control; this aspect, however, will not be considered in these lectures.

Doob's inequality itself appeared for the first time in his classic book [5]. In his original approach, the L^p inequality is derived from a *weak-type* estimate in L^1, which in turn is proved by a *stopping time* technique. This nice argument is commonly found in more recent books as well (e.g. [11]), and it is also the way that the Hardy–Littlewood inequality is usually proved in modern textbooks, although not the proof originally given by Hardy and Littlewood [6].

3. Burkholder's inequality

3.1. **The sign transform of a martingale.** On a measure space $(\Omega, \mathscr{F}, \mu)$ with a filtration $(\mathscr{F}_k)_{k=0}^n$, consider a martingale $f = (f_k)_{k=0}^n$ and denote the corresponding *martingale difference sequence* by
$$d_0 := f_0, \qquad d_k := f_k - f_{k-1}, \quad k = 1, \ldots, n.$$
So here $\mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$ for all $k = 1, \ldots, n$, and $f_k = \sum_{j=0}^k d_j$.

Let $(\varepsilon_k)_{k=0}^n$ be a sequence of signs, $\varepsilon_k \in \{-1, +1\}$. The sign transform of the martingale f is the new martingale $g = (g_k)_{k=0}^n$, where $g_k := \sum_{j=0}^k \varepsilon_j d_j$.

In this chapter, the following fundamental result is proven. It has important consequences, some of which will be considered later on.

3.2. **Theorem** (Burkholder's inequality). *Let $p \in (1,\infty)$. Then there exists a constant β, such that if $f_n \in L^p(\mathscr{F},\mu)$ and g_n is its sign transform, then*

$$\|g_n\|_p \leq \beta \|f_n\|_p.$$

One can take

$$\beta = p^* - 1 := \max\{p, p'\} - 1 = \max\Big\{p-1, \frac{1}{p-1}\Big\},$$

and this constant is optimal.

The core idea of the proof is similar to Burkholder's proof for Doob's inequality: to relate the estimate concerning martingales to the properties of appropriate convex functions. Again it is useful to start with some simplifying preparatory considerations.

3.3. **Lemma** (Reduction to finite measure spaces). *It suffices to prove Burkholder's inequality in the case $\mu(\Omega) < \infty$, i.e., this case already implies the general case.*

Proof. Let $(d_k)_{k=0}^n$ be a martingale difference sequence on a σ-finite space $(\Omega,\mathscr{F},\mu)$. By the standing assumptions, $(\Omega,\mathscr{F}_0,\mu)$ is also σ-finite, and hence there are $E_j \in \mathscr{F}_0$ of finite measure with $E_j \nearrow \Omega$. Then $1_{E_j} d_k \to d_k$ as $j \to \infty$ pointwise, and then in L^p by dominated convergence. Hence we can fix some $E = E_j$ such that $\|1_E d_k - d_k\|_p < \varepsilon$ for all $k = 0,\ldots,n$. Then $(1_E d_k)_{k=0}^n$ is another martingale difference sequence, indeed $\mathbb{E}[1_E d_k|\mathscr{F}_{k-1}] = 1_E \mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$ for all $k \geq 1$ since E is measurable with respect to $\mathscr{F}_0 \subseteq \mathscr{F}_{k-1}$.

Assuming Burkholder's inequality for the finite measure space $(E,\mathscr{F}\cap E,\mu)$, it follows that

$$\begin{aligned}\Big\|\sum_{k=0}^n \varepsilon_k d_k\Big\|_p \leq \Big\|\sum_{k=0}^n \varepsilon_k 1_E d_k\Big\|_p + \sum_{k=0}^n \|d_k - 1_E d_k\|_p &\leq \beta\Big\|\sum_{k=0}^n 1_E d_k\Big\|_p + \sum_{k=0}^n \varepsilon \\ &\leq \beta\Big\|\sum_{k=0}^n d_k\Big\|_p + (n+1)\varepsilon.\end{aligned}$$

Letting $\varepsilon \searrow 0$, the assertion follows. □

3.4. **Lemma** (Reduction to simple martingales). *It suffices to prove Burkholder's inequality in the case when each $f_k \in L^p(\mathscr{F}_k,\mu)$ is a simple function.*

Proof. Let $(\Omega,\mathscr{F},\mu)$ be a finite measure space, and $(d_k)_{k=0}^n$ be a martingale difference sequence. For $\varepsilon > 0$ and each $k = 0,1,\ldots,n$, there are $\mathscr{F}_k$-simple functions s_k, such that $\|d_k - s_k\|_p < \varepsilon$.

Let $\mathscr{G}_k := \sigma(s_0,\ldots,s_k) \subseteq \mathscr{F}_k$ be the smallest σ-algebra for which $s_0,\ldots,s_k$ are measurable. Since each s_j is simple, $\mathscr{G}_k$ contains only finitely many sets. Clearly $(\mathscr{G}_k)_{k=0}^\infty$ is a filtration. The functions s_k may fail to be martingale differences, but the new functions

$$u_k := s_k - \mathbb{E}[s_k|\mathscr{G}_{k-1}]$$

clearly are. Since $\mathscr{G}_{k-1}$ is finite, the conditional expectation, and then u_k, is again a simple function. (Note that, in general, the conditional expectation of a simple function need not be simple; see exercises.)

There holds

$$\mathbb{E}[d_k|\mathscr{G}_{k-1}] = \mathbb{E}[\mathbb{E}[d_k|\mathscr{F}_{k-1}]|\mathscr{G}_{k-1}] = \mathbb{E}[0|\mathscr{G}_{k-1}] = 0,$$

and hence

$$\|\mathbb{E}[s_k|\mathscr{G}^0_{k-1}]\|_p = \|\mathbb{E}[s_k - d_k|\mathscr{G}^0_{k-1}]\|_p \leq \|s_k - d_k\|_p < \varepsilon;$$

thus $\|u_k - d_k\|_p < 2\varepsilon$. So u_k is a simple martingale difference sequence which approximates the original one. Then

$$\Big|\Big\|\sum_{k=0}^n \varepsilon_k d_k\Big\|_p - \Big\|\sum_{k=0}^n \varepsilon_k u_k\Big\|_p\Big| \leq \Big\|\sum_{k=0}^n \varepsilon_k [d_k - u_k]\Big\|_p \leq \sum_{k=0}^n \|d_k - u_k\|_p \leq 2\varepsilon(n+1)$$

which holds in particular in the case $\varepsilon_k \equiv 1$.

If Burkholder's inequality holds in the simple martingale case

$$\Big\| \sum_{k=0}^{n} \varepsilon_k u_k \Big\|_p \le \beta \Big\| \sum_{k=0}^{n} u_k \Big\|_p,$$

then

$$\begin{aligned}\Big\| \sum_{k=0}^{n} \varepsilon_k d_k \Big\|_p &\le \Big\| \sum_{k=0}^{n} \varepsilon_k u_k \Big\|_p + 2(n+1)\varepsilon \le \beta \Big\| \sum_{k=0}^{n} u_k \Big\|_p + 2(n+1)\varepsilon \\ &\le \beta\Big(\Big\| \sum_{k=0}^{n} d_k \Big\|_p + 2(n+1)\varepsilon \Big) + 2(n+1)\varepsilon = \beta \Big\| \sum_{k=0}^{n} d_k \Big\|_p + 2(\beta+1)(n+1)\varepsilon.\end{aligned}$$

The claim follows with $\varepsilon \searrow 0$. □

3.5. **Zigzag martingales.** Let $Z = (Z_k)_{k=0}^n = (X_k, Y_k)_{k=0}^n$ be a sequence of pairs of functions (X_k, Y_k). Then Z is called a two-dimensional martingale if both X_k ja Y_k are martingales adapted to the same filtration. Z is called a *zigzag martingale* if it has the following additional property: for all $k \ge 1$ either $X_k - X_{k-1} = 0$ or $Y_k - Y_{k-1} = 0$.

Starting from a martingale and its sign transform

$$f_k = \sum_{j=0}^{k} d_j, \qquad g_k = \sum_{j=0}^{k} \varepsilon_j d_j,$$

one defines new martingales

$$X_k := g_k + f_k = \sum_{j=0}^{k} (\varepsilon_j + 1) d_j, \qquad Y_k := g_k - f_k = \sum_{j=0}^{k} (\varepsilon_j - 1) d_j,$$

and the two-dimensional martingale $Z_k := (X_k, Y_k)$. This clearly has the zigzag property, since for all k either $\varepsilon_k = 1$, and then $Y_k - Y_{k-1} = 0$, or $\varepsilon_k = -1$, and then $X_k - X_{k-1} = 0$.

3.6. **Lemma.** *Let $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a bi-concave function, i.e., both $x \mapsto u(x, y)$ and $y \mapsto u(x, y)$ are concave functions $\mathbb{R} \to \mathbb{R}$. If Z is a simple zigzag martingale, there holds*

$$\int u(Z_k)\, \mathrm{d}\mu \le \int u(Z_{k-1})\, \mathrm{d}\mu.$$

Proof. By symmetry one may suppose that

$$X_k = X_{k-1} = \sum_{r=1}^{s} a_r 1_{A_r}, \qquad A_r \in \mathscr{F}_{k-1}^0.$$

Hence

$$\begin{aligned}\int u(Z_k)\, \mathrm{d}\mu &= \int u(X_{k-1}, Y_k)\, \mathrm{d}\mu = \sum_{r=1}^{s} \int_{A_r} u(a_r, Y_k)\, \mathrm{d}\mu = \sum_{r=1}^{s} \int_{A_r} \mathbb{E}[u(a_r, Y_k) | \mathscr{F}_{k-1}]\, \mathrm{d}\mu \\ &\le \sum_{r=1}^{s} \int_{A_r} u(a_r, \mathbb{E}[Y_k | \mathscr{F}_{k-1}]\, \mathrm{d}\mu = \sum_{r=1}^{s} \int_{A_r} u(a_r, Y_{k-1})\, \mathrm{d}\mu = \int u(X_{k-1}, Y_{k-1})\, \mathrm{d}\mu = \int u(Z_{k-1})\, \mathrm{d}\mu,\end{aligned}$$

where Jensen's inequality for conditional expectations was used. □

3.7. **Theorem.** *Burkholder's inequality $\|g_n\|_p \le \beta \|f_n\|_p$ holds for a constant β, if and only if there exists a bi-concave function $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ with*

$$u(x, y) \ge F(x, y) := \Big| \frac{x+y}{2} \Big|^p - \beta^p \Big| \frac{x-y}{2} \Big|^p.$$

Hence finding a bi-concave function u as above is not just one possible way of proving Burkholder's inequality, but this task is equivalent to the original problem! Also notice that this theorem by itself does not yet tell whether Burkholder's inequality is true or not, but only transforms it to a new question. There still remains the problem of finding the auxiliary function u, and this will be dealt with after the proof of the theorem.

3.8. **Additional symmetries for the function u.** Observe first that if there is some function u fulfilling the above conditions, there is even such a u which additionally satisfies

$$u(\alpha x, \alpha y) = |\alpha|^p u(x, y)$$

for all $\alpha \in \mathbb{R}$. Namely, define $u_\alpha(x, y) := |\alpha|^{-p} u(\alpha x, \alpha y)$ for $\alpha \neq 0$. Also this function is bi-concave and satisfies $u_\alpha \geq F$. Hence even $\tilde{u} := \inf_{\alpha \neq 0} u_\alpha$ verifies the same conditions; here one appeals to the fact that the infimum of concave functions is again concave. Now $\tilde{u}$ satisfies the above homogeneity property for $\alpha \neq 0$ by a simple change of variable. But then also

$$\tilde{u}(0 \cdot x, 0 \cdot y) = \tilde{u}(0, 0) = \tilde{u}(\alpha \cdot 0, \alpha \cdot 0) = |\alpha|^p \tilde{u}(0, 0) \to 0 = |0|^p \tilde{u}(x, y)$$

as $\alpha \to 0$, and the claim follows for all $\alpha \in \mathbb{R}$.

Furthermore one can require that $u(x, y) = u(y, x)$. If this condition is not already satisfied, then the new function $\frac{1}{2}[u(x, y) + u(y, x)]$ will do; it is still bi-concave and at least as large as $F(x, y) = F(y, x)$.

3.9. **Burkholder's inequality with the help of u.** Suppose that a function u exists. Let $Z = (X, Y)$ be the zigzag martingale related to a simple martingale f and its sign transform. Another way of writing the inequality to be proven is as follows:

$$\int \Big(\Big| \frac{X_n + Y_n}{2} \Big|^p - \beta^p \Big| \frac{X_n - Y_n}{2} \Big|^p \Big) \, \mathrm{d}\mu = \int F(X_n, Y_n) \, \mathrm{d}\mu \leq 0.$$

By assumption and Lemma 3.6, there holds

$$\int F(Z_n) \, \mathrm{d}\mu \leq \int u(Z_n) \, \mathrm{d}\mu \leq \int u(Z_{n-1}) \, \mathrm{d}\mu \leq \ldots \leq \int u(Z_0) \, \mathrm{d}\mu.$$

Here either $Z_0 = (X_0, 0)$ or $Z_0 = (0, Y_0)$, assume for example the first case. Thanks to the additional properties of u,

$$u(x, 0) = u(-x, 0) = \frac{u(x, 0) + u(-x, 0)}{2} \leq u\Big(\frac{x - x}{2}, 0\Big) = u(0, 0) = 0;$$

and the assertion is proven.

3.10. **Constructing u from the martingales.** For each pair of points $(x, y) \in \mathbb{R} \times \mathbb{R}$, consider the set

$$\mathbb{S}(x, y) := \{Z = (Z_k)_{k=0}^n \text{ simple zigzag martingale, } Z_k : [0, 1) \to \mathbb{R} \times \mathbb{R}, \ Z_0 = (x, y)\}.$$

The filtration has not been fixed, i.e., two zigzag martingales $Z, Z' \in \mathbb{S}(x, y)$ may be adapted to different filtrations of $\big([0, 1), \mathscr{B}([0, 1)), \mathrm{d}t\big)$. Also the parameter $n \in \mathbb{N}$ may vary and attain arbitrarily large values, but it is finite for each fixed Z. Nevertheless, the notation $Z_\infty := Z_n$ will be used for simplicity when $Z = (Z_k)_{k=0}^n$.

Now define the function

$$U(x, y) := \sup \Big\{ \int_0^1 F(Z_\infty) \, \mathrm{d}t : Z \in \mathbb{S}(x, y) \Big\} \in (-\infty, \infty].$$

The following properties should be checked:

- $U(x, y) \geq F(x, y)$,
- U is bi-concave, and
- U is real-valued, i.e., $U(x, y) < \infty$ everywhere.

The first property is clear, since the collection $\mathbb{S}(x, y)$ contains the trivial zigzag martingale $Z = (Z_0) = (Z_k)_{k=0}^0$, for which $Z_\infty = (x, y)$, and hence $U(x, y) \geq \int_0^1 F(x, y) \, \mathrm{d}t = F(x, y)$. The last point, which one could easily forget to think about, is in fact quite essential. Indeed, it would be easy to fulfill the other two properties with the trivail choice $U \equiv \infty$ identically, but this function would not be of any use.

3.11. **Bi-concavity.** Let $y, x_1, x_2 \in \mathbb{R}$, $\alpha \in (0,1)$ be arbitrary, $x = \alpha x_1 + (1-\alpha)x_2$, and also $m_1, m_2 \in \mathbb{R}$, such that $U(x_i, y) > m_i$. The claim is that $U(x,y) > \alpha m_1 + (1-\alpha)m_2$, which would imply $U(x,y) \geq \alpha U(x_1,y) + (1-\alpha)U(x_2,y)$ and similarly in the y-coordinate by completely symmetric considerations.

By the choice of the m_i, there are $Z^i \in \mathbb{S}(x_i, y)$, such that $\int_0^1 F(Z^i_\infty)\,\mathrm{d}t > m_i$. These zigzag martingales may even be chosen in such a way that

$$(*) \qquad Y^i_{2k+1} - Y^i_{2k} = 0 = X^i_{2k+2} - X^i_{2k+1}$$

and they have the same length, i.e., $Z^i = (Z^i_k)_{k=0}^n$ for the same $n \in \mathbb{N}$.

