Chapter VI
Some Extensions and Applications of Martingale Theory

Many of the results obtained in § 5.2 admit easy extensions to both infinite measures and Banach space valued random variables. Furthermore, in many applications, these extensions play a useful, and occasionally essential, role. In the first section of this chapter, I will develop some of these extensions, and in the second section I will show how these extensions can be used to derive Birkhoff's Individual Ergodic Theorem. The final section is devoted Burkholder's Inequality for martingales, an estimate which is second in importance only to Doob's Inequality.

§ 6.1 Some Extensions

Throughout the discussion which follows, $(\Omega, \mathcal{F}, \mu)$ will be a measure space and $\{\mathcal{F}_n : n \in \mathbb{N}\}$ will be a non-decreasing sequence of sub-σ-algebras with the property that $\mu \restriction \mathcal{F}_0$ is σ-finite. In particular, this means that the conditional expectation of a locally μ-integrable random variable given $\mathcal{F}_n$ is well-defined (cf. Theorem 5.1.12) even if the random variable takes values in a separable Banach space E. Thus, I will say that the sequence $\{X_n; n \in \mathbb{N}\}$ of E-valued random variables is a **μ-martingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$, or, more briefly, that the triple $(X_n, \mathcal{F}_n, \mu)$ is a **martingale**, if $\{X_n : n \in \mathbb{N}\}$ is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, each X_n is locally μ-integrable, and

$$X_{n-1} = \mathbb{E}^{\mu}\left[X_n \middle| \mathcal{F}_{n-1}\right] \quad \text{(a.e., } \mu) \quad \text{for each } n \in \mathbb{Z}^+.$$

Furthermore, when $E = \mathbb{R}$, I will say that $\{X_n : n \in \mathbb{N}\}$ is a **μ-submartingale with respect to** $\{\mathcal{F}_n : n \in \mathbb{N}\}$ (equivalently, the triple $(X_n, \mathcal{F}_n, \mu)$ is a **submartingale**) if $\{X_n : n \in \mathbb{N}\}$ is $\{\mathcal{F}_n : n \in \mathbb{N}\}$-progressively measurable, each X_n is locally μ-integrable, and

$$X_{n-1} \le \mathbb{E}^{\mu}\left[X_n \middle| \mathcal{F}_{n-1}\right] \quad \text{(a.e., } \mu) \quad \text{for each } n \in \mathbb{Z}^+.$$

§ 6.1.1. Martingale Theory for a σ-Finite Measure Space. Without any real effort, I can now prove the following variants of each of the basic results in § 5.2.

THEOREM 6.1.1. *Let $(X_n, \mathcal{F}_n, \mu)$ be an $\mathbb{R}$-valued μ-submartingale. Then, for each $N \in \mathbb{N}$ and $A \in \mathcal{F}_0$ on which X_N is μ-integrable:*

$$\mu\left(\left\{\max_{0\le n\le N} X_n \ge \alpha\right\} \cap A\right) \le \frac{1}{\alpha}\,\mathbb{E}^\mu\left[X_N,\, \left\{\max_{0\le n\le N} X_n \ge \alpha\right\} \cap A\right] \tag{6.1.2}$$

for all $\alpha \in (0,\infty)$; and so, when all the X_n's are non-negative, for every $p \in (1,\infty)$ and $A \in \mathcal{F}_0$:

$$\mathbb{E}^\mu\left[\sup_{n\in\mathbb{N}} |X_n|^p,\, A\right]^{\frac{1}{p}} \le \frac{p}{p-1} \sup_{n\in\mathbb{N}} \mathbb{E}^\mu\big[|X_n|^p,\, A\big]^{\frac{1}{p}}.$$

Furthermore, for each stopping time ζ, $(X_{n\wedge\zeta}, \mathcal{F}_n, \mu)$ is a submartingale or a martingale depending on whether $(X_n, \mathcal{F}_n, \mu)$ is a submartingale or a martingale. In addition, for any pair of bounded stopping times $\zeta \le \zeta'$,

$$X_\zeta \le \mathbb{E}^\mu\big[X_{\zeta'}\big|\mathcal{F}_\zeta\big] \quad \text{(a.e., }\mu\text{)},$$

and the inequality is an equality in the martingale case. Finally, given $a < b$ and $A \in \mathcal{F}_0$,

$$\mathbb{E}^\mu\big[U_{[a,b]},\, A\big] \le \sup_{n\in\mathbb{N}} \frac{\mathbb{E}^\mu\big[(X_n - a)^+,\, A\big]}{b-a},$$

where $U_{[a,b]}(\omega)$ denotes the precise number of times that $\{X_n(\omega) : n \ge 1\}$ upcrosses $[a,b]$ (cf. the discussion preceding Theorem 5.2.15), and therefore

$$\sup_{n\in\mathbb{N}} \mathbb{E}^\mu\big[X_n^+,\, A\big] < \infty \text{ for every } A \in \mathcal{F}_0 \text{ with } \mu(A) < \infty$$
$$\implies X_n \longrightarrow X \quad \text{(a.e., }\mu\text{)},$$

where X is $\bigvee_0^\infty \mathcal{F}_n$-measurable and locally μ-integrable. In fact, in the case of martingales, there is a $\bigvee_0^\infty \mathcal{F}_n$-measurable, locally μ-integrable X such that

$$X_n = \mathbb{E}^\mu\big[X\big|\mathcal{F}_n\big] \quad \text{(a.e., }\mu\text{)} \quad \text{for all } n \in \mathbb{N}$$

if and only if $\{X_n : n \ge 0\}$ is uniformly μ-integrable on each $A \in \mathcal{F}_0$ with $\mu(A) < \infty$, in which case X is μ-integrable if and only if $X_n \longrightarrow X$ in $L^1(\mu;\mathbb{R})$. On the other hand, when $p \in (1,\infty)$, $X \in L^p(\mu;\mathbb{R})$ if and only if $\{X_n : n \ge 0\}$ is bounded in $L^p(\mu;\mathbb{R})$, in which case, $X_n \longrightarrow X$ in $L^p(\mu;\mathbb{R})$.

PROOF: Obviously, there is no problem unless $\mu(\Omega) = \infty$. However, even then, each of these results follows immediately from its counterpart in § 5.2 once one makes the following trivial observation. Namely, given $\Omega' \in \mathcal{F}_0$ with $\mu(\Omega') \in (0,\infty)$, set

$$\mathcal{F}' = \mathcal{F}[\Omega'], \quad \mathcal{F}'_n = \mathcal{F}_n[\Omega'], \quad X'_n = X_n \restriction \Omega', \quad \text{and } \mathbb{P} = \frac{\mu \restriction \mathcal{F}'}{\mu(\Omega')}.$$

Then $(X'_n, \mathcal{F}'_n, \mathbb{P}')$ is a submartingale or martingale depending on whether the original $(X_n, \mathcal{F}_n, \mu)$ was a submartingale or martingale. Hence, when $\mu(\Omega) = \infty$, simply choose a sequence $\{\Omega_k : k \geq 1\}$ of mutually disjoint, μ-finite elements of $\mathcal{F}_0$ so that $\Omega = \bigcup_1^\infty \Omega_k$, work on each Ω_k separately, and, at the end, sum the results. $\square$

I will now spend a little time seeing how Theorem 6.1.1 can be applied to give a simple proof of the **Hardy–Littlewood maximal inequality**. To state their result, define the maximal function $\mathbf{M}f$ for $f \in L^1(\mathbb{R}^N; \mathbb{R})$ by

$$\mathbf{M}f(\mathbf{x}) = \sup_{Q \ni \mathbf{x}} \frac{1}{|Q|} \int_Q |f(\mathbf{y})| \, d\mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^N,$$

where Q is used to denote a generic **cube**

$$Q = \prod_{j=1}^N [a_j, a_j + r) \quad \text{with } \mathbf{a} \in \mathbb{R}^N \text{ and } r > 0. \tag{6.1.3}$$

As is easily checked, $\mathbf{M}f : \mathbb{R}^N \longrightarrow [0, \infty]$ is lower semicontinuous and therefore certainly Borel measurable. Furthermore, if we restrict our attention to *nicely meshed* families of cubes, then it is easy to relate $\mathbf{M}f$ to martingales. More precisely, for each $n \in \mathbb{Z}$, the n**th standard dyadic partition of** $\mathbb{R}^N$ is the partition $\mathcal{P}_n$ of $\mathbb{R}^N$ into the cubes

$$C_n(\mathbf{k}) \equiv \prod_{i=1}^N \left[\frac{k_i}{2^n}, \frac{k_i + 1}{2^n}\right), \quad \mathbf{k} \in \mathbb{Z}^N. \tag{6.1.4}$$

These partitions are nicely meshed in the sense that the $(n+1)$st is a refinement of the nth. Equivalently, if $\mathcal{F}_n$ denotes the σ-algebra over $\mathbb{R}^N$ generated by the partition $\mathcal{P}_n$, then $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. Moreover, if $f \in L^1(\mathbb{R}^N; \mathbb{R})$ and

$$X_n^f(\mathbf{x}) \equiv 2^{nN} \int_{C_n(\mathbf{k})} \bigl|f(\mathbf{y})\bigr| \, d\mathbf{y} \quad \text{for } \mathbf{x} \in C_n(\mathbf{k}) \text{ and } \mathbf{k} \in \mathbb{Z}^N,$$

then, for each $n \in \mathbb{Z}$,

$$X_n^f = \mathbb{E}^{\lambda_{\mathbb{R}^N}}\bigl[|f| \bigm| \mathcal{F}_n\bigr] \quad (\text{a.e.}, \lambda_{\mathbb{R}^N}),$$

where $\lambda_{\mathbb{R}^N}$ denotes Lebesgue measure on $\mathbb{R}^N$. In particular, for each $m \in \mathbb{Z}$,

$$\bigl(X_{m+n}^f, \mathcal{F}_{m+n}, \lambda_{\mathbb{R}^N}\bigr), \quad n \in \mathbb{N},$$

is a non-negative martingale; and so, by applying (6.1.2) for each $m \in \mathbb{Z}$ and then letting $m \searrow -\infty$, we see that

$$\Bigl|\Bigl\{\mathbf{x} : \mathbf{M}^{(\mathbf{0})} f(\mathbf{x}) \geq \alpha\Bigr\}\Bigr| \leq \frac{1}{\alpha} \int\limits_{\{\mathbf{M}^{(\mathbf{0})} f \geq \alpha\}} |f(\mathbf{y})| \, dy, \quad \alpha \in (0, \infty), \tag{6.1.5}$$

where

$$\mathbf{M}^{(\mathbf{0})}f(\mathbf{x}) = \sup\left\{\frac{1}{|Q|}\int_Q |f(\mathbf{y})|\,d\mathbf{y} : \mathbf{x} \in Q \in \bigcup_{n\in\mathbb{Z}} \mathcal{P}_n\right\}$$

and I have used $|\Gamma|$ to denote $\lambda_{\mathbb{R}^N}(\Gamma)$, the Lebesgue measure of Γ.