Proof. In case $Y^i_{2k+1} - Y^i_{2k} \neq 0$ for some k, one can "augment a zero step", i.e., to define a new martingale $\tilde{Z}^i_j := Z^i_j$ when $0 \leq j \leq 2k$ and $\tilde{Z}^i_j := Z^i_{j-1}$ when $j > 2k$. Now $\tilde{Z}^i \in \mathbb{S}(x,y)$ is another zigzag martingale with $\tilde{Z}^i_\infty = Z^i_\infty$, but in addition $\tilde{Z}^i_{2k+1} = \tilde{Z}^i_{2k}$, so this holds in particular for the y-component. By repeating this operation finitely many times, starting from the smallest k and taking into account the similar situations in the x-component, one gets a new $\tilde{Z}^i \in \mathbb{S}(x,y)$ which fulfills the requirement $(*)$ and also $\tilde{Z}^i_\infty = Z^i_\infty$, so in particular $\int_0^1 F(\tilde{Z}^i_\infty)\,\mathrm{d}t > m_i$. Finally, if the martingales $\tilde{Z}^1$ ja $\tilde{Z}^2$ have different lenghts, zero steps can be similarly augmented at the end of the shorter one. □

Then one defines

$$Z_0(t) := (x,y), \qquad Z_{k+1}(t) := \begin{cases} Z^1_k(t/\alpha), & t \in [0,\alpha), \\ Z^2_k\big((t-\alpha)/(1-\alpha)\big), & t \in [\alpha,1), \end{cases} \quad k = 1, \ldots, n.$$

Clearly

$$\begin{aligned} \int_0^1 F(Z_\infty)\,\mathrm{d}t &= \int_0^\alpha F \circ Z^1_\infty(t/\alpha)\,\mathrm{d}t + \int_\alpha^1 F \circ Z^2_\infty\big((t-\alpha)/(1-\alpha)\big)\,\mathrm{d}t \\ &= \alpha \int_0^1 F(Z^1_\infty)\,\mathrm{d}t + (1-\alpha)\int_0^1 F(Z^2_\infty)\,\mathrm{d}t > \alpha m_1 + (1-\alpha)m_2, \end{aligned}$$

so it remains to prove that $Z = (Z_k)_{k=0}^{n+1} \in \mathbb{S}(x,y)$.

Proof. Let $(\mathscr{F}^i_k)_{k=0}^n$ be the filtration related to the martingale $Z^i = (Z^i_k)_{k=0}^n$. Set

$$\mathscr{F}_0 := \{\varnothing, [0,1)\}, \qquad \mathscr{F}_{k+1} := \sigma\big(\alpha \cdot \mathscr{F}^1_k, \alpha + (1-\alpha)\cdot \mathscr{F}^2_k\big).$$

Then clearly the Z defined above is adapted to $(\mathscr{F}_k)_{k=0}^n$, it remains to prove that it is a zigzag martingale. When $k \geq 1$, there holds

$$\begin{aligned} \mathbb{E}[Z_{k+1}|\mathscr{F}_k] &= \mathbb{E}[1_{[0,\alpha)} Z^1_k(\cdot/\alpha)\big|\alpha \cdot \mathscr{F}^1_{k-1}] + \mathbb{E}[1_{[\alpha,1)} Z^2_k\big((\cdot-\alpha)/(1-\alpha)\big)\big|\alpha + (1-\alpha)\cdot\mathscr{F}^2_{k-1}] \\ &= 1_{[0,\alpha)}\mathbb{E}[Z^1_k|\mathscr{F}^1_{k-1}](\cdot/\alpha) + 1_{[\alpha,1)}\mathbb{E}[Z^2_k|\mathscr{F}^2_{k-1}]\big((\cdot-\alpha)/(1-\alpha)\big) \\ &= 1_{[0,\alpha)} Z^1_{k-1}(\cdot/\alpha) + 1_{[\alpha,1)} Z^2_{k-1}\big((\cdot-\alpha)/(1-\alpha)\big) = Z_k, \end{aligned}$$

where most of the steps where essentially changes of variables, which are easy to verify directly from the definition of the conditional expectation, especially since the functions and the σ-algebras above are simple.

When $k = 0$, there holds

$$\begin{aligned} \mathbb{E}[Z_1|\mathscr{F}_0] &= \mathbb{E}[1_{[0,\alpha)} Z^1_0|\mathscr{F}_0] + \mathbb{E}[1_{[\alpha,1)} Z^2_0|\mathscr{F}_0] = \mathbb{E}[1_{[0,\alpha)}|\mathscr{F}_0](x_1,y) + \mathbb{E}[1_{[\alpha,1)}|\mathscr{F}_0](x_2,y) \\ &= \alpha(x_1,y) + (1-\alpha)(x_2,y) = (x,y), \end{aligned}$$

since $Z^i_0 \equiv (x_i, y)$, and the conditional expectation of a function with respect to the trivial σ-algebra is its average on the whole space.

The general zigzag property would not necessarily be inherited by Z from the Z^i, but now that there holds the stronger property $(*)$, it is easy to see that a similar condition (with shifted indices) is also true for Z. □

3.12. **Finiteness of values.** So far the assumed inequality of Burkholder has not been used in the consideration of the properties of U, so it should happen in this last phase. A central step in proving the finiteness is to show that

$$U(0,0) \leq 0.$$

Proof. Let $Z = (Z_k)_{k=0}^n \in \mathbb{S}(0,0)$, in particular $Z_0 = (X_0, Y_0) = (0,0)$. Define the signs ε_k and the martingale differences d_k by:

$$(\varepsilon_k, d_k) := \begin{cases} (+1, +\frac{1}{2}[X_k - X_{k-1}]) & \text{if } X_k - X_{k-1} \neq 0, \\ (-1, -\frac{1}{2}[Y_k - Y_{k-1}]) & \text{otherwise,} \end{cases}$$

and $f_k := \sum_{j=1}^k d_j$, $g_k := \sum_{j=1}^k \varepsilon_j d_j$. By the zigzag property, for all $k = 1, \ldots, n$ there holds

$$X_k - X_{k-1} = (\varepsilon_k + 1) d_k, \qquad Y_k - Y_{k-1} = (\varepsilon_k - 1) d_k,$$

and summing up,

$$X_k = g_k + f_k, \qquad Y_k = g_k - f_k.$$

Thus in particular

$$\int_0^1 F(Z_n)\,\mathrm{d}t = \int_0^1 \Big|\frac{X_n + Y_n}{2}\Big|^p - \beta^p \Big|\frac{X_n - Y_n}{2}\Big|^p \,\mathrm{d}t = \|g_n\|_p^p - \beta^p \|f_n\|_p^p \leq 0.$$

Upon taking the supremum over all $Z \in \mathbb{S}(0,0)$, the assertion follows. □

The proof of everywhere finiteness is finished by simple concavity considerations: for all $x, y \in \mathbb{R}$ there holds

$$[U(x,y) + U(-x,y)] + [U(x,-y) + U(-x,-y)] \leq 2U(0,y) + 2U(0,-y) \leq 4U(0,0) \leq 0 < \infty,$$

so all the terms on the left need to be finite.

3.13. **Search for the auxiliary function: reduction to a one-variable problem.** By Theorem 3.7, one should find a bi-concave function $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, such that $u \geq F$. By section 3.8, one can additionally demand that $u(x,y) = u(y,x)$ and $u(\alpha x, \alpha y) = |\alpha|^p u(x,y)$ for all $\alpha, x, y \in \mathbb{R}$. But then, if $y \neq 0$,

$$u(x,y) = u(y \cdot x/y, y \cdot 1) = |y|^p u(x/y, 1) =: |y|^p w(x/y),$$

where

$$(*) \qquad \begin{aligned} &w \text{ is concave}, \qquad w(x) = u(x,1) \geq F(x,1) =: f(x), \\ &w(1/x) = u(1/x, 1) = u(1/x \cdot 1, 1/x \cdot x) = |x|^{-p} u(1,x) = |x|^{-p} u(x,1) = |x|^{-p} w(x). \end{aligned}$$

Conversely, if w is a function verifying the above listed properties, one can define

$$u(x,y) := \begin{cases} |y|^p w(x/y), & \text{jos } y \neq 0, \\ |x|^p w(0), & \text{jos } y = 0. \end{cases}$$

Let us prove that then u again satisfies the properties required for it, so searching for w is completely equivalent to searching for u.

Observe first that u has the symmetries $u(y,x) = u(x,y)$ and $u(\alpha x, \alpha y) = |\alpha|^p u(x,y)$.

Proof. If $x \neq 0 \neq y$, then

$$u(y,x) = |x|^p w(y/x) = |x|^p |y/x|^p w(x/y) = |y|^p w(x/y) = u(x,y),$$

and if in addition $\alpha \neq 0$

$$u(\alpha x, \alpha y) = |\alpha y|^p w(\alpha x / \alpha y) = |\alpha|^p |y|^p w(x/y) = |\alpha|^p u(x,y).$$

If $x \neq 0 = y$ and $\alpha \neq 0$, then

$$u(0,x) = |x|^p w(0/x) = |x|^p w(0) = u(x,0), \qquad u(0, \alpha x) = |\alpha|^p |x|^p w(0) = |\alpha|^p u(0,x).$$

Finally $u(0,0) = |0|^p w(0) = 0$, so $u(0 \cdot x, 0 \cdot y) = 0 = |0|^p u(x,y)$. □

Another basic observation is the inequality $w(0) \leq 0$.

Proof. Assume, contrary to the claim, that $w(0) > 0$. Since a concave function is continuous, there holds $w(t) \geq \delta > 0$ for all $|t| \leq \varepsilon$. Thus $w(t^{-1}) = |t|^{-p} w(t) \geq \delta |t|^{-p} \to \infty$, when $t \to 0$. In particular, for M there is an N, such that $w(x) \geq M$, when $|x| \geq N$. But by concavity the holds $w(x) \geq M$ also when $(-N, N)$, and hence on all of $\mathbb{R}$. Thus $w \geq M$ everywhere, and by the arbitrariness of M it follows that $w \equiv +\infty$. This contradicts the real-valuedness of w. □

From the previous observation it follows that $x \mapsto u(x, y)$ is concave for $y = 0$. For other values of y, this follows directly from the concavity of w. The symmetry $u(x, y) = u(y, x)$ implies that u is also concave as a function of y for a fixed x.

If $y \neq 0$, then

$$u(x, y) = |y|^p w(x/y) \geq |y|^p F(x/y, 1) = F(x, y)$$

by the obvious homogeneity of the function F, and finally

$$u(x, 0) = |x|^p w(0) \geq |x|^p F(0, 1) = |x|^p F(1, 0) = F(x, 0),$$

which one could also deduce by taking the limit $y \to 0$, since both $y \mapsto u(x, y)$ (being concave) and $y \mapsto F(x, y)$ (obviously) are continuous.

3.14. **The search for w: miscellanea.** Whether or not it is possible to find $w \geq f$ with the required properties depends on the function f, and then on the constant β in its definition. Let us note that it is necessary that $\beta \geq 1$. Indeed, if $\beta < 1$, the $w(x) \geq f(x) \to \infty$ as $x \to \pm\infty$, and this is impossible for a concave function w, as we saw in Section 3.13.

The function f, and then the problem of finding a dominating w, is a little different depending on whether $p \in (1, 2)$, $p = 2$, or $p \in (2, \infty)$. Let us notice that the case $p = 2$, unlike the other two, is trivial. (This is not surprising, since Burkholder's inequality itself is also easy in this case, see Exercise 2.) Namely, then

$$f(x) = \Big(\frac{x+1}{2}\Big)^2 - \beta^p \Big(\frac{x-1}{2}\Big)^2 = \frac{1}{4}\Big((1 - \beta^2)x^2 + 2(1 + \beta^2)x + (1 - \beta^2)\Big)$$

is already concave for all $\beta \geq 1$. With the minimal choice $\beta = 1$, it follows that $f(x) = x$, and one can take $w(x) = x$, $u(x, y) = xy$.

In the sequel, we will concentrate on the case $p \in (2, \infty)$, leaving the treatment of $p \in (1, 2)$ for the exercises. Note that, just for the proof of Burkholder's inequality, one would not need to repeat all the concave function constructions, but one could derive the inequality for $p \in (1, 2)$ from the case $p \in (2, \infty)$ by a duality argument, see Exercise 1.

3.15. **The shape of the function f.** Let us compute the derivatives

$$\begin{aligned} f(x) &= \Big|\frac{x+1}{2}\Big|^p - \beta^p \Big|\frac{x-1}{2}\Big|^p, \\ f'(x) &= \frac{p}{2}\Big(\operatorname{sgn}(x+1)\Big|\frac{x+1}{2}\Big|^{p-1} - \beta^p \operatorname{sgn}(x-1)\Big|\frac{x-1}{2}\Big|^{p-1}\Big), \\ f''(x) &= \frac{p(p-1)}{4}\Big(\Big|\frac{x+1}{2}\Big|^{p-2} - \beta^p \Big|\frac{x-1}{2}\Big|^{p-2}\Big). \end{aligned}$$

For $k = 0, 2$, it follows that

$$f^{(k)}(x) = 0 \ \Leftrightarrow\ \Big|\frac{x+1}{2}\Big|^{p-k} - \beta^p \Big|\frac{x-1}{2}\Big|^{p-k} \ \Leftrightarrow\ |x+1| = \beta_k |x-1|, \quad \beta_k := \beta^{p/(p-k)}.$$

For $k = 1$, one has the same condition as above, and in addition $\operatorname{sgn}(x+1) = \operatorname{sgn}(x-1)$, i.e., $|x| > 1$. Solving for x,

$$|x+1| = \beta_k|x-1| \ \Leftrightarrow\ \begin{cases} x + 1 = \beta_k(1 - x) \\ x + 1 = \beta_k(x + 1) \end{cases} \ \Leftrightarrow\ \begin{cases} x = x_k \\ x = 1/x_k \end{cases} \quad x_k := \frac{\beta_k - 1}{\beta_k + 1} \in [0, 1).$$

Since $\beta \geq 1$ and $p > 2$, hence $1 = p/(p-0) < p/(p-1) < p/(p-2)$, it follows that $1 \leq \beta_0 \leq \beta_1 \leq \beta_2$ and then $0 \leq x_0 \leq x_1 \leq x_2 < 1 < 1/x_2 \leq 1/x_1 \leq 1/x_0$, and all inequalities are strict if $\beta > 1$. It will be shown shortly that this is the case.

So $f^{(k)}$ has zeros at x_k and $1/x_k$ for $k = 0, 2$ and only at $1/x_1$ for $k = 1$. A routine investigation of the signs reveals that $f^{(k)}$ is positive on $(x_k, 1/x_k)$ and negative on $[x_k, 1/x_k]^\complement$ for $k = 0, 2$, while f' is positive on $(-\infty, 1/x_1)$ and negative on $(1/x_1, \infty)$. The fact that $f'' > 0$ on $(x_2, 1/x_2)$ shows that f itself does not qualify for w, since it fails to be concave on that interval.

3.16. **Lemma** ($\beta \geq p-1$)**.** *If there exists a concave $w \geq f$ with $w(x) = |x|^p w(1/x)$, then $\beta \geq p-1$.*

Proof. Note that $0 = f(x_0) \leq w(x_0)$ and $1 = f(1) \leq w(1)$.

For $x \in (0,1)$, consider the difference quotient

$$\frac{w(1)-w(x)}{1-x} = \frac{w(1)-x^p w(1/x)}{1-x} = w(1)\frac{1-x^p}{1-x} + x^{p-1}\frac{w(1)-w(1/x)}{1/x-1}.$$

A concave function has the one-sided derivatives $D_{\pm}w(x) = \lim_{y\to x\pm}(w(y)-w(x))/(y-x)$ at every point. So in particular, taking the limit $x \nearrow 1$ above, it follows that

$$D_-w(1) = pw(1) - D_+w(1)$$

By the condition $w(x_0) \geq 0$ and concavity, there holds for all $x \in (x_0, 1)$,

$$(*) \qquad \frac{w(1)}{1-x_0} \geq \frac{w(1)-w(x_0)}{1-x_0} \geq \frac{w(1)-w(x)}{1-x} \geq D_-w(1) \geq \frac{1}{2}\big(D_-w(1)+D_+w(1)\big) = \frac{p}{2}w(1).$$

Dividing by $w(1) > 0$, it follows that $p/2 \leq 1/(1-x_0) = (\beta+1)/2$, and this completes the proof. □

3.17. **The form of w on $[x_0, 1]$.** From now on, we fix $\beta := p-1$ and try to find a dominating function w. With this choice of β, the left and the right sides of $(*)$ in the proof of Lemma 3.16 are equal, and hence there must be equality at every step. This gives various pieces of useful information.