At first sight, one might hope that it should be possible to pass directly from (6.1.5) to analogous estimates on the level sets of $\mathbf{M}f$. However, the passage from (6.1.5) to control on $\mathbf{M}f$ is not so easy as it might appear at first: the "sup" in the definition of $\mathbf{M}f$ involves many more cubes than the one in the definition of $\mathbf{M}^{(\mathbf{0})}f$. For this reason I will have to introduce additional families of meshed partitions. Namely, for each $\boldsymbol{\eta} \in \{0,1\}^N$, set

$$\mathcal{P}_n(\boldsymbol{\eta}) = \left\{\frac{(-1)^n\boldsymbol{\eta}}{3\times 2^n} + C_n(\mathbf{k}) : \mathbf{k} \in \mathbb{Z}^N\right\},$$

where $C_n(\mathbf{k})$ is the cube described in (6.1.4). It is then an easy matter to check that, for each $\boldsymbol{\eta} \in \{0,1\}^N$, $\{\mathcal{P}_n(\boldsymbol{\eta}) : n \in \mathbb{Z}\}$ is a family of meshed partitions of $\mathbb{R}^N$. Furthermore, if

$$\left[\mathbf{M}^{(\boldsymbol{\eta})}f\right](\mathbf{x}) = \sup\left\{\frac{1}{|Q|}\int_Q \left|f(\mathbf{y})\right|d\mathbf{y} : \mathbf{x} \in Q \in \bigcup_{n\in\mathbb{Z}} \mathcal{P}_n(\boldsymbol{\eta})\right\}, \quad \mathbf{x} \in \mathbb{R}^N,$$

then exactly the same argument which (when $\boldsymbol{\eta} = \mathbf{0}$) led us to (6.1.5) can now be used to get

$$(*) \qquad \left|\left\{\mathbf{x} \in \mathbb{R}^N : \mathbf{M}^{(\boldsymbol{\eta})}f(\mathbf{x}) \geq \alpha\right\}\right| \leq \frac{1}{\alpha}\int\limits_{\{\mathbf{M}^{(\boldsymbol{\eta})}f\geq\alpha\}} \left|f(\mathbf{y})\right|d\mathbf{y},$$

for each $\boldsymbol{\eta} \in \{0,1\}^N$ and $\alpha \in (0,\infty)$. Finally, if Q is given by (6.1.3) and $r \leq \frac{1}{3\,2^n}$, then it is possible to find an $\boldsymbol{\eta} \in \{0,1\}^N$ and a $C \in \mathcal{P}_n(\boldsymbol{\eta})$ for which $Q \subseteq C$. (To see this, first reduce to the case when $N = 1$.) Hence,

$$\max_{\boldsymbol{\eta}\in\{0,1\}^N} \mathbf{M}^{(\boldsymbol{\eta})}f \leq \mathbf{M}f \leq 6^N \max_{\boldsymbol{\eta}\in\{0,1\}^N} \mathbf{M}^{(\boldsymbol{\eta})}f.$$

After combining this with the estimate in (*), we arrive at the following version of the **Hardy–Littlewood inequality**

$$(6.1.6) \qquad \left|\left\{\mathbf{x} \in \mathbb{R}^N : \mathbf{M}f(\mathbf{x}) \geq \alpha\right\}\right| \leq \frac{(12)^N}{\alpha}\int_{\mathbb{R}^N} |f(\mathbf{y})|\,d\mathbf{y}.$$

At the same time, (*) implies that

$$\max_{\boldsymbol{\eta}\in\{0,1\}^N} \left\|\mathbf{M}^{(\boldsymbol{\eta})}f\right\|_{L^p(\mathbb{R}^N;\mathbb{R})} \leq \frac{p}{p-1}\,\|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad p \in (1,\infty].$$

To check this, first note that it suffices to do so when f vanishes outside of the ball $B(\mathbf{0},R)$ for some $R>0$. Second, assuming that $f=0$ off of $B(\mathbf{0},R)$, observe that (*) implies that

$$\left|\left\{\mathbf{x}\in B(\mathbf{0},R):\,\mathbf{M}^{(\boldsymbol{\eta})}f(\mathbf{x})\geq\alpha\right\}\right|\leq\frac{1}{\alpha}\int\limits_{\{\mathbf{M}^{(\boldsymbol{\eta})\cap B(\mathbf{0},R)}f\geq\alpha\}}\big|f(\mathbf{y})\big|\,d\mathbf{y}.$$

Next, even the result in Exercise 1.4.18 was stated for probability measures, it applies equally well to any finite measure. Thus, we now know that

$$\|Mf\|_{L^p(\mathbb{R}^N;\mathbb{R})}=\lim_{R\to\infty}\left(\int_{B(\mathbf{0},R)}Mf^p(\mathbf{x})\,d\mathbf{x}\right)^{\frac{1}{p}}\leq\frac{p}{p-1}\|f\|_{L^p(\mathbb{R}^N;\mathbb{R})},$$

and so we can repeat the argument just made to obtain

$$\big\|\mathbf{M}f\big\|_{L^p(\mathbb{R}^N;\mathbb{R})}\leq\frac{(12)^Np}{p-1}\,\|f\|_{L^p(\mathbb{R}^N;\mathbb{R})},\quad p\in(1,\infty]. \tag{6.1.7}$$

In this connection, notice that there is no hope of getting this sort of estimate when $p=1$, since it is clear that

$$\varliminf_{|\mathbf{x}|\to\infty}|\mathbf{x}|^N\,\mathbf{M}f(\mathbf{x})>0$$

whenever f does not vanish $\lambda_{\mathbb{R}^N}$-almost everywhere.

The inequality in (6.1.6) plays the same role in classical analysis as Doob's inequality plays in martingale theory. For example, by essentially the same argument as I used to pass from Doob's Inequality to Corollary 5.2.4, we obtain the following famous **Lebesgue Differentiation Theorem**.

THEOREM 6.1.8. *For each $f\in L^1(\mathbb{R}^N;\mathbb{R})$,*

$$\lim_{B\searrow\{\mathbf{x}\}}\frac{1}{|B|}\int_B\big|f(\mathbf{y})-f(\mathbf{x})\big|\,d\mathbf{y}=0 \quad \text{for } \lambda_{\mathbb{R}^N}\text{-almost every } \mathbf{x}\in\mathbb{R}^N, \tag{6.1.9}$$

where, for each $\mathbf{x}\in\mathbb{R}^N$, the limit is taken over balls B which contain $\mathbf{x}$ and tend to $\mathbf{x}$ in the sense that their radii shrink to 0. In particular,

$$f(\mathbf{x})=\lim_{B\searrow\{\mathbf{x}\}}\frac{1}{|B|}\int_B f(\mathbf{y})\,d\mathbf{y}\quad\text{for } \lambda_{\mathbb{R}^N}\text{-almost every } \mathbf{x}\in\mathbb{R}^N.$$

PROOF: I begin with the observation that, for each $f\in L^1(\mathbb{R}^N;\mathbb{R})$,

$$\tilde{\mathbf{M}}f(\mathbf{x})\equiv\sup_{B\ni\mathbf{x}}\frac{1}{|B|}\int_B\big|f(\mathbf{y})\big|\,d\mathbf{y}\leq\kappa_N\mathbf{M}f(\mathbf{x}),\quad\mathbf{x}\in\mathbb{R}^N$$

where $\kappa_n = \frac{2^N}{\mathbf{\Omega}_N}$ with $\mathbf{\Omega}_N = \big|B(\mathbf{0},1)\big|$. Second, notice that (6.1.9) for *every* $\mathbf{x} \in \mathbb{R}^N$ is trivial when $f \in C_c(\mathbb{R}^N;\mathbb{R})$. Hence, all that remains is to check that if $f_n \longrightarrow f$ in $L^1(\mathbb{R}^N;\mathbb{R})$ and if (6.1.9) holds for each f_n, then it holds for f. To this end, let $\epsilon > 0$ be given and check that, because of the preceding and (6.1.6),

$$\begin{aligned}
&\left|\left\{\mathbf{x} : \varlimsup_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B \big|f(\mathbf{y}) - f(\mathbf{x})\big|\,d\mathbf{y} \geq \epsilon\right\}\right| \\
&\quad \leq \left|\left\{\mathbf{x} : \tilde{\mathbf{M}}(f - f_n)(\mathbf{x}) \geq \frac{\epsilon}{3}\right\}\right| \\
&\qquad + \left|\left\{\mathbf{x} : \varlimsup_{B\searrow\{\mathbf{x}\}} \frac{1}{|B|}\int_B \big|f_n(\mathbf{y}) - f_n(\mathbf{x})\big|\,d\mathbf{y} \geq \frac{\epsilon}{3}\right\}\right| \\
&\qquad\quad + \left|\left\{\mathbf{x} : \big|f_n(\mathbf{x}) - f(\mathbf{x})\big| \geq \frac{\epsilon}{3}\right\}\right| \\
&\quad \leq \frac{3}{\epsilon}\big(1 + (12)^N\kappa_N\big)\|f - f_n\|_{L^1(\mathbb{R}^N)}
\end{aligned}$$

for every $n \in \mathbb{Z}^+$. Hence, after letting $n \to \infty$, we see that (6.1.9) also holds for f. $\square$

Although applications like Lebesgue's Differentiation Theorem might make one think that (6.1.6) is most interesting because of what it says about averages over small cubes, its implications for large cubes are also significant. In fact, as I will show in § 6.2, it allows one to prove Birkhoff's Individual Ergodic Theorem (cf. Theorem 6.2.8), which may be viewed as a result about *differentiation at infinity.* The link between ergodic theory and the Hardy–Littlewood Inequality is provided by the following deterministic version of the Maximal Ergodic Lemma (cf. Lemma 6.2.2). Namely, let $\{a_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^N\}$ be a summable subset of $[0,\infty)$, and set

$$\overline{S}_n(\mathbf{k}) = \frac{1}{(2n)^N}\sum_{\mathbf{j}\in Q_n} a_{\mathbf{j}+\mathbf{k}}, \quad n \in \mathbb{N} \text{ and } \mathbf{k} \in \mathbb{Z}^N,$$

where $Q_n = \{\mathbf{j} \in \mathbb{Z}^N : -n \leq j_i < n \text{ for } 1 \leq i \leq N\}$. By applying (6.1.6) and (6.1.7) to the function f given by (cf. (6.1.4)) $f(\mathbf{x}) = a_{\mathbf{k}}$ when $\mathbf{x} \in C_0(\mathbf{k})$, we see that

$$\text{(6.1.10)}\qquad \operatorname{card}\left\{\mathbf{k} \in \mathbb{Z}^N : \sup_{n\in\mathbb{Z}^+} \overline{S}_n(\mathbf{k}) \geq \alpha\right\} \leq \frac{(12)^N}{\alpha}\sum_{\mathbf{k}\in\mathbb{Z}^N} a_{\mathbf{k}}, \quad \alpha \in (0,\infty)$$

and

$$\text{(6.1.11)}\qquad \left(\sum_{\mathbf{k}\in\mathbb{Z}^N}\sup_{n\in\mathbb{Z}^+}|\overline{S}_n(\mathbf{k})|^p\right)^{\frac{1}{p}} \leq \frac{(12)^N p}{p-1}\left(\sum_{\mathbf{k}\in\mathbb{Z}^N}|a_{\mathbf{k}}|^p\right)^{\frac{1}{p}}$$

for each $p \in (1, \infty]$.

The inequality in (6.1.10) is called **Hardy's Inequality**. Actually, Hardy worked in one dimension and was drawn to this line of research by his passion for the game of cricket. What Hardy wanted to find is the optimal order in which to arrange batters to maximize the average score per inning. Thus, he worked with a non-negative sequence $\{a_k : k \geq 0\}$ in which a_k represented the expected number of runs scored by player k, and what he showed is that, for each $\alpha \in (0, \infty)$,

$$\left|\left\{k \in \mathbb{N} : \sup_{n \in \mathbb{Z}^+} \overline{S}_n(k) \geq \alpha\right\}\right|$$

is maximized when $\{a_k\}_0^\infty$ is non-increasing; from which it is an easy application of Markov's inequality to prove that

$$\left|\left\{k \in \mathbb{N} : \sup_{n \in \mathbb{Z}^+} \overline{S}_n(k) \geq \alpha\right\}\right| \leq \frac{1}{\alpha} \sum_0^\infty a_k, \quad \alpha \in (0, \infty).$$

Although this sharpened result can also be obtained as a corollary the *Sunrise Lemma*,* Hardy's approach remains the most appealing.

§ **6.1.2. Banach Space Valued Martingales.** I turn next to Banach space valued martingales. Actually, everything except the easiest aspects of this topic becomes extremely complicated and technical very quickly, and, for this reason, I will restrict my attention to those results which do not involve any deep properties of the geometry of Banach spaces. In fact, the only general theory with which I will deal is contained in the following.