First, there must hold $w(x_0) = 0$ and $D_-w(1) = D_+w(1)$. Hence the derivative $w'(1)$ exists and equals $w'(1) = p/2 \cdot w(1)$. Finally, from the equality $(w(1)-w(x))/(1-x) = p/2 \cdot w(1)$, it follows that

$$w(x) = \alpha_p\Big(\frac{p}{2}x - \big(\frac{p}{2}-1\big)\Big), \quad x \in [x_0, 1] = [1-\frac{2}{p}, 1], \qquad \alpha_p := w(1).$$

So w is an affine function on this interval, and it remains to determine the constant α_p. The following Lemma 3.18 implies that there must hold

$$\begin{aligned}\frac{p}{2}\alpha_p = f'(x_0) = w'(x_0) &= \frac{p}{2}\Big(\big(\frac{x_0+1}{2}\big)^{p-1} + \beta^p\big(\frac{1-x_0}{2}\big)^{p-1}\Big)\\ &= \frac{p}{2}\Big(\big(\frac{p-1}{p}\big)^{p-1} + (p-1)^p\big(\frac{1}{p}\big)^{p-1}\Big) = \frac{p}{2}\Big(\frac{p-1}{p}\Big)^{p-1}\big(1+(p-1)\big),\end{aligned}$$

and hence

$$\alpha_p = p\Big(\frac{p-1}{p}\Big)^{p-1}.$$

The lemma still has to be proved:

3.18. **Lemma.** *Let w be concave, $w \geq f$ and $w(x_0) = f(x_0)$ at a point x_0, where f is differentiable. Then also w is differentiable at x_0 and $w'(x_0) = f'(x_0)$.*

Proof. Let $h > 0$. Then

$$\begin{aligned}0 \leq w(x_0 \pm h) - f(x_0 \pm h) &= [w(x_0) \pm hD_{\pm}w(x_0) + o(h)] - [f(x_0) \pm hf'(x_0) + o(h)]\\ &= \pm h[D_{\pm}w(x_0) - f'(x_0)] + o(h).\end{aligned}$$

For the $+$ case, $D_+w(x_0) < f'(x_0)$ would lead to a contradiction, since then the right side would be negative for small h. Similarly, $D_-w(x_0) > f'(x_0)$ would make a contradiction in the $-$ case. Hence

$$D_-w(x_0) \leq f'(x_0) \leq D_+w(x_0) \leq D_-w(x_0),$$

where the last inequality was due to concavity. Thus all the expressions are equal. □

3.19. **Checking that the obtained w works.** So we have seen that if a function w exists, it has to be $w(x) = \alpha_p\big(p/2 \cdot x - (p/2 - 1)\big)$ on the interval $[x_0, 1]$. But does this function satisfy the condition $w \geq f$? This has to be checked.

The second derivative f'' is negative on (x_0, x_2) and positive on $(x_2, 1)$, thus f' is decreasing on the first interval and increasing on the second. Hence, on the whole interval,

$$f' \leq \max\{f'(x_0), f'(1)\} = \max\{w'(x_0), p/2\} = \max\{\alpha_p, 1\} \cdot p/2 = \alpha_p \cdot p/2 \equiv w'.$$

The estimate $\alpha_p \geq 1$, which was used above, is left as an exercise.

So $(f - w)' \leq 0$ and hence $f - w$ is decreasing, thus $f - w \leq (f - w)(x_0) = 0$, and this completes the check.

3.20. **First solution for w.** With the function w determined on $[x_0, 1]$, it is also determined on $[1, 1/x_0]$ due to the requirement that $w(x) = |x|^p w(1/x)$. This gives

$$\begin{aligned} w(x) &= \alpha_p\Big(\frac{p}{2}x^{p-1} - \big(\frac{p}{2} - 1\big)x^p\Big), \qquad x \in [1, 1/x_0] = [1, \frac{p}{p-2}] \\ w'(x) &= \alpha_p\Big(\frac{p}{2}(p-1)x^{p-2} - \frac{p-2}{2}px^{p-1}\Big), \\ w''(x) &= \frac{1}{2}\alpha_p p(p-1)(p-2)x^{p-3}(1-x) \leq 0, \end{aligned}$$

since $p > 2$ and $x \geq 1$. So the function is concave on $[1, 1/x_0]$.

To check that w, which is concave on $[x_0, 1]$ and $[1, 1/x_0]$, is actually concave on the whole interval $[x_0, 1/x_0]$, one has to compare the one-sided derivatives at 1. But substituting above,

$$D_+w(1) = \alpha_p\Big(\frac{p}{2}(p-1) - \frac{p}{2}(p-2)\Big) = \alpha_p\frac{p}{2} = D_-w(1),$$

so this is fine.

Moreover, since $w(x) \geq f(x)$ on $[x_0, 1]$, it follows at once that

$$w(x) = x^p w(x^{-1}) \geq x^p f(x^{-1}) = f(x)$$

also fot $x \in [1, 1/x_0]$. In particular, $w(1/x_0) = f(1/x_0) = 0$. Moreover,

$$(*) \qquad -x^{-2}w'(x^{-1}) = D[w(x^{-1})] = D[x^{-p}w(x)] = -px^{-p-1}w(x) + x^{-p}w'(x).$$

Since the same computation holds with f in place of w, and since $w^{(k)}(x_0) = f^{(k)}(x_0)$ for $k = 0, 1$, it follows that also $w'(1/x_0) = f'(1/x_0)$.

With the function w defined on $[x_0, 1/x_0]$ as above, it is seen that the following function provides the required concave dominant of f on all $\mathbb{R}$:

$$\tilde{w}(x) := \begin{cases} w(x), & x \in [x_0, 1/x_0] = [\frac{p-2}{p}, \frac{p}{p-2}] \\ f(x), & \text{otherwise.} \end{cases}$$

Indeed, we have seen that it is concave on the intervals $(-\infty, x_0]$, $[x_0, 1/x_0]$ and $[1/x_0, \infty)$ and the one-sided derivatives agree at x_0 and $1/x_0$, so it is actually concave on $\mathbb{R}$. The function $\tilde{w}$ also dominates f, and it satisfies $\tilde{w}(x) = |x|^p\tilde{w}(1/x)$, since both f and w do. This completes the search for the function w, and hence the proof of Burkholder's inequality in the case $p \in (2, \infty)$.

The function $\tilde{w}$ is actually the smallest possible solution of the problem. Indeed, on the interval $[x_0, 1/x_0]$ the function was uniquely determined, and on the rest of $\mathbb{R}$, where there could be some freedom, it clearly cannot be smaller, since otherwise it would not dominate f.

3.21. **Second solution for w.** The function w is not uniquely determined outside $[x_0, 1/x_0]$, so if we do not care about having the minimal solution, there is some freedom of choice. Consider the function

$$\begin{aligned} w(x) &:= \alpha_p\Big(\Big|\frac{x+1}{2}\Big| - (p-1)\Big|\frac{x-1}{2}\Big|\Big)\max\{1,|x|\}^{p-1} \\ &= \alpha_p \times \begin{cases} \Big(-\frac{p-2}{2}|x|^p - \frac{p}{2}|x|^{p-1}\Big), & x\in(-\infty,-1], \\ \Big(\frac{p}{2}x - \frac{p-2}{2}\Big), & x\in[-1,1], \\ \Big(-\frac{p-2}{2}x^p + \frac{p}{2}x^{p-1}\Big), & x\in[1,\infty), \end{cases} \end{aligned}$$

which coincides with the earlier definition on the interval $[x_0, 1/x_0]$. It also satisfies the property $w(x) = |x|^p w(1/x)$, as one readily checks.

Clearly w is concave on $[-1,1]$, and the proof above for the concavity on $[1,1/x_0]$ actually works for $[1,\infty)$, since the function has the same expression on this whole half-line. For $x\in(-\infty,-1)$,

$$w''(x) = -\frac{1}{2}\alpha_p p(p-1)(p-2)\big(|x|^{p-2} + |x|^{p-3}\big) \le 0,$$

so it is also concave there. That $D_- w(1) = D_+ w(1)$ was already checked above, and

$$D_- w(-1) = \alpha_p\frac{1}{2}\Big((p-2)p + p(p-1)\Big) = \alpha_p\frac{p}{2}(2p-3) > \alpha_p\frac{p}{2}(4-3) = D_+ w(-1).$$

So the derivative has a discontinuity, but the inequality is to the right direction for concavity.

It remains to prove that $w\ge f$ everywhere. This was already done for $[x_0,1/x_0]$. As for $[-1,x_0]$, there holds $f''\le 0$, and hence f' is decreasing, thus $f'\ge f'(x_0) = w'(x_0)\equiv w'$. Then $f-w$ is increasing, thus $f-w\le (f-w)(x_0) = 0$, and this interval is also dealt with. Finally, the mapping $x\mapsto x^{-1}$ transforms $(-\infty,1]$ onto $[-1,0)$ and $[1/x_0,\infty)$ onto $(0,x_0]$ so the domination on these remaining intervals follows as in $(*)$ of Section 3.20.

3.22. **The biconcave function u.** With simple algebra one checks that

$$\max\{1,|x|\} = \Big|\frac{x+1}{2}\Big| + \Big|\frac{x-1}{2}\Big|,$$

and hence

$$w(x) = \alpha_p\Big(\Big|\frac{x+1}{2}\Big| - \beta_p\Big|\frac{x-1}{2}\Big|\Big)\Big(\Big|\frac{x+1}{2}\Big| + \Big|\frac{x-1}{2}\Big|\Big)^{p-1},$$

where $\beta_p = p-1$. From here one also gets the original biconcave function u, the search for which was reduced to the search for w:

$$u(x,y) = |y|^p w(x/y) = \alpha_p\Big(\Big|\frac{x+y}{2}\Big| - \beta_p\Big|\frac{x-y}{2}\Big|\Big)\Big(\Big|\frac{x+y}{2}\Big| + \Big|\frac{x-y}{2}\Big|\Big)^{p-1}.$$

Note that these formulae also work for $p=2$, since then $\alpha_2=\beta_2=1$ and

$$w(x) = \Big|\frac{x+1}{2}\Big|^2 - \Big|\frac{x-1}{2}\Big|^2 = x, \qquad u(x,y) = xy.$$

As it turns out (see the exercises), they also work for $p\in(1,2)$, provided that one chooses the constants α_p and β_p appropriately. The definitions

$$\alpha_p := p\Big(1-\frac{1}{p^*}\Big)^{p-1}, \qquad \beta_p := p^*-1$$

are good for all $p\in(1,\infty)$.

3.23. **Extensions of Burkholder's inequality.** Burkholder's inequality, as stated and proved above, can be generalized in various ways. The sign transform (multiplication of the martingale differences by $\varepsilon_k\in\{-1,+1\}$) can be replaced by much more general *martingale transforms*, where one multiplies d_k by a *predictable* function $v_k\in L^\infty(\mathscr{F}_{k-1},\mu)$ (with $\mathscr{F}_{-1}:=\mathscr{F}_0$). Observe from the tower property that $v_k d_k$ is again a martingale difference sequence. This version will not be treated in these lectures, however.

The following simpler extension is easy to prove: For all real numbers $\lambda_k\in[-1,1]$, there holds

$$\Big\|\sum_{k=0}^n \lambda_k d_k\Big\|_p \le \beta\Big\|\sum_{k=0}^n d_k\Big\|_p.$$

Proof. 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 $\lambda_k = \sum_{j=1}^\infty \varepsilon_{kj}2^{-j}$. Hence

$$\Big\|\sum_{k=0}^n \lambda_k d_k\Big\|_p = \Big\|\sum_{k=0}^n\sum_{j=1}^\infty \varepsilon_{kj}2^{-j}d_k\Big\|_p \le \sum_{j=1}^\infty 2^{-j}\Big\|\sum_{k=0}^n \varepsilon_{kj}d_k\Big\|_p \le \sum_{j=1}^\infty 2^{-j}\beta\Big\|\sum_{k=0}^n d_k\Big\|_p = \beta\Big\|\sum_{k=0}^n d_k\Big\|_p.$$

□

One can also prove the corresponding results for infinite series:

3.24. **Corollary.** *Let $\sum_{k=0}^\infty d_k$ be a series of martingale differences which converges in $L^p(\mathscr{F},\mu)$. Then, for all $\lambda_k \in [-1,1]$, the series $\sum_{k=0}^\infty \lambda_k d_k$ also converges and satisfies*

$$\Big\|\sum_{k=0}^\infty \lambda_k d_k\Big\|_p \le \beta\Big\|\sum_{k=0}^\infty d_k\Big\|_p.$$

Proof. Given $\varepsilon > 0$, there is $N(\varepsilon)$ so that $\|\sum_{k=m}^n d_k\|_p < \varepsilon$ whenever $n \ge m \ge N(\varepsilon)$. By (a reindexed version of) Burkholder's inequality from Section 3.23, this implies

$$\Big\|\sum_{k=m}^n \lambda_k d_k\Big\|_p < \beta\varepsilon$$

for all $n \ge m \ge N(\varepsilon)$, and this means that the series $\sum_{k=0}^\infty \lambda_k d_k$ converges. Moreover,

$$\Big\|\sum_{k=0}^\infty \lambda_k d_k\Big\|_p = \lim_{n\to\infty}\Big\|\sum_{k=0}^n \lambda_k d_k\Big\|_p \le \beta \lim_{n\to\infty}\Big\|\sum_{k=0}^n d_k\Big\|_p = \beta\Big\|\sum_{k=0}^\infty d_k\Big\|_p,$$

and this completes the proof. □

In a similar way one can prove the result for doubly infinite series,

$$\Big\|\sum_{k=-\infty}^\infty \lambda_k d_k\Big\|_p \le \beta\Big\|\sum_{k=-\infty}^\infty d_k\Big\|_p,$$

whenever the series on the right converges in $L^p(\mathscr{F},\mu)$.

3.25. **Burkholder's inequality with random signs.** If one writes down Burkholder's inequality with the new martingale difference sequence $\varepsilon_k d_k$ in place of d_k, one finds that on the left one gets $\varepsilon_k \cdot \varepsilon_k d_k = d_k$, and hence

$$\Big\|\sum_k d_k\Big\|_p \le \beta\Big\|\sum_k \varepsilon_k d_k\Big\|_p.$$

(The summation could be any of the possibilities $1 \le k \le n$, $1 \le k < \infty$ or $-\infty < k < \infty$ considered above.)

So there is actually a two-sided inequality

$$\beta^{-1}\Big\|\sum_k d_k\Big\|_p \le \Big\|\sum_k \varepsilon_k d_k\Big\|_p \le \beta\Big\|\sum_k d_k\Big\|_p$$

valid for any fixed sequence of signs $\varepsilon_k \in \{-1,+1\}$. Then it also holds if one takes the average over all possible choices of signs!

To be precise, let $(\varepsilon_k)_{k\in\mathbb{Z}}$ denote a sequence of independent random variables on some probability space, with distribution $\mathbb{P}(\varepsilon_k = +1) = \mathbb{P}(\varepsilon_k = -1) = \frac{1}{2}$ and write $\mathbb{E}$ for the expectation, i.e., integral over the probability space with respect to $\mathrm{d}\mathbb{P}$. Raising the above two-sided inequality to power p and taking the expectation with respect to the signs ε_k, it follows that

$$(*) \qquad \beta^{-1}\Big\|\sum_k d_k\Big\|_p \le \Big(\int_\Omega \mathbb{E}\Big|\sum_k \varepsilon_k d_k(x)\Big|^p \mathrm{d}\mu(x)\Big)^{1/p} \le \beta\Big\|\sum_k d_k\Big\|_p.$$

In the situation of Theorem 2.15, when $d_k = \mathbb{E}[f|\mathscr{F}_k] - \mathbb{E}[f|\mathscr{F}_{k-1}]$, $k \in \mathbb{Z}$, there holds $\sum_k d_k = f$, and the middle expression in $(*)$ is seen to give a new equivalent norm for $L^p(\mathscr{F},\mu)$.

3.26. **Exercises.**

1. Let β_p denote the optimal constant in Burkholder's inequality in L^p. Without using any knowledge about the value of β_p, prove that $\beta_{p'} = \beta_p$ for all $p \in (1, \infty)$. (Hint: express g_n as a linear transformation $T_\varepsilon f_n$ of f_n, where T_ε is a suitable linear combination of conditional expectations. What does the adjoint operator look like?)
2. Give a short proof of Burkholder's inequality for $p = 2$, and show that it is actually an equality in this case. (Hint: orthogonality.)
3. Let $\phi : [0, 1] \to [0, 1]$ be a Borel measurable function. Let $A := \{(x, y) \in [0, 1]^2 : y \leq \phi(x)\}$. Let $\mathscr{G} := \mathscr{B}([0, 1]) \times \{\varnothing, [0, 1]\}$ be the product of the Borel σ-algbera in the first coordinate and the trivial σ-algebra in the second. Find the conditional expectation $\mathbb{E}[1_A|\mathscr{G}]$. Conclude that the conditional expectation of a simple function is not necessarily simple.
4. Let $p \in (1, \infty)$ and suppose that β is a number such that $\|f_n^*\|_p \leq \beta \|f_n\|_p$ for all simple martingales $(f_k)_{k=0}^n$. (By Doob's inequality, this is true with $\beta = p'$, but the value is not needed in this exercise.) Let $F : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ be the function $F(x, y) := y^p - \beta^p |x|^p$, and
$$U(x, y) := \sup\Big\{ \int F(f_\infty, \max\{f^*, y\})\,\mathrm{d}t : f = (f_k)_{k=0}^n \text{ simple martingale,}$$
$$f_k : [0, 1) \to \mathbb{R},\ f_0 \equiv x\Big\},$$
where $f_\infty := f_n$ if $f = (f_k)_{k=0}^n$. Check the following properties for all $x \in \mathbb{R}$, $y \in [0, \infty)$ and $\alpha \in (0, \infty)$:
$$U(0, 0) \leq 0, \qquad U(\alpha x, \alpha y) = \alpha^p U(x, y),$$
$$U(x, \max\{|x|, y\}) = U(x, y) \geq F(x, \max\{|x|, y\}).$$
Also prove that $x \mapsto U(x, y)$ is concave for every $y \in [0, \infty)$.
5. For $p \in [2, \infty)$, let $\alpha_p := p\Big(\frac{p-1}{p}\Big)^{p-1}$. Prove the estimate $\alpha_p \geq 1$, which was used in Section 3.19.

The following exercises deal with the search for a concave function $w : \mathbb{R} \to \mathbb{R}$ such that $w(x) = |x|^p w(1/x)$ for all $x \neq 0$, and
$$w(x) \geq f(x) := \Big|\frac{x+1}{2}\Big|^p - \beta^p \Big|\frac{x-1}{2}\Big|^p,$$
in the case when $p \in (1, 2)$. A general hint: use similar ideas as in the case $p \in (2, \infty)$ on the lecture, but note that some details are a little different.

6. Find the zeros of f and its first and second derivative, and sketch the graph of f.
7. Prove that if w exists, then $\beta \geq (p-1)^{-1}$. (Hint: Consider the one-sided derivatives of w at the point $x = -1$. In order to divide a certain inequality by a constant, you will need to know that $w(-1) < 0$. Conclude this from the concavity and what you know about $w(1)$ and $w(0)$; see section 3.13 in the notes.)
8. From now on, fix $\beta := (p-1)^{-1}$. Show that w (if it exists) must be an affine function (i.e., of the form $ax + b$) on the interval $[-1, x_0]$, where x_0 is the smaller zero of f. Find the formula for w on this interval. (Hint: the derivatives of w and f must be equal at a certain point.)
9. Check that the function w, which you obtained in the previous exercise, satisfies the inequality $w \geq f$ on $[-1, x_0]$.
10. Where else is the function w uniquely determined by the conditions posed on it? Write down the formula for w (maybe piecewise defined). How could you define w on the rest of $\mathbb{R}$ to satisfy the required conditions?

3.27. **References.** The presentation follows Burkholder's summer school lectures [4] quite closely, with the addition of some details here and there. The inequality (but not yet with the optimal constant) was originally proved in [2]. (It is only contained as part of the proof of Theorem 9 in [2], not as a separate result!) The optimal constant was obtained in [3].

4. Petermichl's dyadic shift and the Hilbert transform

4.1. **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 2.16(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.

4.2. **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 2.16(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 $j \to -\infty$.

4.3. **Haar functions.** Let $L^p(\mathbb{R}) := L^p(\mathscr{B}(\mathbb{R}), \mathrm{d}x)$. By Theorem 2.15, 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 3.25), it follows that

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

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

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

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 4.3 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

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

4.5. **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 $f \in L^p(\mathbb{R})$,

$$Hf := \lim_{\substack{\varepsilon\to 0\\ R\to\infty}} H_{\varepsilon,R}f$$

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.

4.6. **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 $\text{Ш}f = H_I$, but if $f = h_J$, where $J \notin \mathscr{D}$ is a slightly translated or dilated version of I, then $\text{Ш}f$ has a much more complicated expression.

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

4.7. **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:

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

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

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

is pointwise dominated by $2M^{\beta,r}f(x)$ and converges a.e. to $\text{Ш}^{\beta,r}f(x)$ 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 $\text{Ш}^{\beta,r}f(x)$.

To see that $(\beta,r)\mapsto \text{Ш}^{\beta,r}f(x)$ 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 $\text{Ш}^{\beta,r}$, there holds

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

In particular, this shows that

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

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

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

which, unravelling the definition of $\text{Ш}_{-m,n}^{\beta,r}$, is the same as $(*)$. Finally, since the right side above is dominated by $\int \mathrm{d}r/r\int\mathrm{d}\mu(\beta)M^{\beta,r}\{\text{Ш}^{\beta,r}f\}\in L^p(\mathbb{R})$, 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

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

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

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

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

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

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

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

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

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

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

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

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.

4.9. **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 $\langle \text{Ш} \rangle f$. 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

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

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

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

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 $[-1,1]$.

4.10. **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

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

□

4.11. **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 2.16(6), $|\phi_R * f| \leq C'Mf \in L^p(\mathbb{R})$, and hence the remaining claim follows from dominated convergence. □

4.12. **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

$$\begin{aligned}
\phi_\varepsilon * f(x) - af(x) &= \int\phi(y)[f(x-\varepsilon y)-f(x)]\,\mathrm{d}y, \\
\|\phi_\varepsilon * f - af\|_p &\leq \int|\phi(y)|\cdot\|f(\cdot-\varepsilon y)-f\|_p\,\mathrm{d}y.
\end{aligned}$$

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 $0<\varepsilon \le |y|^{-1}$, $g(\cdot-\varepsilon y)-g$ 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

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

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

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

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

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

Hence

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

which can be made arbitrarily small. □

Now the proof of

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

is complete.

4.13. **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 4.3 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: same trick as in 3.23.)
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 4.4.)

4.14. **References.** Petermichl's representation for the Hilbert transform as an average of the dyadic shifts is from [9]. The precise form of the representation and its proof given above are somewhat modified from the original ones, so that the convergence to the Hilbert transform is obtained in a stronger sense. The proof of the L^p boundedness of the dyadic shift, and the notation 'Ш', are taken from [8]. 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.

The L^p boundedness of the Hilbert transform is originally a classical result of M. Riesz [10]. 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.

5. Back to Burkholder's inequality

5.1. **The Fourier transform.** For $f \in L^1(\mathbb{R})$ and $\xi \in \mathbb{R}$, one defines

$$\hat{f}(\xi) := \int_{\mathbb{R}} f(x) e^{-i2\pi x\cdot\xi}\, \mathrm{d}x.$$

By taking absolute values inside the norm, it is clear that $|\hat{f}(\xi)| \leq \|f\|_1$ for all ξ. If $\xi \to \xi_0$, the continuity of the exponential function implies that $e^{-i2\pi x\cdot\xi} \to e^{-i2\pi x\cdot\xi_0}$ for every x, and it follows from dominated convergence that $\hat{f}(\xi) \to \hat{f}(\xi_0)$. So the function $\hat{f}$ is continuous, and in particular measurable. If $f, g \in L^1(\mathbb{R})$, then $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$.

Important properties of the Fourier transform follow just from the knowledge of the transform of one particular function:

5.2. **Lemma.** *The function $\phi(x) := e^{-\pi x^2}$ satisfies $\hat{\phi} = \phi$.*

Proof. Note that $\hat{\phi}(0)$ is the familiar Gaussian integral,

$$\hat{\phi}(0)^2 = \int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\pi(x^2+y^2)}\, \mathrm{d}x\, \mathrm{d}y = \int_0^\infty \int_0^{2\pi} e^{-\pi r^2}\, \mathrm{d}\theta r\, \mathrm{d}r = \int_0^\infty 2\pi r e^{-\pi r^2}\, \mathrm{d}r = \Big|_0^\infty - e^{-\pi r^2} = 1,$$

and hence $\hat{\phi}(0) = 1$, since it is clearly positive. Now there are (at least) two ways to finish the proof.

(*By Cauchy's theorem for complex path integrals.*) Completing the square

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\pi(x^2+i2x\cdot\xi+(i\xi)^2-(i\xi)^2)}\, \mathrm{d}x = \int_{-\infty}^{\infty} e^{-\pi(x+i\xi)^2}\, \mathrm{d}x \cdot e^{-\pi\xi^2} = \int_{-\infty+i\xi}^{\infty+i\xi} e^{-\pi z^2}\, \mathrm{d}z \cdot e^{-\pi\xi^2}.$$

By Cauchy's theorem, it is easy to check that one can shift the integration path back to the real axis, and this was just computed above.

(*By the uniqueness theory of ordinary differential equations.*) Notice that $\phi'(x) = -2\pi x\phi(x)$. Taking the Fourier transform of both sides and integrating by parts, it follows that $i2\pi\xi\hat{\phi}(\xi) = -i\hat{\phi}'(\xi)$. Hence both ϕ and $\hat{\phi}$ are solutions of the differential equation

$$u'(x) = -2\pi x u(x), \qquad u(0) = 1,$$

and therefore must be equal. □

By Lemma 5.2 (interchanging the roles of x and ξ)

$$\phi(x) = \hat{\phi}(x) = \int_{\mathbb{R}} \phi(\xi) e^{-i2\pi x\cdot\xi}\, \mathrm{d}\xi.$$

Since ϕ is real-valued, taking complex conjugates of both sides one can replace $-i$ by $+i$ in the exponent. Substituting x/ε in place of x and changing integration variables, one further obtains

$$(*) \qquad \phi_\varepsilon(x) = \frac{1}{\varepsilon}\phi\Big(\frac{x}{\varepsilon}\Big) = \int_{\mathbb{R}} \phi(\varepsilon\xi) e^{i2\pi x\cdot\xi}\, \mathrm{d}\xi.$$

5.3. **Theorem** (Fourier inversion). *Suppose that both $f \in L^1(\mathbb{R})$ and $\hat{f} \in L^1(\mathbb{R})$. Then, a.e.,*

$$f(x) = \int_{\mathbb{R}} \hat{f}(\xi) e^{i2\pi x\cdot\xi} \,\mathrm{d}\xi.$$

Proof. By Lemma 4.12, $\phi_\varepsilon * f \to f$ in $L^1(\mathbb{R})$, and hence a subsequence converges pointwise a.e. Taking the limit along this subsequence and using the above formula $(*)$ for ϕ_ε, it follows that

$$\begin{aligned} f(x) &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \phi_\varepsilon(y) f(x-y)\,\mathrm{d}y \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}}\int_{\mathbb{R}} \phi(\varepsilon\xi) e^{i2\pi y\cdot\xi}\,\mathrm{d}\xi f(x-y)\,\mathrm{d}y \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \phi(\varepsilon\xi) \int_{\mathbb{R}} e^{-i2\pi(x-y)\cdot\xi} f(x-y)\,\mathrm{d}y\, e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi \\ &= \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}} \phi(\varepsilon\xi)\hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi = \int_{\mathbb{R}} \hat{f}(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi, \end{aligned}$$

where the second to last equality was the definition of $\hat{f}(\xi)$, and the last one was dominated convergence based on the fact that $\phi(\varepsilon\xi) \to \phi(0) = 1$ at every ξ. □

5.4. **Corollary** (Uniqueness of the representation). *If $f \in L^1(\mathbb{R})$ has the representation*

$$f(x) = \int_{\mathbb{R}} g(\xi) e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi$$

for some $g \in L^1(\mathbb{R})$, then $g = \hat{f}$.

Proof. Let $h(x) := g(-x)$. A change of variables shows that $f = \hat{h}$. Hence $h, \hat{h} \in L^1$ and thus $g(-\xi) = h(\xi) = \int \hat{h}(x) e^{i2\pi x\cdot\xi}\,\mathrm{d}x = \int f(x) e^{i2\pi x\cdot\xi}\,\mathrm{d}x$. □

5.5. **The Fourier transform of the Hilbert transform.** Let $f \in L^1(\mathbb{R})$ and consider the truncated Hilbert transform $H_{\varepsilon,R} f$. This is $k_{\varepsilon,R} * f$, where $k_{\varepsilon,R}(x) = 1_{\varepsilon<|x|<R}/\pi x$. Hence $\widehat{H_{\varepsilon,R} f}(\xi) = \hat{k}_{\varepsilon,R}(\xi)\hat{f}(\xi)$. So one should compute

$$\pi\hat{k}_{\varepsilon,R}(\xi) = \int_{\varepsilon<|x|<R} e^{-i2\pi x\cdot\xi}\frac{\mathrm{d}x}{x} = -\operatorname{sgn}(\xi)\int_{i[-R,-\varepsilon]\cup i[\varepsilon,R]} e^{2\pi|\xi|z}\frac{\mathrm{d}z}{z}$$

By Cauchy's theorem, the integration path may be shifted from the union of the two vertical line segment to the semicircles on the left halfplane connecting their endpoints; thus the above integral is equal to

$$\int_{\pi/2}^{3\pi/2} \big(e^{2\pi|\xi|\varepsilon e^{i\theta}} - e^{2\pi|\xi|Re^{i\theta}}\big) i\,\mathrm{d}\theta.$$

The second term is dominated by

$$\begin{aligned} \Big|\int_{\pi/2}^{3\pi/2} e^{2\pi|\xi|Re^{i\theta}}\,\mathrm{d}\theta\Big| &\le \int_{\pi/2}^{3\pi/2} e^{2\pi|\xi|R\cos\theta}\,\mathrm{d}\theta = 2\int_0^{\pi/2} e^{-2\pi|\xi|R\sin\theta}\,\mathrm{d}\theta \\ &\le 2\int_0^{\pi/2} e^{-4|\xi|R\theta}\,\mathrm{d}\theta = 2\frac{1-e^{-2\pi|\xi|R}}{4|\xi|R}, \end{aligned}$$

where the estimate $\sin\theta \ge 2/\pi\cdot\theta$ for $\theta \in [0,\pi/2]$ was used. The result is uniformly bounded by π (since $1 - e^{-x} \le x$) and tends to zero as $R \to \infty$.

The same upper bound applies to the first term, which clearly tends to

$$\int_{\pi/2}^{3\pi/2} i\,\mathrm{d}\theta = i\pi$$

as $\epsilon \searrow 0$. So the conclusion is that

$$\hat{k}_{\varepsilon,R}(\xi) = -i\operatorname{sgn}(\xi) m_{\varepsilon,R}(\xi),$$

where $|m_{\varepsilon,R}(\xi)| \le 2$ and $m_{\varepsilon,R}(\xi) \to 1$ as $\varepsilon \searrow 0$ and $R \nearrow \infty$.