THEOREM 6.1.12. *Let E be a separable Banach space and $\big(X_n, \mathcal{F}_n, \mu\big)$ an E-valued martingale. Then $\big(\|X_n\|_E, \mathcal{F}_n, \mu\big)$ is a non-negative submartingale and therefore, for each $N \in \mathbb{Z}^+$ and all $\alpha \in (0, \infty)$,*

$$\mu\left(\sup_{0 \leq n \leq N} \|X_n\|_E \geq \alpha\right) \leq \frac{1}{\alpha} \mathbb{E}^\mu\left[\|X_N\|_E, \sup_{0 \leq n \leq N} \|X_n\|_E \geq \alpha\right]. \tag{6.1.13}$$

In particular, for each $p \in (1, \infty]$,

$$\left\|\sup_{n \in \mathbb{N}} \|X_n\|_E\right\|_{L^p(\mu;E)} \leq \frac{p}{p-1} \sup_{n \in \mathbb{N}} \|X_n\|_{L^p(\mu;E)}. \tag{6.1.14}$$

Finally, if $X_n = \mathbb{E}^\mu[X \,|\, \mathcal{F}_n]$ where $X \in L^p(\mu; E)$ for some $p \in [1, \infty)$, then

$$X_n \longrightarrow \mathbb{E}^\mu\left[X \,\middle|\, \bigvee_0^\infty \mathcal{F}_n\right] \text{ both (a.e., } \mu) \text{ and in } L^p(\mu; E).$$

---

* See Lemma 3.4.5 in my *A Concise Introduction to the Theory of Integration*, 3rd edition, published by Birkhauser in 1998.

Proof: The fact $\big(\|X_n\|_E, \mathcal{F}_n, \mu\big)$ is a submartingale is an easy application of the inequality in (5.1.14); and, given this fact, the inequalities in (6.1.13) and (6.1.14) follow from the corresponding inequalities in Theorem 6.1.1.

While proving the convergence statement, I may and will assume that $\mathcal{F} = \bigvee_0^\infty \mathcal{F}_n$. Now let $X \in L^p(\mu; E)$ be given, and set $X_n = \mathbb{E}^\mu[X|\mathcal{F}_n]$, $n \in \mathbb{N}$. Because of (6.1.13) and (6.1.14), we know (cf. the proofs of Corollary 5.2.4 and Theorem 6.1.8) that the set of X for which $X_n \longrightarrow X$ (a.e., μ) is a closed subset of $L^p(\mu; E)$. Moreover, if X is μ-simple, then the μ-almost everywhere convergence of X_n to X follows easily from the $\mathbb{R}$-valued result. Hence, we now know that $X_n \longrightarrow X$ (a.s, μ) for each $X \in L^1(\mu; E)$. In addition, because of (6.1.14), when $p \in (1, \infty)$, the convergence in $L^p(\mu; E)$ follows by Lebesgue's Dominated Convergence Theorem. Finally, to prove the convergence in $L^1(\mu; E)$ when $X \in L^1(\mu; E)$, note that, by Fatou's Lemma,

$$\|X\|_{L^1(\mu;E)} \leq \varliminf_{n\to\infty} \|X_n\|_{L^1(\mu;E)};$$

whereas (5.1.14) guarantees that

$$\|X\|_{L^1(\mu;E)} \geq \varlimsup_{n\to\infty} \|X_n\|_{L^1(\mu;E)}.$$

Hence, because

$$\Big| \|X_n\|_E - \|X\|_E - \|X_n - X\|_E \Big| \leq 2\|X\|_E,$$

the convergence in $L^1(\mu; E)$ is again an application of Lebesgue's Dominated Convergence Theorem. $\square$

Going beyond the convergence result in Theorem 6.1.12 to get an analog of Doob's Martingale Convergence Theorem is hard. For one thing, a naïve analog is not even true for general separable Banach spaces, and a rather deep analysis of the geometry of Banach spaces is required in order to determine exactly when it is true. (See Exercise 6.1.18 for a case in which it is.)

Exercises for § 6.1

Exercise 6.1.15. In this exercise we will develop Jensen's inequality in the Banach space setting. Thus, $(\Omega, \mathcal{F}, \mathbb{P})$ will be a probability space, C will be a closed, convex subset of the separable Banach space E, and X will be a C-valued element of $L^1(\mathbb{P}; E)$.

(**i**) Show that there exists a sequence $\{X_n : n \geq 1\}$ of C-valued, simple functions which tend to X both $\mathbb{P}$-almost surely and in $L^1(\mathbb{P}; E)$.

(**ii**) Show that $\mathbb{E}^{\mathbb{P}}[X] \in C$ and that

$$\mathbb{E}^{\mathbb{P}}\big[g(X)\big] \leq g\big(\mathbb{E}^{\mathbb{P}}[X]\big)$$

for every continuous, concave $g : C \longrightarrow [0, \infty)$.

(**iii**) Given a sub-σ-algebra Σ of $\mathcal{F}$, follow the argument in Corollary 5.2.8 to show that there exists a sequence $\{\mathcal{P}_n\}_0^\infty$ of finite, Σ-measurable partitions with the property that

$$\sum_{A\in\mathcal{P}_n} \frac{\mathbb{E}^{\mathbb{P}}[X,\, A]}{\mathbb{P}(A)}\,\mathbf{1}_A \longrightarrow \mathbb{E}^{\mathbb{P}}[X|\Sigma] \quad \text{both } \mathbb{P}\text{-almost surely and in } L^1(\mathbb{P};E).$$

In particular, conclude that there is a representative X_Σ of $\mathbb{E}^{\mathbb{P}}[X|\Sigma]$ which is C-valued and that

$$\mathbb{E}^{\mathbb{P}}\big[g(X)\big|\Sigma\big] \le g\big(X_\Sigma\big) \quad (\text{a.s.}, \mathbb{P})$$

for each continuous, convex $g : C \longrightarrow [0,\infty)$.

Exercise 6.1.16. Again let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and E a separable Banach space. Further, suppose that $\{\mathcal{F}_n : n \ge 0\}$ is a *non-increasing* sequence of sub-σ-algebras of $\mathcal{F}$, and set $\mathcal{F}_\infty = \bigcap_0^\infty \mathcal{F}_n$. Finally, let $X \in L^1(\mathbb{P};E)$.

(**i**) Show that

$$\mathbb{E}^{\mathbb{P}}\big[X\big|\mathcal{F}_n\big] \longrightarrow \mathbb{E}^{\mathbb{P}}[X|\mathcal{F}_\infty] \quad \text{both } \mathbb{P}\text{-almost surely and in } L^p(\mathbb{P};E)$$

for any $p \in [1,\infty)$ with $X \in L^p(\mathbb{P};E)$.

Hint: Use (6.1.13) and the approximation result in Theorem 5.1.10 to reduce to the case when X is simple. When X is simple, get the result as an application of the convergence theorem for $\mathbb{R}$-valued, reversed martingales in Theorem 5.2.21.

(**ii**) Using part (**i**) and following the line of reasoning suggested at the end of § 5.2.4, give a proof of The Strong Law of Large Numbers for Banach space-valued random variables.* (See Exercises 6.2.19 and 9.1.18 for an entirely different approaches.)

Exercise 6.1.17. As we saw in the proof of Theorem 6.1.8, the Hardy–Littlewood maximal function can be used to dominate other quantities of interest. As further indication of its importance, I will use it in this exercise to prove the analogue of Theorem 6.1.8 for a large class of approximate identities. That is, let $\psi \in L^1(\mathbb{R}^N;\mathbb{R})$ with $\int_{\mathbb{R}^N}\psi(\mathbf{x})\,d\mathbf{x} = 1$ be given, and set

$$\psi_t(\mathbf{x}) = t^{-N}\psi\left(\tfrac{\mathbf{x}}{t}\right), \quad t \in (0,\infty) \text{ and } \mathbf{x} \in \mathbb{R}^N.$$

Then $\{\psi_t : t > 0\}$ forms an **approximate identity** in the sense that, as tempered distributions, $\psi_t \longrightarrow \delta_{\mathbf{0}}$ as $t \searrow 0$. In fact, because

$$\|\psi_t \star f\|_{L^p(\mathbb{R}^N;\mathbb{R})} \le \|\psi\|_{L^1(\mathbb{R}^N;\mathbb{R})}\,\|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad t \in (0,\infty) \text{ and } p \in [1,\infty],$$

* This proof, which seems to have been the first, of the Strong Law for Banach space was given by E. Mourier in "Eléments aléatoires dans un espace de Banach," *Ann. Inst. Poincaré.*

and

$$\psi_t \star f(\mathbf{x}) = \int_{\mathbb{R}^N} \psi(\mathbf{y})\, f(\mathbf{x} - t\mathbf{y})\, d\mathbf{y},$$

it is easy to see that, for each $p \in [1, \infty)$,

$$\lim_{t \searrow 0} \big\|\psi_t \star f - f\big\|_{L^p(\mathbb{R}^N;\mathbb{R})} = 0$$

first for $f \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$ and then for all $f \in L^p(\mathbb{R}^N;\mathbb{R})$.

The purpose of this exercise is to sharpen the preceding under the assumption that

$$\psi(\mathbf{x}) = \alpha\big(|\mathbf{x}|\big), \quad \mathbf{x} \in \mathbb{R}^N \setminus \{\mathbf{0}\} \quad \text{for some } \alpha \in C^1\big((0,\infty);\mathbb{R}\big) \text{ with}$$
$$A \equiv \int_{(0,\infty)} r^N |\alpha'(r)|\, dr < \infty.$$

Notice that when α is non-negative and non-increasing, integration by parts shows that $A = N$.

(**i**) Let $f \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$ be given, and set

$$\tilde{f}(r, \mathbf{x}) = \frac{1}{|B(\mathbf{x}, r)|} \int_{B(\mathbf{x},r)} f(\mathbf{y})\, d\mathbf{y} \quad \text{for } r \in (0,\infty) \text{ and } \mathbf{x} \in \mathbb{R}^N.$$

Using integration by parts and the given hypotheses, show that

$$\psi_t \star f(\mathbf{x}) = -\tfrac{1}{N} \int_{(0,\infty)} r^N \alpha'(r)\, \tilde{f}(tr, \mathbf{x})\, dr,$$

and conclude that

$$\big|\psi_t \star f(\mathbf{x})\big| \le \tfrac{A}{N}\, \tilde{\mathbf{M}} f(\mathbf{x}),$$

where $\tilde{\mathbf{M}} f$ is the quantity introduced at the beginning of the proof of Theorem 6.1.8. In particular, conclude that there is a constant $K_N \in (0,\infty)$, depending only on $N \in \mathbb{Z}^+$, such that

$$\mathbf{M}_\psi f(\mathbf{x}) \equiv \sup_{t \in (0,\infty)} \big|\psi_t \star f(\mathbf{x})\big| \le K_N A\, \mathbf{M} f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^N.$$

(**ii**) Starting from the conclusion in (**i**), show that

$$\big|\{\mathbf{x} : \mathbf{M}_\psi f(\mathbf{x}) \ge R\}\big| \le \frac{(12)^N K_N A \|f\|_{L^1(\mathbb{R}^N)}}{R}, \quad f \in L^1(\mathbb{R}^N;\mathbb{R}),$$

and that for $p \in (1, \infty]$,

$$\|\mathbf{M}_\psi f\|_{L^p(\mathbb{R}^N;\mathbb{R})} \le \frac{(12)^N K_N A\, p}{p-1} \|f\|_{L^p(\mathbb{R}^N;\mathbb{R})}, \quad f \in L^p(\mathbb{R}^N;\mathbb{R}).$$

Finally, proceeding as in the proof of Theorem 6.1.8, use the first of these to prove that, for $f \in L^1(\mathbb{R}^N;\mathbb{R})$ and Lebesgue almost every $\mathbf{x} \in \mathbb{R}^N$:

$$\begin{aligned}&\varlimsup_{t \searrow 0} \big|\psi_t \star f(\mathbf{x}) - f(\mathbf{x})\big| \\ &\quad \le \varlimsup_{t \searrow 0} \int_{\mathbb{R}^N} \big|\psi_t(\mathbf{y})\big(f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x})\big)\big|\, d\mathbf{y} = 0.\end{aligned}$$

Two of the most familiar examples to which the preceding applies are the Gauss kernel $g_t(\mathbf{x}) = (2\pi t)^{-\frac{N}{2}} \exp\big(-\frac{|\mathbf{x}|^2}{2}\big)$ and the Poisson kernel (cf. (3.2.45)) $\Pi_t^{\mathbb{R}^N}$. In both these cases, $A = N$.

Exercise 6.1.18. Let E be a separable Hilbert space and $(X_n, \mathcal{F}, \mathbb{P})$ an E-valued martingale on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the condition

$$\sup_{n \in \mathbb{Z}^+} \mathbb{E}^{\mathbb{P}}\big[\|X_n\|_E^2\big] < \infty.$$

Proceeding as in (**i**) of Exercise 5.2.28, first prove that there is a $\bigvee_1^\infty \mathcal{F}_n$-measurable $X \in L^2(\mathbb{P}; E)$ to which $\{X_n : n \ge 1\}$ converges in $L^2(\mathbb{P}; E)$, next check that

$$X_n = \mathbb{E}^{\mathbb{P}}\big[X \big| \mathcal{F}_n\big] \quad \text{(a.s., } \mathbb{P}) \text{ for each } n \in \mathbb{Z}^+,$$

and finally apply the last part of Theorem 6.1.12 to see that $X_n \longrightarrow X$ $\mathbb{P}$-almost surely.

§ 6.2 Elements of Ergodic Theory

Among the two or three most important general results about dynamical systems is D. Birkhoff's Individual Ergodic Theorem. In this section, I will present a generalization, due to Wiener, of Birkhoff's basic theorem.

The setting in which I will prove the Ergodic Theorem will be the following. $(\Omega, \mathcal{F}, \mu)$ will be a σ-finite measure space on which there exits a semigroup $\{\mathbf{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ of measurable, μ-**measure preserving transformations.** That is, for each $\mathbf{k} \in \mathbb{N}^N$, $\mathbf{\Sigma}^{\mathbf{k}}$ is an $\mathcal{F}$-measurable map from Ω into itself, $\mathbf{\Sigma}^{\mathbf{0}}$ is the identity map, $\mathbf{\Sigma}^{\mathbf{k}+\boldsymbol{\ell}} = \mathbf{\Sigma}^{\mathbf{k}} \circ \mathbf{\Sigma}^{\boldsymbol{\ell}}$ for all $\mathbf{k}$, $\boldsymbol{\ell} \in \mathbb{N}^N$, and

$$\mu(\Gamma) = \mu\big((\mathbf{\Sigma}^{\mathbf{k}})^{-1}(\Gamma)\big) \quad \text{for all } \mathbf{k} \in \mathbb{N} \text{ and } \Gamma \in \mathcal{F}.$$

Further, E will be a separable Banach space with norm $\|\cdot\|_E$; and, given a function $F : \Omega \longrightarrow E$, I will be considering the averages

$$\mathbf{A}_n F(\omega) \equiv \frac{1}{n^N} \sum_{\mathbf{k} \in Q_n^+} F \circ \mathbf{\Sigma}^{\mathbf{k}}(\omega), \quad n \in \mathbb{Z}^+, \tag{6.2.1)6.1.1$$

where Q_n^+ is the cube $\{\mathbf{k} \in \mathbb{N}^N : \|\mathbf{k}\|_\infty < n\}$ and $\|\mathbf{k}\|_\infty \equiv \max_{1\le j\le N} k_j$. My goal (cf. Theorem 6.2.8 below) is to show that for each $p \in [1,\infty)$ and $F \in L^p(\mu;E)$, $\{\mathbf{A}_n F : n \ge 1\}$ converges μ-almost everywhere. In fact, when either μ is finite or $p \in (1,\infty)$, I will show that the convergence is also in $L^p(\mu;E)$.

§ **6.2.1. The Maximal Ergodic Lemma.** Because he was thinking in terms of dynamical systems and therefore did not take full advantage of measure theory, Birkhoff's own proof of his theorem is rather cumbersome. Later, F. Riesz discovered a proof which has become the model for all later proofs. Specifically, he introduced what is now called the Maximal Ergodic Inequality, which is an inequality that plays the same role here that Doob's Inequality played in the derivation of Corollary 5.2.4. In order to cover Wiener's extension of Birkhoff's theorem, I will derive a multi-parameter version of the Maximal Ergodic Inequality, which, as the proof shows, is really just a clever application of Hardy's Inequality.*

LEMMA 6.2.2 (**Maximal Ergodic Lemma**). *For each $n \in \mathbb{Z}^+$ and $p \in [1,\infty]$, $\mathbf{A}_n$ is a contraction on $L^p(\mu;E)$. Moreover, for each $F \in L^p(\mu;E)$:*

$$\mu\left(\sup_{n\ge1}\|\mathbf{A}_nF\|_E \ge \lambda\right) \le \frac{(24)^N}{\lambda}\|F\|_{L^1(\mu;E)}, \quad \lambda\in(0,\infty), \tag{6.2.3}$$

or

$$\left\|\sup_{n\ge1}\|\mathbf{A}_nF\|_E\right\|_{L^p(\mu)} \le \frac{(24)^N p}{p-1}\|F\|_{L^p(\mu;E)}, \tag{6.2.4}$$

depending on whether $p = 1$ or $p \in (1,\infty)$.

PROOF: First observe that, because $\|\mathbf{A}_nF\|_E \le \mathbf{A}_n\|F\|_E$, it suffices to prove all of these assertions in the case when $E = \mathbb{R}$ and F is non-negative. Thus, I will restrict myself to this case. Since $F \circ \mathbf{\Sigma}^{\mathbf{k}}$ has the same distribution as F itself, the first assertion is trivial. To prove (6.2.3) and (6.2.4), let $n \in \mathbb{Z}^+$ be given, apply (6.1.10) and (6.1.11) to

$$a_{\mathbf{k}}(\omega) \equiv \begin{cases} F\circ\mathbf{\Sigma}^{\mathbf{k}}(\omega) & \text{if } \mathbf{k}\in Q_{2n}^+ \\ 0 & \text{if } \mathbf{k}\notin Q_{2n}^+, \end{cases}$$

and conclude that

$$C_n(\omega) \equiv \left|\left\{\mathbf{k}\in Q_n^+ : \max_{1\le m\le n}\mathbf{A}_m\big(F\circ\mathbf{\Sigma}^{\mathbf{k}}\big)(\omega)\ge\lambda\right\}\right| \le \frac{(12)^N}{\lambda}\sum_{\mathbf{k}\in Q_{2n}^+}F\circ\mathbf{\Sigma}^{\mathbf{k}}(\omega)$$

* The idea of using Hardy's Inequality was suggested to P. Hartman by J. von Neumann and appears for the first time in Hartman's "On the ergodic theorem," *Am. J. Math.* **69**: 193–199 (1947).

and

$$\sum_{\mathbf{k}\in Q_n^+} \max_{1\le m\le n} \Big(\mathbf{A}_m\big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}\big)(\omega)\Big)^p \le \left(\frac{(12)^N p}{p-1}\right)^p \sum_{\mathbf{k}\in Q_{2n}^+} \Big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}(\omega)\Big)^p.$$

Hence, by Tonelli's Theorem,

$$\begin{aligned}\sum_{\mathbf{k}\in Q_n^+} \mu\left(\max_{1\le m\le n} \mathbf{A}_m\big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}\big)\ge\lambda\right) &= \int C_n(\omega)\,\mu(d\omega)\\ &\le \frac{(12)^N}{\lambda}\sum_{\mathbf{k}\in Q_{2n}^+}\int F\circ\boldsymbol{\Sigma}^{\mathbf{k}} f\,d\mu\end{aligned}$$

and, similarly,

$$\sum_{\mathbf{k}\in Q_n^+}\int \max_{1\le m\le n}\Big(\mathbf{A}_m\big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}\big)\Big)^p d\mu \le \left(\frac{(12)^N p}{p-1}\right)^p \sum_{\mathbf{k}\in Q_{2n}^+}\int\Big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}\Big)^p d\mu.$$

Finally, since the distributions of $\max_{1\le m\le n}\mathbf{A}_m\big(F\circ\boldsymbol{\Sigma}^{\mathbf{k}}\big)$ and $F\circ\boldsymbol{\Sigma}^{\mathbf{k}}$ do not depend on $\mathbf{k}\in\mathbb{N}^N$, the preceding lead immediately to

$$\mu\left(\max_{1\le m\le n}\mathbf{A}_m F\ge\lambda\right)\le\frac{(24)^N}{\lambda}\|F\|_{L^1(\mu)}$$

and

$$\left\|\max_{1\le m\le n}\mathbf{A}_m F\right\|_{L^p(\mu)}\le\frac{2^{\frac{N}{p}}(12)^N p}{p-1}\|F\|_{L^p(\mu)}$$

for all $n\in\mathbb{Z}^+$. Thus, (6.2.3) and (6.2.4) follow after one lets $n\to\infty$. $\square$

Given (6.2.3) and (6.2.4), I adopt again the strategy used in the proof of Corollary 5.2.4. That is, we must begin by finding a dense subset of each L^p-space on which the desired convergence results can be checked by hand, and for this purpose I will have to introduce the notion of invariance.

A set $\Gamma\in\mathcal{F}$ is said to be **invariant**, and I write $\Gamma\in\mathfrak{I}$, if $\Gamma=(\boldsymbol{\Sigma}^{\mathbf{k}})^{-1}(\Gamma)$ for every $\mathbf{k}\in\mathbb{N}^N$. As is easily checked, $\mathfrak{I}$ is a sub-σ-algebra of $\mathcal{F}$. In addition, it is clear that $\Gamma\in\mathcal{F}$ is invariant if $\Gamma=(\boldsymbol{\Sigma}^{\mathbf{e}_j})^{-1}(\Gamma)$ for each $1\le j\le N$, where $\{\mathbf{e}_i : 1\le i\le N\}$ is the standard orthonormal basis in $\mathbb{R}^N$. Finally, if $\overline{\mathfrak{I}}$ is the μ-completion of $\mathfrak{I}$ relative to $\mathcal{F}$ in the sense that $\Gamma\in\overline{\mathfrak{I}}$ if and only if $\Gamma\in\mathcal{F}$ and there is $\tilde{\Gamma}\in\mathfrak{I}$ such that $\mu(\Gamma\Delta\tilde{\Gamma})=0$ ($A\Delta B\equiv(A\setminus B)\cup(B\setminus A)$ is the **symmetric difference** between the sets A and B), then an $\mathcal{F}$-measurable $F:\Omega\longrightarrow E$ is $\overline{\mathfrak{I}}$-measurable if and only if $F=F\circ\boldsymbol{\Sigma}^{\mathbf{k}}$ (a.e., μ) for each $\mathbf{k}\in\mathbb{N}^N$. Indeed, one need only check this equivalence for indicator functions of sets. But if $\Gamma \in \mathcal{F}$ and $\mu(\Gamma \Delta \tilde{\Gamma}) = 0$ for some $\tilde{\Gamma} \in \mathfrak{I}$, then

$$\mu\Big(\Gamma \Delta (\mathbf{\Sigma}^{\mathbf{k}})^{-1}(\Gamma)\Big) \le \mu\big((\mathbf{\Sigma}^{\mathbf{k}})^{-1}(\Gamma \Delta \tilde{\Gamma})\big) + \mu(\Gamma \Delta \tilde{\Gamma}) = 0,$$

and so $\Gamma \in \overline{\mathfrak{I}}$. Conversely, if $\Gamma \in \overline{\mathfrak{I}}$, set

$$\tilde{\Gamma} = \bigcup_{\mathbf{k} \in \mathbb{N}^N} (\mathbf{\Sigma}^{\mathbf{k}})^{-1}(\Gamma),$$

and check that $\tilde{\Gamma} \in \mathfrak{I}$ and $\mu(\Gamma \Delta \tilde{\Gamma}) = 0$.