Suppose then that $f, \hat{f} \in L^1(\mathbb{R})$. Then clearly $H_{\varepsilon,R} f = k_{\varepsilon,R} * f$ and $\widehat{H_{\varepsilon,R} f} = \hat{k}_{\varepsilon,R}\hat{f}$ belong to $L^1(\mathbb{R})$ as well. Hence

$$H_{\varepsilon,R} f(x) = \int_{\mathbb{R}} \widehat{H_{\varepsilon,R} f}(\xi) e^{i2\pi x\cdot\xi} \,\mathrm{d}\xi = \int_{\mathbb{R}} -i \operatorname{sgn}(\xi) m_{\varepsilon,R}(\xi)\hat{f}(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi.$$

As $\varepsilon \searrow 0$, $R \nearrow \infty$, one can use dominated convergence on the right, and the convergence $H_{\varepsilon,R} f \to Hf$ on the left to conclude that

$$Hf(x) = \int_{\mathbb{R}} -i\operatorname{sgn}(\xi)\hat{f}(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi$$

for all $f \in L^1$ with $\hat{f} \in L^1$, and so

$$(*) \qquad \widehat{Hf}(\xi) = -i\operatorname{sgn}(\xi)\hat{f}(\xi)$$

by the uniqueness of the representation. For such functions f, it also follows that

$$Hf = \begin{cases} -if, & \text{if } \{\hat{f} \neq 0\} \subseteq [0, +\infty), \\ +if, & \text{if } \{\hat{f} \neq 0\} \subseteq (-\infty, 0]. \end{cases}$$

5.6. **Theorem** (Conjugation of trigonometric polynomials). *Let $f(x) := \sum_{k\neq 0} a(k)e^{i2\pi xk}$ be a trigonometric polynomial (a finite sum) and $g(x) := \sum_{k\neq 0} -i\operatorname{sgn}(k)a(k)e^{i2\pi xk}$ its conjugate polynomial. Then $\|g\|_p \leq \alpha\|f\|_p$, where the L^p norms are taken on $[0, 1)$, and α is the norm of the Hilbert transform on $L^p(\mathbb{R})$.*

Proof. Let $\phi \not\equiv 0$ be a function on $\mathbb{R}$ with $|\phi(x)| \leq C(1 + |x|)^{-1-\delta}$ and $\{\hat{\phi} \neq 0\} \subset [-R, R]$ for some $R, \delta > 0$. (See the exercises for the existence of such a ϕ.) Let $f^\varepsilon(x) := f(x)\phi(\varepsilon x)$. Then $f^\varepsilon \in L^1(\mathbb{R})$ and

$$\hat{f}^\varepsilon(\xi) = \sum_{k\neq 0} a(k)\int_{\mathbb{R}} \phi(\varepsilon x)e^{i2\pi x(k-\xi)}\,\mathrm{d}x = \sum_{k\neq 0} a(k)\frac{1}{\varepsilon}\hat{\phi}\Big(\frac{\xi - k}{\varepsilon}\Big),$$

so that also $\hat{f}^\varepsilon \in L^1(\mathbb{R})$. One defines g^ε similarly and makes the same observations.

Now $\hat{\phi}((\xi - k)/\varepsilon) \neq 0$ only if $|\xi - k| \leq \varepsilon R$, which implies in particular that ξ and k have the same sign if $\varepsilon R < 1$. Hence one finds that

$$-i\operatorname{sgn}(\xi)\hat{f}^\varepsilon(\xi) = \sum_{k\neq 0} -i\operatorname{sgn}(k)a(k)\frac{1}{\varepsilon}\hat{\phi}\Big(\frac{\xi - k}{\varepsilon}\Big) = \hat{g}^\varepsilon(\xi),$$

and thus $Hf^\varepsilon = g^\varepsilon$. It follows that $\|g^\varepsilon\|_p \leq \alpha\|f^\varepsilon\|_p$, where

$$\|f^\varepsilon\|_p^p = \int_{\mathbb{R}} |f(x)|^p|\phi(\varepsilon x)|^p\,\mathrm{d}x = \frac{1}{\varepsilon}\int_0^1 |f(x)|^p\Big(\sum_{k\in\mathbb{Z}} \varepsilon|\phi(\varepsilon(x + k))|^p\Big)\,\mathrm{d}x$$

using the periodicity $f(x + k) = f(x)$ of f. The quantity in parentheses is bounded uniformly in ε and x and, as a Riemann sum, converges to $\int_{\mathbb{R}} |\phi(y)|^p\,\mathrm{d}y$ as $\varepsilon \searrow 0$. Similar observations hold for g^ε, and hence $\int_0^1 |g(x)|^p\,\mathrm{d}x \leq \alpha^p \int_0^1 |f(x)|^p\,\mathrm{d}x$, as claimed. □

5.7. **Back to Burkholder's inequality.** Suppose now that we had not proved Burkholder's inequality but that we know the L^p boundedness of the Hilbert transform. The aim is to give another proof of Burkholder's inequality based on this information. The formula $(*)$ from Section 5.5 gives some hope by showing that H can be viewed as a kind of a sign transformation, too.

Recall (Lemmas 3.3 and 3.4) that it suffices to prove Burkholder's inequality in the case that $(d_k)_{k=0}^\infty$ is a simple martingale difference sequence on a finite measure space $(\Omega, \mathscr{F}, \mu)$. Since one can take the filtration to be $\mathscr{F}_k := \sigma(d_0, \ldots, d_k)$, it can be assumed simple, too. Replacing the measure μ by $\tilde{\mu}(E) := \mu(E)/\mu(\Omega)$, one may further assume without loss of generality that $\mu(\Omega) = 1$.

Here are some further reductions:

5.8. **Lemma** (Reduction to zero-start). *It suffices to prove Burkholder's inequality in the case when $f_0 = d_0 = 0$.*

Proof. Consider the new space $\tilde{\Omega} := \Omega \times \{-1,+1\}$ with the product σ-algebra and measure, where $\{-1,+1\}$ is equipped with $\nu(\{-1\}) = \nu(\{+1\}) = 1/2$. Let $\mathscr{G}_0 := \sigma(\{-1,+1\})$ and $\mathscr{G}_1 := \sigma(\{-1\},\{+1\})$. A new filtration is given by $\tilde{\mathscr{F}}_k := \mathscr{F}_k \times \mathscr{G}_1$ for $k = 0,\ldots,n$ and $\tilde{\mathscr{F}}_{-1} := \mathscr{F}_0 \times \mathscr{G}_0$, and a new martingale by $\tilde{f}_k(\omega,\eta) := \eta f_k(\omega)$ for $k = 0,\ldots,n$, $\omega \in \Omega$ and $\eta \in \{-1,+1\}$, and $\tilde{f}_{-1}(\omega,\eta) := 0$. The martingale property $\mathbb{E}[\tilde{f}_k|\tilde{\mathscr{F}}_{k-1}]$ is follows from the martingale property of $(f_k)_{k=0}^n$ for $k = 1,\ldots,n$ and from $\int_{\{-1,+1\}} \eta \, \mathrm{d}\nu(\eta) = 0$ for $k = 0$. Moreover,

$$\Big\| \sum_{k=-1}^{n} \varepsilon_k \tilde{d}_k \Big\|_p = \Big\| \eta \sum_{k=0}^{n} \varepsilon_k d_k \Big\|_p = \Big\| \sum_{k=0}^{n} \varepsilon_k d_k \Big\|_p,$$

so Burkholder's inequality for $(\tilde{d}_k)_{k=-1}^n$ implies the same result for $(d_k)_{k=0}^n$. Of course one can shift the indexing of $\tilde{d}_k$ so as to start from zero. □

5.9. **Lemma** (Reduction to one-by-one increments)**.** *It suffices to prove Burkholder's inequality in the case when $\mathscr{F}_k = \sigma(\mathscr{A}_k)$, where $\mathscr{A}_k$ is a partition of Ω with the number of elements $\#\mathscr{A} = k+1$ for every k.*

Proof. Things have already been reduced to the situation where each $\mathscr{F}_k$ is finite, and then it is of the form $\mathscr{F}_k = \sigma(\mathscr{A}_k)$, where the partition $\mathscr{A}_k$ consists of the minimal non-empty sets in $\mathscr{F}_k$. Moreover, since $f_0 = 0$ is a constant, $\mathscr{F}_0 = \sigma(f_0)$ is trivial and hence $\mathscr{A}_0 = \{\Omega\}$, which already satisfies the requirement $\#\mathscr{A}_0 = 1$. The idea of the proof is very simple: one just adds finitely many new σ-algebras between $\mathscr{F}_{k-1}$ and $\mathscr{F}_k$ in order to satisfy the condition that the underlying partitions only grow by one set at a time. Here are the details:

Let $\mathscr{A}_k := \{A_1,\ldots,A_{N(k)}\}$, and

$$\mathscr{F}_{k-1,r} := \sigma(\mathscr{F}_{k-1}, A_1,\ldots,A_r), \qquad r = 0,\ldots,N(k).$$

Then

$$\mathscr{F}_{k-1} = \mathscr{F}_{k-1,0} \subseteq \mathscr{F}_{k-1,1} \subseteq \mathscr{F}_{k-1,2} \subseteq \mathscr{F}_{k-1,N(k)} = \mathscr{F}_k$$

and $\mathscr{F}_{k-1,r+1} = \sigma(\mathscr{F}_{k-1,r}, A_{r+1})$. If $\mathscr{A}_{k-1,r}$ is the collection of minimal non-empty sets in $\mathscr{F}_{k-1,r}$, then depending on whether $A_{r+1} \in \mathscr{F}_{k-1,r} = \sigma(\mathscr{A}_{k-1,r})$ or not, there holds

$$\#\mathscr{A}_{k-1,r+1} - \#\mathscr{A}_{k-1,r} \in \{0,1\}.$$

So the doubly-indexed filtration $(\mathscr{F}_{k-1,r} : k = 1,\ldots,n; r = 0,\ldots,N(k))$ (where the pairs (i,j) are ordered so that $(i,j) < (i',j')$ iff $i < i'$ or $[i = i'$ and $j < j']$) has the property that the associated partitions increase by at most one at the time. By relabelling this filtration as $(\mathscr{G}_j)_{j=0}^M$, where all the repetitions of the same σ-algebra have been removed, one obtains a filtration with the required property. The original σ-algebras form a subfiltration, $\mathscr{F}_k = \mathscr{G}_{J(k)}$ for some $0 = J(0) \leq J(1) \leq \ldots \leq J(n) = M$.

Then one can simply define a new martingale by $g_j := \mathbb{E}[f_n|\mathscr{G}_j]$, and observe that (recalling $f_0 = g_0 = 0$)

$$\sum_{k=1}^{n} \varepsilon_k (f_k - f_{k-1}) = \sum_{k=1}^{n} \varepsilon_k \sum_{j=J(k-1)+1}^{J(k)} (g_j - g_{j-1}) = \sum_{j=1}^{M} \varepsilon'_j (g_j - g_{j-1}),$$

where $\varepsilon'_j := \varepsilon_k$ for $j \in (J(k-1), J(k)]$. □

5.10. **From an abstract space to** $[0,1)^n$**.** In order to make use of the Hilbert transform, one still has to be able to reduce martingales on an abstract measure space $(\Omega, \mathscr{F}, \mu)$ to ones involving a real variable. But this is easily done after the reductions in Lemmas 5.8 and 5.9. Recall the situation: There are finite partitions $\mathscr{A}_k = \{A_{k0},\ldots,A_{kk}\}$ of Ω, and $\mathscr{A}_k$ is a refinement of $\mathscr{A}_{k-1}$; also $d_0 = 0$. The refinement property can be stated as follows: for all $k \in \{1,\ldots,n\}$, there are $r,s \in \{0,\ldots,k\}$, $t \in \{0,\ldots,k-1\}$ and a bijection $\pi : \{0,\ldots,k\} \setminus \{r,s\} \to \{0,\ldots,k-1\} \setminus \{t\}$ such that

$$A_{kj} = A_{k-1,\pi(j)}, \quad j \notin \{r,s\}, \qquad A_{kr} \cup A_{ks} = A_{k-1,t} \quad \text{(disjoint union)}.$$

One defines a sequence of sets with the same measure-theoretic properties as follows:

$$Q_{10} := I_{10} := \big[0, \mu(A_{10})\big), \qquad Q_{11} := I_{11} := \big[\mu(A_{10}), 1\big),$$

and then inductively

$$\begin{aligned} Q_{kj} &:= Q_{k-1,\pi(j)} \times [0,1), \quad j \notin \{r,s\}, \\ Q_{kr} &:= Q_{k-1,t} \times I_{kr} := Q_{k-1,t} \times \Big[0, \frac{\mu(A_{kr})}{\mu(A_{k-1,t})}\Big), \\ Q_{ks} &:= Q_{k-1,t} \times I_{ks} := Q_{k-1,t} \times \Big[\frac{\mu(A_{kr})}{\mu(A_{k-1,t})}, 1\Big). \end{aligned}$$

Thus $Q_{kj} \subseteq [0,1)^k$ is a Cartesian product of intervals, and by induction one checks $(Q_{kj})_{j=0}^k$ is a partition of $[0,1)^k$ with $|Q_{kj}| = \mu(A_{kj})$. Let further $R_{kj} := Q_{kj} \times [0,1)^{n-k}$, so that $(R_{kj})_{j=0}^k$ is a partition of $[0,1)^n$ which refines $(R_{k-1,j})_{j=0}^{k-1}$.

Now consider a martingale difference sequence $(d_k)_{k=0}^n$ on $(\Omega, \mathscr{F}, \mu)$, adapted to $(\mathscr{F}_k)_{k=0}^n$, where $\mathscr{F}_k = \sigma(\mathscr{A}_k)$. By $\mathscr{F}_k$-measurability,

$$d_k = \sum_{j=0}^k a_{kj} 1_{A_{kj}} = 1_{A_{k-1,t}}\big(a_{kr}1_{A_{kr}} + a_{ks}1_{A_{ks}}\big) + \sum_{j \notin \{r,s\}} a_{kj} 1_{A_{k-1,\pi(j)}}.$$

The condition that $\mathbb{E}[d_k|\mathscr{F}_{k-1}] = 0$ says that $a_{kj} = 0$ for $j \notin \{r,s\}$ and $a_{kr}\mu(A_{kr}) + a_{ks}\mu(A_{ks}) = 0$, so that

$$d_k = a_{kr}1_{A_{kr}} + a_{ks}1_{A_{ks}}.$$

Defining new functions on $[0,1)^n$ by

$$\tilde{d}_k := a_{kr}1_{R_{kr}} + a_{ks}1_{R_{ks}},$$

these have exactly the same measure-theoretic properties as the d_k, so proving Burkholder's inequality amounts to showing that

$$\Big\|\sum_{k=1}^n \varepsilon_k \tilde{d}_k\Big\|_p \leq C \Big\|\sum_{k=1}^n \tilde{d}_k\Big\|_p.$$

Here

$$\begin{aligned} \tilde{d}_k(x) &= a_{kr}1_{R_{kr}}(x) + a_{ks}1_{R_{ks}}(x) \\ &= 1_{Q_{k-1,t}}(x_1, \ldots, x_{k-1})\big(a_{kr}1_{I_{kr}}(x_k) + a_{ks}1_{I_{ks}}(x_k)\big) = \prod_{j=1}^k \phi_{kj}(x_j), \end{aligned}$$

where the ϕ_{kj} are indicators of some intervals for $j = 1, \ldots, k-1$, and $\int \phi_{kk}(t)\,\mathrm{d}t = 0$.

So Burkholder's inequality will follow from:

5.11. **Theorem** (Bourgain). *Let $\phi_{kj} \in L^p(0,1)$ for $1 \leq j \leq k \leq n$ and $\varepsilon_k \in \{-1,+1\}$. Then for all $p \in (1,\infty)$,*

$$\Big\|\sum_{k=1}^n \varepsilon_k \prod_{j=1}^k \phi_{kj}(x_j)\Big\|_p \leq \alpha^2 \Big\|\sum_{k=1}^n \prod_{j=1}^k \phi_{kj}(x_j)\Big\|_p, \qquad \text{if } \int_0^1 \phi_{kk}\,\mathrm{d}t = 0.$$

where $\alpha = \alpha_p$ is the norm of the Hilbert transform on $L^p(\mathbb{R})$.

Note that this does not recover the optimal constant in Burkholder's inequality, even if one used the (known) best constant for the Hilbert transform.