Lemma 6.2.5. *Let $\mathfrak{I}(E)$ be the subspace of $\overline{\mathfrak{I}}$-measurable elements of $L^2(\mu; E)$. Then, $\mathfrak{I}(E)$ is a closed linear subspace of $L^2(\mu; E)$. Moreover, if $\Pi_{\mathfrak{I}(\mathbb{R})}$ denotes orthogonal projection from $L^2(\mu; \mathbb{R})$ onto $\mathfrak{I}(\mathbb{R})$, then there exists a unique linear contraction $\Pi_{\mathfrak{I}(E)} : L^2(\mu; E) \longrightarrow \mathfrak{I}(E)$ with the property that $\Pi_{\mathfrak{I}(E)}(\mathbf{a} f) = \mathbf{a}\Pi_{\mathfrak{I}(\mathbb{R})} f$ for $\mathbf{a} \in E$ and $f \in L^2(\mu; \mathbb{R})$. Finally, for each $F \in L^2(\mu; E)$,*

$$\mathbf{A}_n F \longrightarrow \Pi_{\mathfrak{I}(E)} F \quad \text{(a.e., } \mu) \text{ and in } L^2(\mu; E). \tag{6.2.6}$$

Proof: I begin with the case when $E = \mathbb{R}$. The first step is to identify the orthogonal complement $\mathfrak{I}(\mathbb{R})^\perp$ of $\mathfrak{I}(\mathbb{R})$. To this end, let $\mathcal{N}$ denote the subspace of $L^2(\mu; \mathbb{R})$ consisting of elements having the form $g - g \circ \mathbf{\Sigma}^{\mathbf{e}_j}$ for some $g \in L^2(\mu; \mathbb{R}) \cap L^\infty(\mu; \mathbb{R})$ and $1 \le j \le N$. Given $f \in \mathfrak{I}(\mathbb{R})$, observe that

$$\big(f, g - g \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big)_{L^2(\mu;\mathbb{R})} = \big(f, g\big)_{L^2(\mu;\mathbb{R})} - \big(f \circ \mathbf{\Sigma}^{\mathbf{e}_j}, g \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big)_{L^2(\mu;\mathbb{R})} = 0.$$

Hence, $\mathcal{N} \subseteq \mathfrak{I}(\mathbb{R})^\perp$. On the other hand, if $f \in L^2(\mu; \mathbb{R})$ and $f \perp \mathcal{N}$, then it is clear that $f \perp f - f \circ \mathbf{\Sigma}^{\mathbf{e}_j}$ for each $1 \le j \le N$ and therefore that

$$\begin{aligned}
&\big\|f - f \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big\|^2_{L^2(\mu;\mathbb{R})} \\
&\quad = \|f\|^2_{L^2(\mu;\mathbb{R})} - 2\big(f, f \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big)_{L^2(\mu;\mathbb{R})} + \big\|f \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big\|^2_{L^2(\mu;\mathbb{R})} \\
&\quad = 2\Big(\|f\|^2_{L^2(\mu;\mathbb{R})} - \big(f, f \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big)_{L^2(\mu;\mathbb{R})}\Big) = 2\big(f, f - f \circ \mathbf{\Sigma}^{\mathbf{e}_j}\big)_{L^2(\mu;\mathbb{R})} = 0.
\end{aligned}$$

Thus, for each $1 \le j \le N$, $f = f \circ \mathbf{\Sigma}^{\mathbf{e}_j}$ μ-almost everywhere; and, by induction on $\|\mathbf{k}\|_\infty$, one concludes that $f = f \circ \mathbf{\Sigma}^{\mathbf{k}}$ μ-almost everywhere for all $\mathbf{k} \in \mathbb{N}^N$. In other words, we have now shown that $\mathfrak{I}(\mathbb{R}) = \mathcal{N}^\perp$, or, equivalently, that $\overline{\mathcal{N}} = \mathfrak{I}(\mathbb{R})^\perp$.

Continuing with $E = \mathbb{R}$, next note that if $f \in \overline{\mathfrak{I}(\mathbb{R})}$ then $\mathbf{A}_n f = f$ (a.e., μ) for each $n \in \mathbb{Z}^+$. Hence, (6.2.6) is completely trivial in this case. On the other hand, if $g \in L^2(\mu; \mathbb{R}) \cap L^\infty(\mu; \mathbb{R})$ and $f = g - g \circ \mathbf{\Sigma}^{\mathbf{e}_j}$, then

$$n^N \mathbf{A}_n f = \sum_{\{\mathbf{k} \in Q_n^+ : k_j = 0\}} g \circ \mathbf{\Sigma}^{\mathbf{k}} - \sum_{\{\mathbf{k} \in Q_n^+ : k_j = n-1\}} g \circ \mathbf{\Sigma}^{\mathbf{k} + \mathbf{e}_j},$$

and so, with $p \in \{2, \infty\}$,

$$\left\|\mathbf{A}_n f\right\|_{L^p(\mu;\mathbb{R})} \le \frac{2\|g\|_{L^p(\mu;\mathbb{R})}}{n} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Hence, in this case also, (6.2.6) is easy. Finally, to complete the proof for $E = \mathbb{R}$, simply note that, by (6.2.4) with $p = 2$ and $E = \mathbb{R}$, the set of $f \in L^2(\mu;\mathbb{R})$ for which (6.2.6) holds is a closed linear subspace of $L^2(\mu;\mathbb{R})$ and that we have already verified (6.2.6) for $f \in \mathfrak{I}(\mathbb{R})$ and f from a dense subspace of $\mathfrak{I}(\mathbb{R})^\perp$.

Turning to general E's, first note $\Pi_{\mathfrak{I}(E)}F$ is well-defined for μ-simple F's. Indeed, if $F = \sum_1^\ell \mathbf{a}_i \mathbf{1}_{\Gamma_i}$ for some $\{\mathbf{a}_i\}_1^\ell \subseteq E$ and $\{\Gamma_i\}_1^\ell$ of mutually disjoint elements of $\mathcal{F}$ with finite μ-measure, then

$$\Pi_{\mathfrak{I}(E)}F = \sum_1^\ell \mathbf{a}_i \Pi_{\mathfrak{I}(\mathbb{R})} \mathbf{1}_{\Gamma_i}$$

and so

$$\begin{aligned}\left\|\Pi_{\mathfrak{I}(E)}F\right\|^2_{L^2(\mu;E)} &\le \int \left(\sum_1^\ell \|\mathbf{a}_i\|_E \Pi_{\mathfrak{I}(\mathbb{R})} \mathbf{1}_{\Gamma_i}\right)^2 d\mu \\ &= \left\|\Pi_{\mathfrak{I}(\mathbb{R})}\left(\sum_1^\ell \|\mathbf{a}_i\|_E \mathbf{1}_{\Gamma_i}\right)\right\|^2_{L^2(\mu;\mathbb{R})} \le \sum_1^\ell \|\mathbf{a}_i\|_E^2 \mu(\Gamma_i) = \|F\|^2_{L^2(\mu;E)}.\end{aligned}$$

Thus, since the space of μ-simple functions is dense in $L^2(\mu;E)$, it is clear that $\Pi_{\mathfrak{I}(E)}$ not only exists but is also unique.

Finally, to check (6.2.6) for general E's, note that (6.2.6) for E-valued, μ-simple F's is an immediate consequence of (6.2.6) for $E = \mathbb{R}$. Thus, we already know (6.2.6) for a dense subspace of $L^2(\mu;E)$; and so the rest is another elementary application of (6.2.4). $\square$

§ **6.2.2. Birkhoff's Ergodic Theorem.** For general $p \in [1,\infty)$, let $\mathfrak{I}^p(E)$ denote the subspace of $\overline{\mathfrak{I}}$-measurable elements of $L^p(\mu;E)$. Clearly $\mathfrak{I}^p(E)$ is closed for every $p \in [1,\infty)$. Moreover, since

$$\mu(\Omega) < \infty \implies \Pi_{\mathfrak{I}(E)}F = \mathbb{E}^\mu\left[F\middle|\mathfrak{I}\right], \tag{6.2.7}$$

whenever μ is finite $\Pi_{\mathfrak{I}(E)}$ extends automatically as a linear contraction from $L^p(\mu;E)$ onto $\mathfrak{I}^p(E)$ for each $p \in [1,\infty)$, the extension being given by the right-hand side of (6.2.7). However, when $\mu(E) = \infty$, there is a problem. Namely, because $\mu \restriction \mathfrak{I}$ will seldom be σ-finite, it will not be possible to condition μ with respect to $\mathfrak{I}$. Be that as it may, (6.2.6) provides an extension of $\Pi_{\mathfrak{I}(E)}$. Namely, from (6.2.6) and Fatou's Lemma, it is clear that, for each $p \in [1,\infty)$,

$$\left\|\Pi_{\mathfrak{I}(E)}F\right\|_{L^p(\mu;E)} \le \|F\|_{L^p(\mu;E)}, \quad F \in L^p(\mu;E) \cap L^2(\mu;E);$$

and therefore the desired existence of the extension follows by continuity.

Theorem 6.2.8 (**The Individual Ergodic Theorem**). *For each $p \in [1,\infty)$ and $F \in L^p(\mu;E)$:*

$$\mathbf{A}_n F \longrightarrow \Pi_{\mathfrak{I}(E)} F \quad \text{(a.e., } \mu). \tag{6.2.9}$$

Moreover, if either $p \in (1,\infty)$ or $p = 1$ and $\mu(\Omega) < \infty$, then the convergence in (6.2.9) is also in $L^p(\mu;E)$. Finally, if $\mu(\Gamma) \wedge \mu(\Gamma\complement) = 0$ for every $\Gamma \in \mathfrak{I}$, then (6.2.9) can be replaced by

$$\lim_{n\to\infty} \mathbf{A}_n F = \begin{cases} \dfrac{\mathbb{E}^\mu[F]}{\mu(\Omega)} & \text{if } \mu(\Omega) \in (0,\infty) \\ 0 & \text{if } \mu(\Omega) = \infty \end{cases} \quad \text{(a.e., } \mu),$$

and the convergence is in $L^p(\mu;E)$ when either $p \in (1,\infty)$ or $p = 1$ and $\mu(\Omega) < \infty$.

Proof: As I said above, the proof is now an easy application of the strategy used to prove Corollary 5.2.4. Namely, by (6.2.3), the set of $F \in L^1(\mu;E)$ for which (6.2.9) holds is closed and, by (6.2.6), it includes $L^1(\mu;E) \cap L^\infty(\mu;E)$. Hence, (6.2.9) is proved for $p = 1$. On the other hand, when $p \in (1,\infty)$, (6.2.4) applies and shows first that the set of $F \in L^p(\mu;E)$ for which (6.2.9) holds is closed in $L^p(\mu;E)$ and second that μ-almost everywhere convergence already implies convergence in $L^p(\mu;E)$. Hence, we have proved that (6.2.9) holds and that the convergence is in $L^p(\mu;E)$ when $p \in (1,\infty)$. In addition, when $\mu(\Gamma) \wedge \mu(\Gamma\complement) = 0$ for all $\Gamma \in \mathfrak{I}$, it is clear that the only elements of $\mathfrak{I}^p(E)$ are μ-almost everywhere constant, which, in the case when $\mu(\Omega) < \infty$ means (cf. (6.2.7)) that $\Pi_{\mathfrak{I}(E)} F = \frac{\mathbb{E}^\mu[F]}{\mu(\Omega)}$, and, when $\mu(\Omega) = \infty$, means that $\mathfrak{I}^p(E) = \{0\}$ for all $p \in [1,\infty)$.