Essentially the same proof would show that the estimate holds more generally for functions of the form $\phi_{k0}(x_1, \ldots, x_{k-1})\phi_{kk}(x_k)$, where $\phi_{k0}(x_1, \ldots, x_{k-1})$ need not split as a product $\prod_{j=1}^{k-1} \phi_{kj}(x_j)$ but, as observed above, the above formulation suffices for the proof of Burkholder's inequality.

5.12. **Reduction to trigonometric polynomials.** By known density results, one may find trigonometric polynomials $\tilde{\phi}_{kj}$ so that $\|\tilde{\phi}_{kj} - \phi_{kj}\|_p < \delta$. One may further choose $\tilde{\phi}_{kk}$ to have vanishing integral. Indeed, since $\int \phi_{kk} = 0$, it follows that

$$\Big| \int_0^1 \tilde{\phi}_{kk}\,\mathrm{d}t \Big| = \Big| \int_0^1 (\tilde{\phi}_{kk} - \phi_{kk})\,\mathrm{d}t \Big| \leq \|\tilde{\phi}_{kk} - \phi_{kk}\|_p < \delta,$$

so $\tilde{\phi}_{kk} - \int_0^1 \tilde{\phi}_{kk}\,\mathrm{d}t$ is still a good approximation to ϕ_{kk}.

Finally, one observes that

$$\begin{aligned} \Big\| \prod_{j=1}^{k} \phi_{kj}(x_j) - \prod_{j=1}^{k} \tilde{\phi}_{kj}(x_j) \Big\|_p &\leq \sum_{r=1}^{k} \Big\| \prod_{j=1}^{r-1} \phi_{kj}(x_j) \times \big(\phi_{kr}(x_r) - \tilde{\phi}_{kr}(x_r)\big) \times \prod_{j=r+1}^{k} \tilde{\phi}_{kj}(x_j) \Big\|_p \\ &= \sum_{r=1}^{k} \prod_{j=1}^{r-1} \|\phi_{kj}\|_p \times \|\phi_{kr} - \tilde{\phi}_{kr}\|_p \times \prod_{j=r+1}^{k} \|\tilde{\phi}_{kj}\|_p, \end{aligned}$$

and this can made arbitrarily small by making δ small.

Hence it suffices to prove Theorem 5.11 for trigonometric polynomials.

5.13. **Bourgain's transformation.** The inequality to be proved still involves n real variables, whereas the Hilbert transform (or the analogous conjugation operation from Theorem 5.6) acts on functions of a single variable. Thus, for fixed $x \in [0,1)^n$ and $y \in \mathbb{Z}^n$, one defines a new periodic function of $t \in \mathbb{R}$ by

$$f(t) := \sum_{k=1}^{n} \prod_{j=1}^{k} \phi_{kj}(x_j + y_j t).$$

Since

$$\phi_{kj}(u) = \sum_{-M \leq m \leq M} a_{kj}(m) e^{i2\pi m u}, \qquad a_{kk}(0) = \int_0^1 \phi_{kk}\,\mathrm{d}u = 0,$$

(one can choose a fixed M, since there are only finitely many ϕ_{kj}) it follows that

$$f(t) = \sum_{k=1}^{n} \sum_{-M \leq m_{k1},\dots,m_{kk} \leq M} \Big(\prod_{j=1}^{k} a_{kj}(m_{kj}) e^{i2\pi x_j m_{kj}} \Big) e^{i2\pi (\sum_{j=1}^{k} y_j m_{kj}) t}.$$

The conjugate trigonometric polynomial $g(t)$ is obtained by multiplying each term in the multiple sum by $-i\,\mathrm{sgn}(\sum_{j=1}^{k} y_j m_{kj})$. Note that here $m_{kk} \neq 0$, since $a_{kk}(0) = 0$. Also $|\sum_{j=1}^{k-1} y_j m_{kj}| \leq M \sum_{j=1}^{k-1} |y_j|$. Now choose the y_k's inductively so that (e.g.) $|y_1| = 1$ and

$$(*) \qquad |y_k| > M \sum_{j=1}^{k-1} |y_j|, \qquad k > 1.$$

This ensures that $\mathrm{sgn}(\sum_{j=1}^{k} y_j m_{kj}) = \mathrm{sgn}(y_k m_{kk}) = \mathrm{sgn}(y_k)\,\mathrm{sgn}(m_{kk})$. Thus the conjugate function $g(t)$ factorizes into the form

$$g(t) = \sum_{k=1}^{n} \mathrm{sgn}(y_k) \prod_{j=1}^{k-1} \phi_{kj}(x_j + y_j t) \psi_{kk}(x_k + y_k t),$$

where

$$\psi_{kk}(u) = \sum_{-M \leq m \leq M} -i\,\mathrm{sgn}(m) a_{kk}(m) e^{i2\pi m u}$$

is the conjugate function of ϕ_{kk}.

5.14. **Completion of the proof.** By Theorem 5.6 there holds $\|g\|_p \leq \alpha \|f\|_p$, i.e.,

$$\begin{aligned}\int_0^1 \Big| \sum_{k=1}^n \operatorname{sgn}(y_k) \prod_{j=1}^{k-1} \phi_{kj}(x_j + y_j t)\psi_{kk}(x_k + y_k t)\Big|^p \,\mathrm{d}t \\ \leq \alpha^p \int_0^1 \Big| \sum_{k=1}^n \prod_{j=1}^{k-1} \phi_{kj}(x_j + y_j t)\phi_{kk}(x_k + y_k t)\Big|^p \,\mathrm{d}t.\end{aligned}$$

Integrate this with respect to $\mathrm{d}x$ over $[0,1)^n$, change the order of integration (bringing the $\mathrm{d}x$ integral inside) and make the change of variables $x_j + y_j t \leftrightarrow x_j$. By periodicity, integration over $x_j \in [y_j t, y_j t + 1)$ gives the same result as over $x_j \in [0,1)$, so the t-dependence disappears, and one is left with

$$\begin{aligned}\int_{[0,1)^n} \Big| \sum_{k=1}^n \operatorname{sgn}(y_k) \prod_{j=1}^{k-1} \phi_{kj}(x_j)\psi_{kk}(x_k)\Big|^p \,\mathrm{d}x \\ \leq \alpha^p \int_{[0,1)^n} \Big| \sum_{k=1}^n \prod_{j=1}^{k-1} \phi_{kj}(x_j)\phi_{kk}(x_k)\Big|^p \,\mathrm{d}x.\end{aligned}$$

Since one is free to choose the signs $\operatorname{sgn}(y_k)$, it follows that

$$(*) \qquad \Big\| \sum_{k=1}^n \varepsilon_k \prod_{j=1}^{k-1} \phi_{kj}(x_j)\psi_{kk}(x_k)\Big\|_p \leq \alpha \Big\| \sum_{k=1}^n \prod_{j=1}^{k-1} \phi_{kj}(x_j)\phi_{kk}(x_k)\Big\|_p.$$

This is not yet the claim, but one can repeat the same argument with $\varepsilon_k \psi_{kk}$ in place of ϕ_{kk}. As one immeadiately checks, the conjugate function of ψ_{kk} is $-\phi_{kk}$. Hence there also holds

$$\Big\| \sum_{k=1}^n \varepsilon'_k \prod_{j=1}^{k-1} \phi_{kj}(x_j)(-\varepsilon_k)\phi_{kk}(x_k)\Big\|_p \leq \alpha \Big\| \sum_{k=1}^n \prod_{j=1}^{k-1} \phi_{kj}(x_j)\varepsilon_k\psi_{kk}(x_k)\Big\|_p.$$

Writing the two inequalities in a row (with ε_k as in the claim and $\varepsilon'_k = -1$), the assertion of Theorem 5.11 follows.

5.15. **Exercises.**

1. 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 $a \leq b \leq 2a$. Prove that $|H_{ab}f| \leq CMf$ 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.)
2. In the proof of Theorem 5.6, one needed an auxiliary function $\phi \not\equiv 0$ such that $|\phi(x)| \leq C(1+|x|)^{-1-\delta}$ and $\{\hat{\phi} \neq 0\} \subseteq [-R, R]$. Check that the following ϕ provides an example of such a function, even with the additional (and sometimes useful) property that $\hat{\phi} \equiv 1$ close to the origin:
$$\phi(x) = \frac{\sin(3\pi x)\sin(\pi x)}{(\pi x)^2}, \qquad \hat{\phi}(\xi) = \min\big\{1, \max\{0, 2 - |\xi|\}\big\}.$$
3. Prove *Plancherel's theorem* in the case of functions $f, \hat{f} \in L^1(\mathbb{R})$: $\|f\|_2 = \|\hat{f}\|_2$. (Hint: both sides are equal to a double integral involving both f and $\hat{f}$.)
4. Suppose that the functions ϕ_{kk}, $k = 1, \ldots, n$, in Theorem 5.11 have the special form $\phi_{kk}(u) = \sum_{m=1}^M a_{kk}(m)e^{i2\pi mu}$ (i.e., trigonometric polynomials with only positive frequencies). Prove that in this case the inequality of the mentioned theorem holds with the constant α instead of α^2 on the right.

5.16. **References.** The basic Fourier analysis presented in this chapter is standard material. The core result is Bourgain's theorem from [1].

Appendix A. Solutions to exercises

A.1. Conditional expectation.

1. E.g. $\Omega = \mathbb{R}$, $\mathscr{F}$ = Borel σ-algebra, $\mathscr{G} = \{\varnothing, \mathbb{R}\}$ and μ = Lebesgue measure. Another possibility is e.g. $\mathscr{G} = \{\varnothing, (-\infty, 0), [0, \infty), \mathbb{R}\}$, which shows that $\mathscr{G}$ need not be the trivial σ-algebra.
2. Denote $h_n := 2g - |f - f_n|$, so that $0 \le h_n \in L^1_\sigma(\mathscr{F}, \mu)$ ja $h_n \to 2g \in L^1_\sigma(\mathscr{F}, \mu)$. By the conditional Fatou's lemma
$$\mathbb{E}[2g|\mathscr{G}] \le \liminf_{n\to\infty} \mathbb{E}[h_n|\mathscr{G}] = \mathbb{E}[2g|\mathscr{G}] - \limsup_{n\to\infty} \mathbb{E}[|f - f_n|\,|\mathscr{G}],$$
which implies $\limsup_{n\to\infty} \mathbb{E}[|f - f_n|\,|\mathscr{G}] \le 0$. Finally $|\mathbb{E}[f_n|\mathscr{G}] - \mathbb{E}[f|\mathscr{G}]| \le \mathbb{E}[|f_n - f|\,|\mathscr{G}]$.
3. Let $H \in \mathscr{H}^0 \subseteq \mathscr{G}^0$. Using the definition of the conditional expectation three times, it follows that
$$\int_H \mathbb{E}\big(\mathbb{E}[f|\mathscr{G}]\big|\mathscr{H}\big)\,\mathrm{d}\mu = \int_H \mathbb{E}[f|\mathscr{G}]\,\mathrm{d}\mu = \int_H f\,\mathrm{d}\mu = \int_H \mathbb{E}[f|\mathscr{H}]\,\mathrm{d}\mu,$$
and Lemma 1.2 completes the argument.
4. The familiar inequality from the hint follows e.g. from Jensen's inequality $\phi(x/p + y/p') \le \phi(x)/p + \phi(y)/p'$ applied to the convex function $\phi(x) = e^x$ and the values $x := \log a^p$, $y =: \log b^{p'}$. Another possibility is to move all the terms on one side of the inequality and investigate the resulting expression, say, as a function of $a \in [0, \infty)$ for a fixed b. Checking the non-negativity is a high school level exercise in differentiation — the extremal values are reached at the endpoints of the interval or at the zeros of the derivative.

   From the hint it follows directly that
$$|\mathbb{E}[f \cdot g|\mathscr{G}]| \le \mathbb{E}[|f| \cdot |g|\,|\mathscr{G}] \le \frac{1}{p}\mathbb{E}[|f|^p|\mathscr{G}] + \frac{1}{p'}\mathbb{E}[|g|^{p'}|\mathscr{G}].$$
If one replaces f by the function $\lambda \cdot f$ and g by g/λ, where $\lambda > 0$ is a constant, the left side of the previous inequality stays invariant, but the right side becomes
$$\frac{\lambda^p}{p}\mathbb{E}[|f|^p|\mathscr{G}] + \frac{\lambda^{-p'}}{p'}\mathbb{E}[|g|^{p'}|\mathscr{G}].$$
By minimizing this expression with respect to λ at each $\omega \in \Omega$, the claimed upper bound for $\mathbb{E}[f \cdot g|\mathscr{G}]$ follows.

A.2. Discrete-time martingales and Doob's inequality.

1. We may assume that $I \ne \varnothing$. For all $j \in \mathbb{Z}$, let $N(j) := \{i \in I : i \ge j\}$. This is empty for some j if and only if I is bounded from above, i.e., there exists $\max I$. Set $n(j) := \min N(j)$ if $N(j) \ne \varnothing$, and $n(j) := \max I$ otherwise. With the help of the tower rule one easily checks that $\mathscr{F}_j := \mathscr{F}_{n(j)}$ is a filtration and $f_j := f_{n(j)}$ a martingale adapted to it. Clearly $n(i) = i$ if $i \in I$, so this is an extension of the original one.
2. For the filtration of the hint, there holds
$$\mathbb{E}[f|\mathscr{F}_{-n}](x) = \frac{1}{n\delta}\int_0^{n\delta} f(y)\,\mathrm{d}y, \qquad x \in (0, n\delta],$$
so the corresponding Doob's maximal function satisfies, for $x \in ((n-1)\delta, n\delta]$,
$$Mf(x) \ge \frac{1}{n\delta}\int_0^{n\delta} f(y)\,\mathrm{d}y \ge \frac{1}{x + \delta}\int_0^x f(y)\,\mathrm{d}y =: F_\delta(x).$$
By Doob's inequality $\|F_\delta\|_p \le p'\|f\|_p$, and the claim follows from monotone convergence as $\delta \searrow 0$.
3. Let $f(x) := 1_{(0,1]}(x) \cdot x^\alpha$. This is in $L^p(\mathbb{R}_+)$ if and only if $\alpha > -1/p$. Now
$$F(x) := \frac{1}{x}\int_0^x f(y)\,\mathrm{d}y = \frac{x^\alpha}{1+\alpha} = (1+\alpha)^{-1}f(x), \qquad x \in (0, 1].$$

Hence, if Hardy's inequality holds with some constant C, then $C \geq \|F\|_p/\|f\|_p \geq (1+\alpha)^{-1}$. In the limit $\alpha \searrow -1/p$, there follows that $C \geq (1-1/p)^{-1} = p'$.

4. One has to show that $\sigma(\mathscr{D}_k^\beta) \subseteq \sigma(\mathscr{D}_{k+1}^\beta)$. It suffices to prove that every $J \in \mathscr{D}_k^\beta$ is a (necessarily countable) union of some sets in $\mathscr{D}_{k+1}^\beta$. By definition, $J = I + \sum_{j>k} 2^{-j}\beta_j$ for some $I = 2^{-k}[\ell, \ell+1) \in \mathscr{D}_k$. Clearly $I = I_0 \cup I_1$, where
$$I_i := 2^{-k}[\ell + i/2, \ell + (i+1)/2) = 2^{-(k+1)}[2\ell + i, 2\ell + i + 1) \in \mathscr{D}_{k+1}.$$
Now $J = J_0 \cup J_1$, if we set
$$\begin{aligned} J_i := I_i + \sum_{j>k} 2^{-j}\beta_j &= (I_i + 2^{-(k+1)}\beta_{k+1}) + \sum_{j>k+1} 2^{-j}\beta_j \\ &\in \mathscr{D}_{k+1} + \sum_{j>k+1} 2^{-j}\beta_j = \mathscr{D}_{k+1}^\beta. \end{aligned}$$

5. Consider the endpoints of the intervals $I \in \mathscr{D}_k^0 \cup \mathscr{D}_k^\beta$. They have the form $2^{-k}\ell$ or $2^{-k}\ell + \sum_{j>k} 2^{-j}\beta_j = 2^{-k}\big(\ell + \sum_{j>k} 2^{k-j}\beta_j\big)$, where $\ell \in \mathbb{Z}$. Depending on the parity of k, there holds one of
$$\sum_{j>k} 2^{k-j}\beta_j = \begin{cases} \sum_{j=1,3,5,\ldots} 2^{-j} = 2^{-1}\sum_{j=0}^\infty 4^{-j} = 2/3, \\ \sum_{j=2,4,5,\ldots} 2^{-j} = 4^{-1}\sum_{j=0}^\infty 4^{-j} = 1/3. \end{cases}$$
Thus the endpoints have the form $2^{-k}\ell$ and either $2^{-k}(\ell+1/3)$ or $2^{-k}(\ell+2/3)$; in either case, the minimal distance of two endpoints is $2^{-k}/3$.