In view of the preceding, all that remains is to discuss the $L^1(\mu;E)$-convergence in the case when $p = 1$ and $\mu(\Omega) < \infty$. To this end, observe that, because the $\mathbf{A}_n$'s are all contractions in $L^1(\mu;E)$, it suffices to prove $L^1(\mu;E)$-convergence for E-valued, μ-simple F's. But $L^1(\mu;E)$-convergence for such F's reduces to showing that $\mathbf{A}_n f \longrightarrow \Pi_{\mathfrak{I}(\mathbb{R})} f$ in $L^1(\mu;\mathbb{R})$ for non-negative $f \in L^\infty(\mu;\mathbb{R})$. Finally, if $f \in L^1\big(\mu;[0,\infty)\big)$, then

$$\big\|\mathbf{A}_n f\|_{L^1(\mu)} = \|f\|_{L^1(\mu)} = \big\|\Pi_{\mathfrak{I}(\mathbb{R})} f\|_{L^1(\mu;\mathbb{R})}, \quad n \in \mathbb{Z}^+,$$

where, in the last equality I used (6.2.7); and this, together with (6.2.9), implies (cf. the final step in the proof of Theorem 6.1.12) convergence in $L^1(\mu)$. $\square$

I will say that semigroup $\big\{\mathbf{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\big\}$ is **ergodic** on $(\Omega,\mathcal{F},\mu)$ if, in addition to being μ-measure preserving, $\mu(\Gamma) \wedge \mu(\Gamma\complement) = 0$ for every invariant $\Gamma \in \mathfrak{I}$.

Classic Example. In order to get a feeling for what the Ergodic Theorem is saying, take μ to be Lebesgue measure on the interval $[0,1)$ and, for a given $\alpha \in (0,1)$, define $\mathbf{\Sigma}_\alpha : [0,1) \longrightarrow [0,1)$ so that

$$\mathbf{\Sigma}_\alpha(\omega) \equiv \omega + \alpha - [\omega + \alpha] = \omega + \alpha \bmod 1.$$

If α is rational and m is the smallest element of $\mathbb{Z}^+$ with the property that $m\alpha \in \mathbb{Z}^+$, then it is clear that, for any F on $[0,1)$, $F \circ \mathbf{\Sigma}_\alpha = F$ if and only if F has period $\frac{1}{m}$. Hence, if $F \in L^2\big([0,1);\mathbb{C}\big)$ and

$$c_\ell(F) \equiv \int_{[0,1)} F(\omega) e^{-\sqrt{-1}\,2\pi\ell\omega}\, d\omega, \quad \ell \in \mathbb{Z},$$

then elementary Fourier analysis leads to the conclusion that, in this case:

$$\lim_{n\to\infty} \mathbf{A}_n F(\omega) = \sum_{\ell\in\mathbb{Z}} c_{m\ell}(F) e^{\sqrt{-1}\,2m\ell\pi\omega} \text{ for Lebesgue-almost every } \omega \in [0,1).$$

On the other hand, if α is irrational, then $\big\{\mathbf{\Sigma}_\alpha^k : k \in \mathbb{N}\big\}$ is μ-ergodic on $[0,1)$. To see this, suppose that $F \in \mathfrak{I}(\mathbb{C})$. Then (cf. the preceding and use Parseval's identity)

$$0 = \big\|F - F \circ \mathbf{\Sigma}_\alpha\big\|^2_{L^2([0,1);\mathbb{C})} = \sum_{\ell\in\mathbb{Z}} \big|c_\ell(F) - c_\ell(F \circ \mathbf{\Sigma}_\alpha)\big|^2.$$

But, clearly,

$$c_\ell(F \circ \mathbf{\Sigma}_\alpha) = e^{\sqrt{-1}\,2\pi\ell\alpha} c_\ell(F), \quad \ell \in \mathbb{Z},$$

and so (because α is irrational) $c_\ell(F) = 0$ for each $\ell \neq 0$. In other words, the only elements of $\mathfrak{I}(\mathbb{C})$ are μ-almost everywhere constant. Thus, for each irrational $\alpha \in (0,1)$, $p \in [1,\infty)$, separable Banach space E, and $F \in L^p\big([0,1);E\big)$:

$$\lim_{n\to\infty} \mathbf{A}_n F = \int_{[0,1)} F(\omega)\, d\omega \text{ Lebesgue-almost everywhere and in } L^p(\mu;E).$$

Finally, notice that the situation changes radically when one moves from $[0,1)$ to $[0,\infty)$ and again takes μ to be Lebesgue measure and $\alpha \in (0,1)$ to be irrational. If I extend the definition of $\mathbf{\Sigma}_\alpha$ by taking $\mathbf{\Sigma}_\alpha(\omega) = [\omega] + \mathbf{\Sigma}_\alpha(\omega - [\omega])$ for $\omega \in [0,\infty)$, then it is clear that invariant functions are those which are constant on each interval $[m, m+1)$ and that, Lebesgue almost surely, $\mathbf{A}_n f(\omega) \longrightarrow \int_{[\omega]}^{[\omega]+1} f(\eta)\, d\eta$. On the other hand, if one defines $\mathbf{\Sigma}_\alpha(\omega) = \omega + \alpha$, then every invariant set which has nonzero measure will have infinite measure, and so, now, every choice of $\alpha \in (0,1)$ (not just irrational ones) will give rise to an ergodic system. In particular, one will have, for each $p \in [1,\infty)$ and $F \in L^p(\mu;E)$,

$$\lim_{n\to\infty} \mathbf{A}_n F = 0 \quad \text{Lebesgue-almost everywhere;}$$

and the convergence will be in $L^p(\mu;E)$ when $p \in (1,\infty)$.

§ **6.2.3. Stationary Sequences.** For applications to probability theory, it is useful to reformulate these considerations in terms of stationary families of random variables. Thus, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(E, \mathcal{B})$ a measurable space (E need not be a Banach space). Given a family $\mathfrak{F} = \{X_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ of E-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, I will say that $\mathfrak{F}$ is **$\mathbb{P}$-stationary** (or simply **stationary**) if, for each $\boldsymbol{\ell} \in \mathbb{N}^N$, the family

$$\mathfrak{F}_{\boldsymbol{\ell}} \equiv \left\{X_{\mathbf{k}+\boldsymbol{\ell}} : \mathbf{k} \in \mathbb{N}^N\right\}$$

has the same (joint) distribution under $\mathbb{P}$ as $\mathfrak{F}$ itself. Clearly, one can test for stationarity by checking that distribution of $\mathfrak{F}_{\mathbf{e}_j}$ is the same as that of $\mathfrak{F}$ for each $1 \le j \le N$. In order to apply the considerations of § 6.2.1 to stationary families, note that all questions about the properties of $\mathfrak{F}$ can be phrased in terms of the following **canonical setting** . Namely, set $\mathbf{E} = E^{\mathbb{N}^N}$ and define μ on $\left(\mathbf{E}, \mathcal{B}^{\mathbb{N}^N}\right)$ to be the image measure $\mathfrak{F}_*\mathbb{P}$. In other words, for each $\Gamma \in \mathcal{B}^{\mathbb{N}^N}$, $\mu(\Gamma) = \mathbb{P}\big(\mathfrak{F} \in \Gamma\big)$. Next, for each $\boldsymbol{\ell} \in \mathbb{N}^N$, define $\boldsymbol{\Sigma}^{\boldsymbol{\ell}} : \mathbf{E} \longrightarrow \mathbf{E}$ to be the natural shift transformation on $\mathbf{E}$ given by $\boldsymbol{\Sigma}^{\boldsymbol{\ell}}(\mathbf{x})_{\mathbf{k}} = x_{\mathbf{k}+\boldsymbol{\ell}}$ for all $\mathbf{k} \in \mathbb{N}^N$. Obviously, *stationarity of $\mathfrak{F}$ is equivalent to the statement that $\{\boldsymbol{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ is μ-measure preserving*. Moreover, if $\mathfrak{I}$ is the σ-algebra of **shift invariant** elements $\Gamma \in \mathcal{B}^{\mathbb{N}^N}$ (i.e., $\Gamma = \left(\boldsymbol{\Sigma}^{\mathbf{k}}\right)^{-1}(\Gamma)$ for all $\mathbf{k} \in \mathbb{N}^N$), then, by Theorem 6.2.8, for any separable Banach space B, any $p \in [1, \infty)$, and any $F \in L^p(\mathbb{P}; B)$:

$$\lim_{n\to\infty} \frac{1}{n^N} \sum_{\mathbf{k} \in Q_n^+} F \circ \mathfrak{F}_{\mathbf{k}} = \mathbb{E}^{\mathbb{P}}\left[F \circ \mathfrak{F} \,\middle|\, \mathfrak{F}^{-1}(\mathfrak{I})\right] \text{ (a.s., } \mathbb{P}) \text{ and in } L^p(\mathbb{P}; B).$$

In particular, when $\left\{\boldsymbol{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$ is ergodic on $\left(\mathbf{E}, \mathcal{B}^{\mathbb{N}^N} \mu\right)$, I will say that the family $\mathfrak{F}$ is **ergodic** and conclude that the preceding can be replaced by

$$\lim_{n\to\infty} \frac{1}{n^N} \sum_{\mathbf{k} \in Q_n^+} F \circ \mathfrak{F}_{\mathbf{k}} = \mathbb{E}^{\mathbb{P}}\big[F \circ \mathfrak{F}\big] \quad \text{(a.s., } \mathbb{P}) \text{ and in } L^p(\mathbb{P}; B). \tag{6.2.10}$$

So far I have discussed *one-sided* stationary families, that is families indexed by $\mathbb{N}^N$. However, for various reasons (cf. Theorem 6.2.12 below) it is useful to know that *one can usually embed a one-sided stationary family into a two-sided one*. In terms of the *semigroup* of shifts, this corresponds to the trivial observation that the semigroup $\left\{\boldsymbol{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$ on $\mathbf{E} = E^{\mathbb{N}^N}$ can be viewed as a sub-semigroup of the *group* of shifts $\left\{\boldsymbol{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^N\right\}$ on $\hat{\mathbf{E}} = E^{\mathbb{Z}^N}$. With these comments in mind, I will prove the following.

Lemma 6.2.11. *Assume that E is a complete, separable, metric space and that $\mathfrak{F} = \{X_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ is a stationary family of E-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ and a family $\hat{\mathfrak{F}} = \left\{\hat{X}_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^N\right\}$ with the property that, for each $\boldsymbol{\ell} \in \mathbb{Z}^N$,*

$$\hat{\mathfrak{F}}_{\boldsymbol{\ell}} \equiv \left\{\hat{X}_{\mathbf{k}+\boldsymbol{\ell}} : \mathbf{k} \in \mathbb{N}^N\right\}$$

has the same distribution under $\hat{\mathbb{P}}$ as $\mathfrak{F}$ has under $\mathbb{P}$.

PROOF: When formulated correctly, this theorem is an essentially trivial application of Kolmogorov's Extension Theorem (cf. part (**iii**) of Exercise 9.1.17). Namely, for $n \in \mathbb{N}$, set

$$\Lambda_n = \big\{\mathbf{k} \in \mathbb{Z}^N : k_j \geq -n \text{ for } 1 \leq j \leq N\big\} \quad \text{and} \quad \mathbf{n} = (n, \dots, n)$$

and define $\Phi_n : E^{\Lambda_0} \longrightarrow E^{\Lambda_n}$ so that

$$\Phi_n(\mathbf{x})_{\mathbf{k}} = x_{\mathbf{n}+\mathbf{k}} \quad \text{for } \mathbf{x} \in E^{\Lambda_0} \text{ and } \mathbf{k} \in \Lambda_n.$$

Next, take μ_0 on E^{Λ_0} to be the $\mathbb{P}$-distribution of $\mathfrak{F}$ and, for $n \geq 1$, μ_n on E^{Λ_n} to be $(\Phi_n)_*\mu_0$. Using stationarity, one can easily check that, for each $n \geq 0$ and $\mathbf{k} \in \mathbb{N}^N$, μ_n is invariant under the obvious extension of $\mathbf{\Sigma}^{\mathbf{k}}$ to E^{Λ_n}. In particular, if one identifies $E^{\Lambda_{n+1}}$ with $E^{\Lambda_{n+1}\setminus\Lambda_n} \times E^{\Lambda_n}$, then

$$\mu_{n+1}\big(E^{\Lambda_{n+1}\setminus\Lambda_n} \times \Gamma\big) = \mu_n(\Gamma) \quad \text{for all } \Gamma \in \mathcal{B}_{E^{\Lambda_n}}.$$

Hence the μ_n's are consistently defined of the spaces E^{Λ_n}, and therefore Kolmogorov's Extension Theorem applies and guarantees the existence of a unique Borel probaility measure μ on $E^{\mathbb{Z}^N}$ with the property that

$$\mu\big(E^{\mathbb{Z}^N\setminus\Lambda_n} \times \Gamma\big) = \mu_n(\Gamma) \quad \text{for all } n \geq 0 \text{ and } \Gamma \in \mathcal{B}_{E^{\Lambda_n}}.$$

Moreover, since each μ_n is $\mathbf{\Sigma}^{\mathbf{k}}$-invariant for all $\mathbf{k} \in \mathbb{N}^N$, it is clear that μ is also. Thus, because $\mathbf{\Sigma}^{\mathbf{k}}$ is invertible on $E^{\mathbb{Z}^N}$ and $\mathbf{\Sigma}^{-\mathbf{k}}$ is its inverse, it follows that μ is invariant under $\mathbf{\Sigma}^{\mathbf{k}}$ for all $\mathbf{k} \in \mathbb{Z}^N$.

To complete the proof at this point, simply take $\hat{\Omega} = E^{\mathbb{Z}^N}$, $\hat{\mathcal{F}} = \mathcal{B}_{\hat{\Omega}}$, $\hat{\mathbb{P}} = \mu$, and $\hat{X}_{\mathbf{k}}(\hat{\omega}) = \hat{\omega}_{\mathbf{k}}$ for $\mathbf{k} \in \mathbb{Z}^N$. $\square$

As an example of the advantage which Lemma 6.2.11 affords, I present the following beautiful observation made by M. Kac.

THEOREM 6.2.12. *Let $(E, \mathcal{B})$ be a measurable space and $\{X_k : k \in \mathbb{N}\}$ a stationary sequence of E-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Given $\Gamma \in \mathcal{B}$, define the return time $\rho_\Gamma(\omega) = \inf\{k \geq 1 : X_k(\omega) \in \Gamma\}$. Then, $\mathbb{E}^{\mathbb{P}}\big[\rho_\Gamma,\, X_0 \in \Gamma\big] = \mathbb{P}\big(X_k \in \Gamma$ for some $k \in \mathbb{N}\big)$. In particular, if $\{X_k : k \in \mathbb{N}\}$ is ergodic, then*

$$\mathbb{P}\big(X_0 \in \Gamma\big) > 0 \implies \mathbb{E}^{\mathbb{P}}\big[\rho_\Gamma,\, X_0 \in \Gamma\big] = 1.$$

PROOF: Set $U_k = \mathbf{1}_\Gamma \circ X_k$ for $k \in \mathbb{N}$. Then $\{U_k : k \in \mathbb{N}\}$ is a stationary sequence of $\{0,1\}$-valued random variables. Hence, by Lemma 6.2.11, we can find a probability space $\big(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}\big)$ on which there is a family $\{\hat{U}_k : k \in \mathbb{Z}\}$ of $\{0,1\}$-valued random variables with the property that, for every $n \in \mathbb{Z}$, $\big(\hat{U}_n, \dots, \hat{U}_{n+k}, \dots\big)$ has the same distribution under $\hat{\mathbb{P}}$ as $(U_0, \dots, U_k, \dots)$ has under $\mathbb{P}$. In particular,

$$P\big(\rho_\Gamma \geq 1,\, X_0 \in \Gamma\big) = \hat{\mathbb{P}}\big(\hat{U}_0 = 1\big)$$
$$P\big(\rho_\Gamma \geq n+1,\, X_0 \in \Gamma\big) = \hat{\mathbb{P}}\big(\hat{U}_{-n} = 1, \hat{U}_{-n+1} = 0, \dots, \hat{U}_0 = 0\big), \quad n \in \mathbb{Z}^+.$$

Thus, if

$$\lambda_\Gamma(\hat{\omega}) \equiv \inf\big\{k \in \mathbb{N} : U_{-k}(\hat{\omega}) = 1\big\},$$

then

$$\mathbb{P}\big(\rho_\Gamma \geq n,\, X_0 \in \Gamma\big) = \hat{\mathbb{P}}\big(\lambda_\Gamma = n-1\big), \quad n \in \mathbb{Z}^+;$$

and so

$$\mathbb{E}^{\mathbb{P}}\big[\rho_\Gamma,\, X_0 \in \Gamma\big] = \hat{\mathbb{P}}\big(\lambda_\Gamma < \infty\big).$$

Now observe that

$$\hat{\mathbb{P}}\big(\lambda_\Gamma > n\big) = \hat{\mathbb{P}}\big(\hat{U}_{-n} = 0, \dots, \hat{U}_0 = 0\big) = \mathbb{P}\big(X_0 \notin \Gamma, \dots, X_n \notin \Gamma\big),$$

from which it is clear that

$$\hat{\mathbb{P}}\big(\lambda_\Gamma < \infty\big) = \mathbb{P}\big(\exists k \in \mathbb{N}\ X_k \in \Gamma\big).$$

Finally, assume that $\{X_k : k \in \mathbb{N}\}$ is ergodic and that $\mathbb{P}(X_0 \in \Gamma) > 0$. Because, by (6.2.10), $\sum_0^\infty \mathbf{1}_\Gamma\big(X_k\big) = \infty$ $\mathbb{P}$-almost surely, it follows that, $\mathbb{P}$-almost surely, $X_k \in \Gamma$ for some $k \in \mathbb{N}$. $\square$

When $\{X_n : n \geq 0\}$ is an irreducible, ergodic Markov chain on a countable state space E, then Kac's theorem proves that the stationary measure at the state $x \in E$ is the reciprocal of the expected time that the chain takes to return to x when it starts at x.

§ **6.2.4. Continuous Parameter Ergodic Theory.** I turn now to the setting of continuously parametrized semigroups of transformations. Thus, again $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space and $\big\{\mathbf{\Sigma}^{\mathbf{t}} : \mathbf{t} \in [0,\infty)^N\big\}$ is a measurable semigroup of μ-measure preserving transformations on Ω. That is, $\mathbf{\Sigma}^{\mathbf{0}}$ is the identity, $\mathbf{\Sigma}^{\mathbf{s}+\mathbf{t}} = \mathbf{\Sigma}^{\mathbf{s}} \circ \mathbf{\Sigma}^{\mathbf{t}}$,

$$(\mathbf{t}, \omega) \in [0,\infty)^N \times \Omega \longmapsto \mathbf{\Sigma}^{\mathbf{t}}(\omega) \in \Omega \quad \text{is } \mathcal{B}_{[0,\infty)^N} \times \mathcal{F}\text{-measurable},$$

and $\big(\mathbf{\Sigma}^{\mathbf{t}}\big)_* \mu = \mu$ for every $\mathbf{t} \in [0,\infty)^N$. Next, given an $\mathcal{F}$-measurable F with values in some separable Banach space E, let $\mathfrak{G}(F)$ be the set of $\omega \in \Omega$ with the property that

$$\int_{[0,T)^N} \big\|F \circ \mathbf{\Sigma}^{\mathbf{t}}(\omega)\big\|_E \, d\mathbf{t} < \infty \quad \text{for all } T \in (0,\infty).$$

Clearly,

$$\omega \in \mathfrak{G}(F) \implies \mathbf{\Sigma}^{\mathbf{t}}(\omega) \in \mathfrak{G}(F) \quad \text{for every } \mathbf{t} \in [0,\infty)^N.$$

In addition, if $F \in L^p(\mu;E)$ for some $p \in [1,\infty)$, then

$$\int_\Omega \left(\int_{[0,T)^N} \left\| F \circ \mathbf{\Sigma}^{\mathbf{t}}(\omega) \right\|_E^p \, d\mathbf{t} \right) \mu(d\omega) = T^N \|F\|_{L^p(\mu;E)}^p < \infty,$$

and so

$$F \in \bigcup_{p \in [1,\infty)} L^p(\mu;E) \implies \mu\Big(\mathfrak{G}(F)\complement \Big) = 0.$$

Next, for each $T \in (0,\infty)$, define

$$\mathcal{A}_T F(\omega) = \begin{cases} T^{-N} \int_{[0,T)^N} F \circ \mathbf{\Sigma}^{\mathbf{t}}(\omega)\, d\mathbf{t} & \text{if } \omega \in \mathfrak{G}(F) \\ 0 & \text{if } \omega \notin \mathfrak{G}(F). \end{cases}$$

Note that, as a consequence of the invariance of $\mathfrak{G}(F)$,

$$\big(\mathcal{A}_T F\big) \circ \mathbf{\Sigma}^{\mathbf{t}} = \mathcal{A}_T\big(F \circ \mathbf{\Sigma}^{\mathbf{t}}\big) \quad \text{for all } \mathbf{t} \in [0,\infty)^N.$$

Finally, use $\hat{\mathfrak{I}}$ to denote the σ-algebra of $\Gamma \in \mathcal{F}$ with the property that $\Gamma = (\mathbf{\Sigma}^{\mathbf{t}})^{-1}(\Gamma)$ for each $t \in [0,\infty)^N$, and say that $\{\mathbf{\Sigma}^{\mathbf{t}} : \mathbf{t} \in [0,\infty)^N\}$ is **ergodic** if $\mu(\Gamma) \wedge \mu(\Gamma\complement) = 0$ for every $\Gamma \in \hat{\mathfrak{I}}$.

Theorem 6.2.13. *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space and $\{\mathbf{\Sigma}^{\mathbf{t}} : \mathbf{t} \in [0,\infty)^N\}$ be a measurable semigroup of μ-measure preserving transformations on Ω. Then, for each separable Banach space E, $p \in [1,\infty)$, and $T \in (0,\infty)$, $\mathcal{A}_T$ is a contraction on $L^p(\mu;E)$. Next, set $\Pi_{\hat{\mathfrak{I}}(E)} = \Pi_{\mathfrak{I}(E)} \circ \mathcal{A}_1$, where $\Pi_{\mathfrak{I}(E)}$ is defined in terms of $\{\mathbf{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\}$ as in Theorem 6.2.8. Then, for each $p \in [1,\infty)$ and $F \in L^p(\mu;E)$:*

$$\lim_{T\to\infty} \mathcal{A}_T F = \Pi_{\hat{\mathfrak{I}}(E)} F \quad \text{(a.e., } \mu). \tag{6.2.14}$$

Moreover, if $p \in (1,\infty)$ or $p = 1$ and $\mu(\Omega) < \infty$, then the convergence is also in $L^p(\mu;E)$. In fact, if $\mu(\Omega) < \infty$, then

$$\lim_{T\to\infty} \mathcal{A}_T F = \mathbb{E}^\mu\big[F \,\big|\, \hat{\mathfrak{I}}\big] \quad \text{(a.e., } \mu) \text{ and in } L^p(\mu : E).$$

Finally, if $\{\mathbf{\Sigma}^{\mathbf{t}} : \mathbf{t} \in [0,\infty)^N\}$ is ergodic, then (6.2.14) can be replaced by

$$\lim_{T\to\infty} \mathcal{A}_T F = \frac{\mathbb{E}^\mu[F]}{\mu(\Omega)} \quad \text{(a.e., } \mu),$$

where it is understood that the ratio is 0 when the denominator is infinite.