Let then J be some finite subinterval of $\mathbb{R}$. Choose the unique $k \in \mathbb{Z}$ with $3|J| < 2^{-k} \leq 6|J|$. Since $|J| < 2^{-k}/3$, the interval J can contain at most one endpoint of one interval $I' \in \mathscr{D}_k^0 \cup \mathscr{D}_k^\beta$. If $I' \in \mathscr{D}_k^0$, then J does not contain any endpoints of intervals in $\mathscr{D}_k^\beta$. On the other hand, $\mathscr{D}_k^\beta$ covers all of $\mathbb{R}$, so in particular all of J. Since J is a connected interval and does not contain endpoints of $\mathscr{D}_k^\beta$, it must be completely contained in a single interval $I \in \mathscr{D}_k^\beta$. Symmetrically, if $I' \in \mathscr{D}_k^\beta$, then there exists $I \in \mathscr{D}_k^0$, which contains J.

In any case $J \subset I \in \mathscr{D}_k^0 \cup \mathscr{D}_k^\beta$ and $|I| = 2^{-k} \leq 6|J|$.

6. Let M^0 and M^β be Doob's maximal operators related to the filtrations $(\sigma(\mathscr{D}_k^0))_{k\in\mathbb{Z}}$ and $(\sigma(\mathscr{D}_k^\beta))_{k\in\mathbb{Z}}$. Let $x \in \mathbb{R}$ and $J \ni x$ be a finite subinterval of $\mathbb{R}$. By the previous exercise, there exists $I \in \mathscr{D}^0 \cup \mathscr{D}^\beta$, such that $I \supset J$ and $|I| \leq 6|J|$. Hence
$$\frac{1}{|J|}\int_J |f(y)|\,\mathrm{d}y \leq \frac{6}{|I|}\int_I |f(y)|\,\mathrm{d}y \leq \begin{cases} 6M^0|f|(x), & I \in \mathscr{D}^0, \\ 6M^\beta|f|(x), & I \in \mathscr{D}^\beta. \end{cases}$$
Taking the supremum over all $J \ni x$ on the left, we obtain
$$M_{HL}f(x) \leq 6M^0|f|(x) + 6M^\beta|f|(x)$$
and then by Doob's inequality
$$\|M_{HL}f\|_p \leq 6\|M^0|f|\,\|_p + 6\|M^\beta|f|\,\|_p \leq 12p'\|f\|_p.$$

A.3. **Burkholder's inequality.**

1. Since $d_k = \mathbb{E}[f_n|\mathscr{F}_k] - \mathbb{E}[f_n|\mathscr{F}_{k-1}]$ for $1 \leq k \geq n$ and $d_0 = \mathbb{E}[f_n|\mathscr{F}_0]$, it follows that
$$g_n = \varepsilon_0\mathbb{E}[f_n|\mathscr{F}_0] + \sum_{k=1}^n \varepsilon_k\big(\mathbb{E}[f_n|\mathscr{F}_k] - \mathbb{E}[f_n|\mathscr{F}_{k-1}]\big) =: T_\varepsilon f_n.$$

The conditional expectation operators are selfadjoint. In fact, if $f \in L^p(\mathscr{F},\mu)$ and $h \in L^{p'}(\mathscr{F},\mu)$ and $\mathscr{G} \subseteq \mathscr{F}$ is a sub-σ-algebra with $G_j \in \mathscr{G}^0$ such that $G_j \nearrow \Omega$, then
$$\begin{aligned} &\int_\Omega \mathbb{E}[f|\mathscr{G}]\cdot h\,\mathrm{d}\mu = \lim_{j\to\infty}\int_{G_j} \mathbb{E}[f|\mathscr{G}]\cdot h\,\mathrm{d}\mu = \lim_{j\to\infty}\int_{G_j} \mathbb{E}\{\mathbb{E}[f|\mathscr{G}]\cdot h|\mathscr{G}\}\,\mathrm{d}\mu \\ &= \lim_{j\to\infty}\int_{G_j} \mathbb{E}[f|\mathscr{G}]\cdot\mathbb{E}\{h|\mathscr{G}\}\,\mathrm{d}\mu = \lim_{j\to\infty}\int_{G_j} \mathbb{E}[f\cdot\mathbb{E}\{h|\mathscr{G}\}|\mathscr{G}]\,\mathrm{d}\mu = \int_\Omega f\cdot\mathbb{E}\{h|\mathscr{G}\}\,\mathrm{d}\mu, \end{aligned}$$

where Theorem 1.15 ("pulling out" a $\mathscr{G}$-measurable function) was used twice.

By the linearity of taking the adjoint operator, it follows that $T_\varepsilon^* = T_\varepsilon$. Hence

$$\begin{aligned}\|T_\varepsilon f\|_{p'} &= \sup\Big\{\int_\Omega T_\varepsilon f\cdot h\,\mathrm{d}\mu : \|h\|_p \le 1\Big\} = \sup\Big\{\int_\Omega f\cdot T_\varepsilon h\,\mathrm{d}\mu : \|h\|_p \le 1\Big\}\\ &\le \|f\|_{p'}\sup\Big\{\|T_\varepsilon h\|_p : \|h\|_p \le 1\Big\} \le \|f\|_{p'}\cdot\beta_p.\end{aligned}$$

Taking the supremum over all appropriate filtrations, functions f and sign sequences ε, it follows that $\beta_{p'} \le \beta_p$. By symmetry, also $\beta_p \le \beta_{p'}$.

2. By the construction of the conditional expectation of L^2 functions, there holds $d_k = f_k - f_{k-1} \perp L^2(\mathscr{F}_{k-1},\mu) \supset \{d_0,\ldots,d_{k-1}\}$ for all $1 \le k \le n$. Hence the sequence $(d_k)_{k=0}^n$, and then also $(\varepsilon_k d_k)_{k=0}^n$, is orthogonal in $L^2(\mathscr{F},\mu)$. Thus by Pythagoras' Theorem,

$$\Big\|\sum_{k=0}^n \varepsilon_k d_k\Big\|_2^2 = \sum_{k=0}^n \|\varepsilon_k d_k\|_2^2 = \sum_{k=0}^n \|d_k\|_2^2 = \Big\|\sum_{k=0}^n d_k\Big\|_2^2.$$

3. A $\mathscr{G}$-measurable function on $[0,1]^2$ is one which is constant with respect to the second variable, but otherwise arbitrary. So $\mathbb{E}[1_A|\mathscr{G}](x,y) = g(x)$ for some Borel function g on $[0,1]$. This function must satisfy

$$\int_G g(x)\,\mathrm{d}x = \int_{G\times[0,1]} \mathbb{E}[1_A|\mathscr{G}]\,\mathrm{d}x\,\mathrm{d}y = \int_G\int_0^1 1_A(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_G \phi(x)\,\mathrm{d}x$$

for all Borel sets $G \subseteq [0,1]$. This holds if and only if $g(x) = \phi(x)$ a.e.

Since ϕ was arbitrary, it is seen that $\mathbb{E}[1_A|\mathscr{G}]$ need not be simple.

4. By the assumption that β is a constant for Doob's inequality, $\int F(f_\infty, \max\{f^*, 0\})\,\mathrm{d}t = \int (f^*)^p\,\mathrm{d}t - \beta^p\int|f_\infty|^p\,\mathrm{d}t \le 0$. Since $U(0,0)$ is the supremum over all such quantities with $f_0 = 0$, also $U(0,0) \le 0$.

Martingales g starting from αx and martingales f starting from x are in the obvious one-to-one correspondence via $g = \alpha f$. Then $F(g_\infty, \max\{g^*, \alpha y\}) = F(\alpha f_\infty, \alpha\max\{f^*, y\}) = \alpha^p F(f_\infty, \max\{f^*, y\})$. Integrating over $[0,1)$ and taking the supremum over all f, equivalently all g, it follows that $U(\alpha x, \alpha y) = \alpha^p U(x,y)$.

If $f_0 \equiv x$, then $f^* \ge |f_0| = x$, and hence $\max\{f^*, \max\{|x|, y\}\} = \max\{f^*, |x|, y\} = \max\{f^*, y\}$. Substituting into $\int_0^1 F(f_\infty, \cdot)\,\mathrm{d}t$ and taking the appropriate supremum gives $U(x, \max\{|x|, y\}) = U(x,y)$.

Finally, the concavity. Let $y \in [0,\infty)$, $x_1, x_2 \in \mathbb{R}$, $\alpha \in (0,1)$, $x = \alpha x_1 + (1-\alpha)x_2$, and $m_i < U(x_i, y)$ for $i = 1,2$. Then one can find simple martingales f^i with $f_0^i \equiv x_i$ so that $\int_0^1 F(f_\infty^i, \max\{f^{i*}, y\})\,\mathrm{d}t > m_i$. By augmenting zero steps if necessary, we may assume that the $f^i = (f_k^i)_{k=0}^n$ with the same n. A new martingale f is defined by

$$f_0(t) := x, \qquad f_{k+1}(t) := \begin{cases} f_k(t/\alpha), & t \in [0,\alpha),\\ f_k\big((t-\alpha)/(1-\alpha)\big), & t \in [\alpha, 1).\end{cases}$$

Then $f^*(t) = \max\{f^{1*}(t/\alpha), |x|\}$ for $t \in [0,\alpha)$ and $f^*(t) = \max\{f^{2*}\big((t-\alpha)/(1-\alpha)\big), |x|\}$ for $t \in [\alpha,1)$. Substituting and changing variables,

$$\begin{aligned}U(x,y) &> \int_0^1 F(f_\infty, \max\{f^*, y\})\,\mathrm{d}t\\ &= \alpha\int_0^1 F(f_\infty^1, \max\{f^{1*}, |x|, y\})\,\mathrm{d}t + (1-\alpha)\int_0^1 F(f_\infty^2, \max\{f^{2*}, |x|, y\})\,\mathrm{d}t\end{aligned}$$

Since F is increasing in its second argument, there holds

$$F(f_\infty^i, \max\{f^{i*}, |x|, y\}) \ge F(f_\infty^i, \max\{f^{i*}, y\}),$$

and hence $U(x,y) > \alpha m_1 + (1-\alpha)m_2$. With $m_i \nearrow U(x_i, y)$, this gives concavity.

5. The claim is equivalent to $\log \alpha_p \geq 0$, where $\log \alpha_p = \log p + (p-1)\log(1-1/p) =: g(p)$. Now $g'(p) = 1/p + \log(1-1/p) + (p-1)(1-1/p)^{-1} \cdot 1/p^2 = 2/p + \log(1-1/p) =: h(1/p)$ and $h'(t) = \big(2t + \log(1-t)\big)' = 2 - (1-t)^{-1} = (1-2t)/(1-t) \geq 0$ for $t \in [0,1/2]$. Hence $h(t) \geq h(0) = 0$ for $t \in [0,1/2]$, i.e., $g'(p) = h(1/p) \geq 0$ for $p \in [2,\infty)$, and then $g(p) \geq g(2) = \log 2 + \log(1/2) = 0$ for $p \in [2,\infty)$.
6. One still gets $f^{(k)}(x) = 0$ iff $|x+1| = \beta^{p/(p-k)}|x-1| =: \beta_k|x-1|$ and $|x| > 1$ when $k = 1$. Again $1 \leq \beta_0 \leq \beta_1$, but now $\beta_2 \in (0,1]$ since $p/(p-2) < 0$. Hence, as before, f has the zeros x_0 and $1/x_0$ while f' has the unique zero $1/x_1$, where $x_k = (\beta_k - 1)/(\beta_k + 1)$, but
$$f''(x) = 0 \Leftrightarrow \ |x-1| = \beta^{p/(2-p)}|x+1| =: \tilde{\beta}|x+1|$$
$$\Leftrightarrow x \in \{x_2, 1/x_2\},\ x_2 = -\frac{\tilde{\beta}-1}{\tilde{\beta}+1} \in (-1,0].$$
So the points of interest are $1/x_2 < x_2 \leq 0 \leq x_0 < 1/x_1 \leq 1/x_0$, and f is positive on $(x_0, 1/x_0)$, increasing on $(-\infty, 1/x_1)$ and convex on $(1/x_2, x_2)$, with opposite properties in the interior of the complement.
7. There holds $0 = f(x_0) \leq w(x_0)$ and $1 = f(1) \leq w(1)$. By concavity, $w(-1) + w(1) \leq 2w(0)$, and $w(0) \leq 0$ by the observation made in 3.13; hence $w(-1) \leq -w(1) \leq -1$. For $x \in (-1,0)$,
$$\frac{w(x) - w(-1)}{x-(-1)} = \frac{|x|^p w(1/x) - w(-1)}{x-(-1)} = \frac{|x|^p}{-x}\frac{w(1/x) - w(-1)}{(-1) - 1/x} + \frac{|x|^p - 1}{-(|x|-1)}w(-1),$$
so letting $x \searrow -1$, one gets $D_+w(-1) = -D_-w(-1) - pw(-1)$. By concavity, for all $x \in (-1, x_0)$,
$$(*) \qquad \begin{aligned} \frac{-w(-1)}{x_0 - (-1)} &\leq \frac{w(x_0) - w(-1)}{x_0 - 1} \leq \frac{w(x) - w(-1)}{x - (-1)} \\ &\leq D_+w(-1) \leq \frac{1}{2}(D_+ + D_-)w(-1) = -\frac{p}{2}w(-1). \end{aligned}$$
Dividing by $-w(-1) > 0$, it follows that $p/2 \geq 1/(x_0+1) = (\beta+1)/2\beta \ \Leftrightarrow\ \beta \geq 1/(p-1)$.
8. If $\beta = 1/(p-1)$, then equality holds at every step in $(*)$ above. Thus for all $x \in (-1, x_0)$ (and by continuity at the endpoints),
$$w(x) = w(-1) + w'(-1)(x+1) = -w(-1)\big(p/2 \cdot x + (p/2 - 1)\big).$$
Let $\tilde{\alpha}_p := -w(-1)$. Since $w \geq f$ and $w(x_0) = 0 = f(x_0)$, by Lemma 3.18 there holds
$$w'(x_0) = \tilde{\alpha}_p\frac{p}{2} = f'(x_0) = \frac{p}{2}\Big(\big(\frac{1}{p}\big)^{p-1} + \big(\frac{1}{p-1}\big)^p\big(\frac{p-1}{p}\big)^{p-1}\Big) = \frac{p}{2}\frac{p}{p-1}\big(\frac{1}{p}\big)^{p-1},$$
and hence $\tilde{\alpha}_p = p' \cdot p^{1-p}$.
9. To prove that $w \geq f$ on $[-1, x_0]$, note that $f'' > 0$ on $(-1, x_2)$ and $f'' < 0$ on x_2, x_0; thus f' is increasing on the first and decreasing on the latter interval, and hence
$$f' \geq \min\{f'(-1), f'(x_0)\} = \min\{\frac{p}{2}\beta^p, w'(x_0)\} = \frac{p}{2}\min\{\beta^p, \tilde{\alpha}_p\} = \frac{p}{2}\alpha_p \equiv w',$$
provided that $\beta^p \geq \tilde{\alpha}_p$. Assuming this, it follows that $(f-w)' \geq 0$, hence $f - w \leq f(x_0) - w(x_0) = 0$.

   To finish, one must show $\big(\frac{1}{p-1}\big)^p \geq \frac{p}{p-1}\big(\frac{1}{p}\big)^{p-1}$, equivalently $p^{p-2} \geq (p-1)^{p-1}$, or
$$(p-2)\log p - (p-1)\log(p-1) \geq 0, \qquad p \in (1,2].$$
Proof. Write $g(p)$ for the left side. Then $g(1) = 0 = g(2)$,
$$g'(p) = \log p - \log(p-1) - \frac{2}{p}, \quad g''(p) = \frac{p-2}{p^2(p-1)} \leq 0.$$
Thus g is concave and hence attains its minimum at the endpoints. □

10. The condition $w(x) = |x|^p w(1/x)$ determines w on the intervals $(-\infty, -1]$ (by its values on $[-1, 0)$) and on $[1/x_0, \infty)$ (by the values on $(0, x_0]$). The expression is

$$w(x) = \tilde{\alpha}_p \times \begin{cases} \Big(-\frac{p}{2}|x|^{p-1} - \big(1 - \frac{p}{2}\big)|x|^p \Big), & x \in (-\infty, -1], \\ \Big(\frac{p}{2}x^{p-1} - \big(1 - \frac{p}{2}\big)x^p \Big), & x \in \big[\frac{1}{x_0}, \infty\big) = \big[\frac{p}{2-p}, \infty\big). \end{cases}$$

The concavity of w on these two intervals is verifies from the condition $w'' \leq 0$ (where one uses $p \in (1, 2)$). One also checks that $D_w(-1) = \tilde{\alpha}_p p/2 = D_+ w(-1)$, whence w is concave on $(-\infty, x_0]$. The domination $w \geq f$ on $(-\infty, -1] \cup [1/x_0, \infty)$ follows from the corresponding property on $[-1, x_0]$ and the fact that both w and f satisfy the property $w(x) = |x|^p w(1/x)$.

The function f itself is concave on $[x_0, 1/x_0]$. It was checked that $f(x_0) = w(x_0)$ and $f'(x_0) = w'(x_0)$ which, as in the case $p > 2$, imply the corresponding equalities at $1/x_0$. Hence the function

$$\tilde{w}(x) := \begin{cases} w(x), & x \notin [x_0, 1/x_0] = \big[\frac{2-p}{p}, \frac{p}{2-p}\big], \\ f(x), & x \in [x_0, 1/x_0], \end{cases}$$

satisfies the required properties, and is the smallest possible such function.

Remark. There are other solutions. One can check that on $[x_0, 1/x_0]^{\complement}$, $w(x)$ is given by

$$\begin{aligned} w(x) &= \tilde{\alpha}_p \Big\{ (p-1)\Big|\frac{x+1}{2}\Big| - \Big|\frac{x-1}{2}\Big| \Big\} \max\{1, |x|\}^{p-1} \\ &= \alpha_p \Big\{ \Big|\frac{x+1}{2}\Big| - \frac{1}{p-1}\Big|\frac{x-1}{2}\Big| \Big\} \Big(\Big|\frac{x+1}{2}\Big| + \Big|\frac{x-1}{2}\Big| \Big)^{p-1}, \end{aligned}$$

where $\alpha_p := (p-1)\tilde{\alpha}_p = p(1/p)^{p-1} = p(1 - 1/p^*)^{p-1}$ for $p \in (1, 2)$, where the final expression agrees with the definition of α_p for $p \in (2, \infty)$. Taking the above formula as the definition of $w(x)$ for all $x \in \mathbb{R}$, one can check that this is also a concave function with $w \geq f$ and $w(x) = |x|^p w(1/x)$.

A.4. **Petermichl's dyadic shift and the Hilbert transform.**

1. Write the binary expansion $x = \operatorname{sgn}(x) \sum_{j \in \mathbb{Z}} \alpha_j 2^{-j}$, where $\alpha_j = \alpha_j(x) \in \{0, 1\}$ for all $j \in \mathbb{Z}$ and $\alpha_j = 0$ for $2^{-j} > |x|$. Then

$$\mathscr{D}_j + x = \mathscr{D}_j + \operatorname{sgn}(x) \sum_{i > j} \alpha_i 2^{-i} = \begin{cases} \mathscr{D}_j + \sum_{i>j} \alpha_i 2^{-i}, & x > 0 \\ \mathscr{D}_j + \sum_{i>j} (1 - \alpha_i) 2^{-i}, & x < 0. \end{cases}$$

For the case $x < 0$ one observed that $\sum_{i>j} 2^{-i} = 2^{-j}$ and $\mathscr{D}_j + 2^{-j} = \mathscr{D}_j$.

Thus it follows that $\mathscr{D} + x = \mathscr{D}^{\beta(x)}$, where $\beta_j(x) = \alpha_j(x)$ if $x > 0$ and $\beta_j(x) = 1 - \alpha_j(x)$ if $x < 0$. In both cases, the sequence $(\beta_j(x))_{j \leq j_0(x)}$ is a constant (in the first case, zero; in the second, one) for some $j_0(x)$. Hence any $\mathscr{D}^\beta$, for which $\lim_{j \to -\infty} \beta_j$ does not exist, cannot be of the form $\mathscr{D} + x$ for any $x \in \mathbb{R}$.

2. Write (in this exercise) $\mathscr{D}_j^1$ for the one-dimensional dyadic intervals of lenght 2^{-j}. Then $\mathscr{D}_j = \mathscr{D}_j^1 \times \mathscr{D}_j^1$, and one can define (e.g.) $\mathscr{D}_{j+1/2} := \mathscr{D}_{j+1}^1 \times \mathscr{D}_j^1$. Then, denoting $\mathscr{F}_j = \sigma(\mathscr{D}_j)$,

$$f = \sum_{j=-\infty}^{\infty} \sum_{Q \in \mathscr{D}_j} 1_Q \big(\mathbb{E}[f|\mathscr{F}_{j+1}] - \mathbb{E}[f|\mathscr{F}_{j+1/2}] + \mathbb{E}[f|\mathscr{F}_{j+1/2}] - \mathbb{E}[f|\mathscr{F}_j] \big)$$

For $Q = I \times J \in \mathscr{D}_j$, write $Q_s := I_s \times J$ and $Q_{st} := I_s \times J_t$, $s, t \in \{-, +\}$. Note that $|Q| = 2|Q_s| = 4|Q_{st}|$. Then

$$\begin{aligned}
&1_Q\big(\mathbb{E}[f|\mathscr{F}_{j+1}] - \mathbb{E}[f|\mathscr{F}_{j+1/2}] + \mathbb{E}[f|\mathscr{F}_{j+1/2}] - \mathbb{E}[f|\mathscr{F}_j]\big) \\
&= \sum_{s\in\{+,-\}} \Big(\sum_{t\in\{+,-\}} \frac{1_{Q_{st}}}{|Q_{st}|} \int_{Q_{st}} f - \frac{1_{Q_s}}{|Q_s|} \int_{Q_s} f \Big) + \sum_{s\in\{+,-\}} \frac{1_{Q_s}}{|Q_s|} \int_{Q_s} f - \frac{1_Q}{|Q|} \int_Q f \\
&= \sum_{s\in\{+,-\}} \sum_{t\in\{+,-\}} \frac{1_{Q_{st}}}{|Q_s|} \Big(\int_{Q_{st}} f - \int_{Q_{s,-t}} f \Big) + \sum_{s\in\{+,-\}} \frac{1_{Q_s}}{|Q_s|} \Big(\int_{Q_s} f - \int_{Q_{-s}} f \Big) \\
&= \sum_{s\in\{+,-\}} \frac{1_{Q_{s-}} - 1_{Q_{s+}}}{|Q_s|} \Big(\int_{Q_{s-}} f - \int_{Q_{s+}} f \Big) + \frac{1_{Q_-} - 1_{Q_+}}{|Q|} \Big(\int_{Q_-} f - \int_{Q_+} f \Big) \\
&= \sum_{s\in\{+,-\}} h_{Q_s} \int h_{Q_s} f + h_Q \int h_Q f,
\end{aligned}$$

where

$$\begin{aligned}
h_Q(x,y) &:= \frac{1}{|Q|^{1/2}} \big(1_{Q_-} - 1_{Q_+}\big)(x,y) = h_I(x) \frac{1_J(y)}{|J|^{1/2}} \\
h_{Q_\pm}(x,y) &:= \frac{1}{|Q_\pm|^{1/2}} \big(1_{Q_{\pm-}} - 1_{Q_{\pm+}}\big)(x,y) = \frac{1_{I_\pm}(x)}{|I_\pm|^{1/2}} h_J(y).
\end{aligned}$$

3. As in 3.23, one can represent any $t_k \in [-1, 1]$ in the form $t_k = \sum_{j=1}^\infty 2^{-j} \varepsilon_{kj}$ with $\varepsilon_{kj} \in \{-1, +1\}$. Let M be the maximum of $F(t)$ for $t \in \{-1, 1\}^n$. Then for any $t \in [-1, 1]^n$,

$$|F(t)| = \Big| \sum_{k=1}^n t_k \xi_k \Big| \leq \sum_{j=1}^\infty 2^{-j} \Big| \sum_{k=1}^n \varepsilon_{kj} \xi_k \Big| \leq \sum_{j=1}^\infty 2^{-j} M = M,$$

so M is the maximum in the whole cube $[-1, 1]^n$.

4. Without loss of generality, $\max_{1\leq k\leq n} \|g_k\|_\infty = 1$. Then for each fixed $x \in \mathbb{R}$, $(g_k(x))_{k=1}^n \in [-1, 1]^n$, and hence (using the previous exercise with $\xi_k = \varepsilon_k f_k(x)$ and $t_k = g_k(x)$)

$$\Big| \sum_{k=1}^n \varepsilon_k g_k(x) f_k(x) \Big| \leq \Big| \sum_{k=1}^n \varepsilon_k \eta_k f_k(x) \Big|$$

for some $\eta_k \in \{-1, 1\}^n$. Taking the pth power and expectations of both sides, it follows that

$$\mathbb{E}\Big| \sum_{k=1}^n \varepsilon_k g_k(x) f_k(x) \Big|^p \leq \mathbb{E}\Big| \sum_{k=1}^n \varepsilon_k \eta_k f_k(x) \Big|^p = \mathbb{E}\Big| \sum_{k=1}^n \varepsilon_k f_k(x) \Big|^p$$

since $(\varepsilon_k \eta_k)_{k=1}^n$ and $(\varepsilon_k)_{k=1}^n$ have the same distribution. It remains to integrate over $\mathbb{R}$.

A.5. **Back to Burkholder's inequality.**

1. Since $a \leq b \leq 2a$,

$$\Big| \int_{a<|y|<b} \frac{1}{\pi y} f(x-y)\,\mathrm{d}y \Big| \leq \frac{1}{\pi a} \int_{|y|<b} |f(x-y)|\,\mathrm{d}y \leq \frac{2}{\pi b} \int_{x-b}^{x+b} |f(y)|\,\mathrm{d}y \leq \frac{4}{\pi} Mf(x).$$

If $g \in C_c(\mathbb{R})$, then similarly

$$|H_{ab} g(x)| = \Big| \int_{a<|y|<b} \frac{1}{\pi y} [g(x-y) - g(x)]\,\mathrm{d}y \Big| \leq \frac{4}{\pi} \max_{|y-x|<b} |g(y) - g(x)|,$$

which tends to zero as $b \to 0$.

For a general $f \in L^p(\mathbb{R})$, let $g \in C_c(\mathbb{R})$ with $\|f - g\|_p < \delta$. Then

$$\limsup_{\varepsilon\to 0} |H_{ab} f| \leq \limsup_{\varepsilon\to 0} \big(|H_{ab}(f-g)| + |H_{ab} g| \big) \leq \frac{4}{\pi} M(f-g).$$

Hence for any $\varepsilon > 0$,

$$|\{\limsup_{\varepsilon\to 0} |H_{ab}f| > \varepsilon\}| \le |\{\frac{4}{\pi}M(f-g) > \varepsilon\}| \le C\varepsilon^{-p}\|f-g\|_p^p \le C(\delta/\varepsilon)^p.$$

Since δ and ε were arbitrary, this proves that $\limsup_{\varepsilon\to 0}|H_{ab}f| = 0$ a.e.

2. Clearly $\{\hat\phi \neq 0\} \subseteq [-2,2]$ and $\hat\phi = 1$ on $[-1,1]$. Also $\phi(x)$ decays like $|x|^{-2}$ as $|x|\to\infty$, and it is bounded on compact intervals as a continuous function (recall that $\lim_{x\to 0}\sin x/x = 1$). By Corollary 5.4, it suffices to check that

$$\int_{-\infty}^{\infty} \hat\phi(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi = \phi(x).$$

So one computes

$$\begin{aligned}
\int_{-\infty}^{\infty} \hat\phi(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi &= 2\,\mathrm{Re}\Big(\int_0^1 e^{i2\pi x\xi}\,\mathrm{d}\xi + \int_1^2 (2-\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi\Big)\\
&= 2\,\mathrm{Re}\Big(\Big|_0^1 \frac{e^{i2\pi x\xi}}{i2\pi x\xi} + \Big|_1^2 (2-\xi)\frac{e^{i2\pi x\xi}}{i2\pi x} - \Big|_1^2 (-1)\frac{e^{i2\pi x\xi}}{(i2\pi x)^2}\Big)\\
&= 2\,\mathrm{Re}\Big(\frac{e^{i2\pi x}-1}{i2\pi x} - \frac{e^{i2\pi x}}{i2\pi x} + \frac{e^{i4\pi x}-e^{i2\pi x}}{(i2\pi x)^2}\Big)\\
&= 2\,\mathrm{Re}\,\frac{e^{i3\pi x}(e^{i\pi x}-e^{-i\pi x}}{-(2\pi x)^2} = 2\,\mathrm{Re}\,\frac{e^{i3\pi x}\cdot 2i\sin(\pi x)}{-(2\pi x)^2} = \frac{\sin(3\pi x)\sin(\pi x)}{(\pi x)^2}.
\end{aligned}$$

3. By the inversion formula, Fubini's theorem, and the definition of the Fourier transform,

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\overline{f(x)}\,\mathrm{d}x &= \int_{\mathbb{R}}\int_{\mathbb{R}} \hat f(\xi)e^{i2\pi x\xi}\,\mathrm{d}\xi\overline{f(x)}\,\mathrm{d}x\\
&= \int_{\mathbb{R}} \hat f(\xi)\overline{\int_{\mathbb{R}} f(x)e^{-i2\pi x\xi}\,\mathrm{d}x}\,\mathrm{d}\xi = \int_{\mathbb{R}} \hat f(\xi)\overline{\hat f(\xi)}\,\mathrm{d}\xi,
\end{aligned}$$

where the left side is $\|f\|_2^2$ while the right one is $\|\hat f\|_2^2$.

4. One repeats the same proof as in the general case until formula $(*)$ of Section 5.14. There one observes that, thanks to the special form of the functions ϕ_{kk}, the conjugate functions are $\psi_{kk} = -i\phi_{kk}$. Substituting this into the mentioned inequality, one has the asserted result.

Appendix B. English–Finnish-vocabulary

adapted – mukautettu
conditional – ehdollinen
dilation – venytys
dyadic – dyadinen
expectation – odotusarvo
filtration – suodatus
martingale – martingaali
martingale difference – martingaalierotus
maximal function – maksimifunktio
shift – siirto
stopping time – pysäytysaika
transform – muunnos
translation – siirto
zigzag martingale – siksakmartingaali

References

[1] J. Bourgain. Some remarks on Banach spaces in which martingale difference sequences are unconditional. *Ark. Mat.*, 21(2):163–168, 1983.

[2] D. L. Burkholder. Martingale transforms. *Ann. Math. Statist.*, 37:1494–1504, 1966.

[3] D. L. Burkholder. Boundary value problems and sharp inequalities for martingale transforms. *Ann. Probab.*, 12(3):647–702, 1984.

[4] Donald 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.

[5] J. L. Doob. *Stochastic processes.* John Wiley & Sons Inc., New York, 1953.

[6] G. H. Hardy and J. E. Littlewood. A maximal theorem with function-theoretic applications. *Acta Math.*, 54(1):81–116, 1930.

[7] Tao Mei. BMO is the intersection of two translates of dyadic BMO. *C. R. Math. Acad. Sci. Paris*, 336(12):1003–1006, 2003.
[8] 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.
[9] 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.
[10] Marcel Riesz. Sur les fonctions conjuguées. *Math. Z.*, 27(1):218–244, 1928.
[11] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.