PROOF: The first step is the observation that

$$\mu\left(\sup_{T>0}\big\|\mathcal{A}_T F\big\|_E \ge \lambda\right) \le \frac{(24)^N}{\lambda}\|F\|_{L^1(\mu;E)}, \quad \lambda\in(0,\infty), \tag{6.2.15}$$

and

$$\Big\|\sup_{T>0}\big\|\mathcal{A}_T F\big\|_E\Big\|_{L^p(\mu;E)} \le \frac{(24)^N p}{p-1}\|F\|_{L^p(\mu;E)} \quad \text{for } p\in(1,\infty). \tag{6.2.16}$$

Indeed, because of $(\mathcal{A}_T F)\circ\mathbf{\Sigma}^t = \mathcal{A}_T(F\circ\mathbf{\Sigma}^t)$, (6.2.15) is derived from (6.1.6) in precisely the same way as I derived (6.2.3) from (6.1.10), and (6.2.16) comes from (6.1.7) just as (6.2.4) came from (6.1.7).

Given (6.2.15) and (6.2.16), we know that it suffices to prove (6.2.14) in the case when F is a uniformly bounded element of $L^1(\mu;E)$. But in that case, set $\hat{F} = \mathcal{A}_1 F$ and observe that

$$\big\|T^N\mathcal{A}_T F(\omega) - n^N\mathbf{A}_n\hat{F}(\omega)\big\|_E \le \int\limits_{[0,n+1)^N\setminus[0,n)^N} \big\|F\circ\mathbf{\Sigma}^{\mathbf{t}}(\omega)\big\|_E\,d\mathbf{t}$$

for $n\le T\le n+1$, and conclude that

$$\lim_{n\to\infty}\Big\|\sup_{n\le T\le n+1}\big\|\mathcal{A}_T F - \mathbf{A}_n\hat{F}\big\|_E\Big\|_{L^p(\mu;E)} = 0 \quad \text{for every } p\in[1,\infty].$$

Hence, (6.2.14) follows from (6.2.9). As for case when $\mu(\Omega)<\infty$, all that we have to do is check that $\Pi_{\hat{\mathfrak{I}}(E)}F = \mathbb{E}^\mu\big[F\big|\hat{\mathfrak{I}}\big]$ (a.e., μ). However, from (6.2.14), it is easy to see that $\Pi_{\hat{\mathfrak{I}}(E)}F$ is measurable with respect to the μ-completion of $\hat{\mathfrak{I}}$; and so it suffices to show that

$$\mathbb{E}^\mu\big[F,\,\Gamma\big] = \mathbb{E}^\mu\big[\mathcal{A}_1F,\,\Gamma\big] \quad \text{for all } \Gamma\in\hat{\mathfrak{I}}.$$

But, if $\Gamma\in\hat{\mathfrak{I}}$, then

$$\begin{aligned}\mathbb{E}^\mu\big[\mathcal{A}_1F,\,\Gamma\big] &= \int_{[0,1)^N}\mathbb{E}^\mu\big[F\circ\mathbf{\Sigma}^{\mathbf{t}},\,\Gamma\big]\,d\mathbf{t}\\ &= \int_{[0,1)^N}\mathbb{E}^\mu\big[F\circ\mathbf{\Sigma}^{\mathbf{t}},\,\big(\mathbf{\Sigma}^{\mathbf{t}}\big)^{-1}(\Gamma)\big]\,d\mathbf{t} = \mathbb{E}^\mu[F,\,\Gamma].\end{aligned}$$

Finally, assume that $\big\{\mathbf{\Sigma}^{\mathbf{t}} : \mathbf{t}\in[0,\infty)^N\big\}$ is μ-ergodic. When $\mu(\Omega)<\infty$, the asserted result follows immediately from the preceding; and when $\mu(\Omega)=\infty$, it follows from the fact that $\Pi_{\hat{\mathfrak{I}}(E)}F$ is measurable with respect to the μ-completion of $\hat{\mathfrak{I}}$. $\square$

Exercises for § 6.2

EXERCISE 6.2.17. Given an irrational $\alpha \in (0,1)$ and an $\epsilon \in (0,1)$, let $N_n(\alpha,\epsilon)$ be the number of $1 \le m \le n$ with the property that

$$\left|\alpha - \frac{\ell}{m}\right| \le \frac{\epsilon}{2m} \quad \text{for some } \ell \in \mathbb{Z}.$$

As an application of the considerations in the Classic Example given at the end of § 6.1, show that

$$\varliminf_{n\to\infty} \frac{N_n(\alpha,\epsilon)}{n} \ge \epsilon.$$

Hint: Let $\delta \in \left(0, \frac{\epsilon}{2}\right)$ be given, take f equal to the indicator function of $[0,\delta) \cup (1-\delta, 1)$, and observe that $N_n(\alpha,\epsilon) \ge \sum_{k=1}^n f \circ \mathbf{\Sigma}_\alpha^k(\omega)$ so long as $0 \le \omega \le \frac{\epsilon}{2} - \delta$.

EXERCISE 6.2.18. Assume that $\mu(\Omega) < \infty$ and that $\left\{\mathbf{\Sigma}^{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$ is ergodic. Given a non-negative $\mathcal{F}$-measurable function f, show that

$$\begin{aligned}\varlimsup_{n\to\infty} \mathbf{A}_n f < \infty \text{ on a set of positive } \mu\text{-measure} \implies f \in L^1(\mu;\mathbb{R}) \\ \implies \lim_{n\to\infty} \mathbf{A}_n f = \frac{\mathbb{E}^\mu[f]}{\mu(\Omega)} \quad (\text{a.e., } \mu).\end{aligned}$$

EXERCISE 6.2.19. Let $\mathfrak{F} = \left\{X_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$ be a stationary family of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in the measurable space $(E, \mathcal{B})$; and let $\mathfrak{I}$ denote the σ-algebra of shift invariant $\Gamma \in \mathcal{B}_E^{\mathbb{N}^N}$.

(**i**) Take

$$\mathcal{T} \equiv \bigcap_{n \ge 0} \sigma\big(X_{\mathbf{k}} : k_j \ge n \text{ for all } 1 \le j \le N\big),$$

the tail σ-algebra determined by $\left\{X_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$. Show that $\mathfrak{F}^{-1}(\mathfrak{I}) \subseteq \mathcal{T}$, and conclude that $\left\{X_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}^N\right\}$ is ergodic if $\mathcal{T}$ is $\mathbb{P}$-trivial (i.e., $\mathbb{P}(\Gamma) \in \{0,1\}$ for all $\Gamma \in \mathcal{T}$.)

(**ii**) By combining (**i**), Kolmogorov's 0–1 Law, and the Individual Ergodic Theorem, give another derivation of the Strong Law of Large Numbers for independent, identically distributed, integrable random variables with values in a separable Banach space.

EXERCISE 6.2.20. Let $\left\{X_k : k \in \mathbb{N}\right\}$ be a stationary, ergodic sequence of $\mathbb{R}$-valued, integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Using the reasoning suggested in Exercise 1.4.28, prove Guivarc'h's lemma:

$$\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0 \implies \varliminf_{n\to\infty} \left|\sum_{k=0}^{n-1} X_k\right| < \infty.$$

§ 6.3 Burkholder's Inequality

Given a martingale $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ with $X_0 = 0$ and a sequence $\{\sigma_n : n \geq 0\}$ of bounded functions with the property that σ_n is $\mathcal{F}_n$-measurable for $n \geq 0$, determine $\{Y_n : n \geq 0\}$ by $Y_0 = 0$ and $Y_n - Y_{n-1} = \sigma_{n-1}(X_n - X_{n-1})$ for $n \geq 1$. It is clear that $\big(Y_n, \mathcal{F}_n, \mathbb{P}\big)$ is again a martingale. In addition, if the absolute value of all the σ_n's are bounded by some constant $\sigma < \infty$ and X_n is square $\mathbb{P}$-integrable, then one can easily check that

$$\mathbb{E}^{\mathbb{P}}\big[Y_n^2\big] = \sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\big[\sigma_n^2(X_n - X_{n-1})^2\big] \leq \sigma^2 \sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\big[(X_n - X_{n-1})^2\big] = \sigma^2 \mathbb{E}^{\mathbb{P}}\big[X_n^2\big].$$

On the other hand, it is not at all clear how to compare the size of Y_n to that of X_n in any of the L^p spaces other than $p = 2$.

The problem of finding such a comparison was given a definitive solution by D. Burkholder, and I will present his solution in this section. Actually, Burkholder solved the problem twice. His first solution was a beautiful adaptation of general ideas and results which had been developed over the years to solve related problems in probability theory and analysis, and, as such, did not yield the optimal solution. His second approach is designed specifically to address the problem at hand and bears little or no resemblance to familiar techniques. It is entirely original, remarkably elementary and effective, but somewhat opaque. The approach is the outgrowth of many years of deep thinking which Burkholder devoted to the topic, and the reader who wants to understand the path which led him to it should consult the explanation which he wrote.†

§ 6.3.1. Burkholder's Comparison Theorem. Burkholder's basic result is the following comparison theorem.

Theorem 6.3.1 (**Burkholder**). *Let $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ be a probability space, $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ a non-decreasing sequence of sub-σ-algebras of $\mathcal{F}$, and E and F a pair of (real or complex) separable Hilbert spaces. Next, suppose that $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ and $\big(Y_n, \mathcal{F}_n, \mathbb{P}\big)$ are, respectively, E- and F-valued martingales. If*

$$\|Y_0\|_F \leq \|X_0\|_E \text{ and } \|Y_n - Y_{n-1}\|_F \leq \|X_n - X_{n-1}\|_E,\ n \in \mathbb{Z}^+,$$

P-almost surely, then, for each $p \in (1, \infty)$ and $n \in \mathbb{N}$,

$$\big\|Y_n\big\|_{L^p(\mathbb{P};F)} \leq B_p \big\|X_n\big\|_{L^p(\mathbb{P};E)} \quad \text{where } B_p \equiv (p-1) \vee \frac{1}{p-1}. \tag{6.3.2}$$

As I said before, the derivation of Theorem 6.3.1 is both elementary and mysterious. I begin with the trivial observation that, without loss in generality,

† For those who want to know the secret behind this proof, Burkholder has revealed it in his article "Explorations in martingale theory and its applications" for the 1989 Saint-Flour Ecole d'Eté lectures published by Springer–Verlag, LNM **1464** (1991).

I may assume that both E and F are complex Hilbert spaces, since we can always complexify them, and, in addition, that $E = F$, since, if that is not already the case, we can embed them in $E \oplus F$. Thus, I will be making these assumptions throughout.

The heart of the proof lies in the computations contained in the following two lemmas.

LEMMA 6.3.3. *Let $p \in (1,\infty)$ be given, set*

$$\alpha_p = \begin{cases} p^{2-p}\,(p-1)^{p-1} & \text{if} \quad p \in [2,\infty) \\ p^{2-p} & \text{if} \quad p \in (1,2], \end{cases}$$

and define $u : E^2 \longrightarrow \mathbb{R}$ by (cf. (6.3.2))

$$u(x,y) = \big(\|y\|_E - B_p\|x\|_E\big)\big(\|y\|_E + \|x\|_E\big)^{p-1}.$$

Then

$$\|y\|_E^p - \big(B_p\|x\|_E\big)^p \le \alpha_p\, u(x,y), \quad (x,y) \in E^2.$$

PROOF: When $p = 2$, there is nothing to do. Thus, I will assume that $p \in (1,\infty) \setminus \{2\}$.

Observe that it suffices to show that for all $(x,y) \in E^2$ satisfying $\|x\|_E + \|y\|_E = 1$:

$$(*) \qquad \|y\|_E^p - \big((p-1)\|x\|_E\big)^p \begin{cases} \le p^{2-p}\,(p-1)^{p-1}\big(\|y\|_E - (p-1)\|x\|_E\big) \\ \ge p^{2-p}\,(p-1)^{p-1}\big(\|y\|_E - (p-1)\|x\|_E\big). \end{cases}$$

depending on whether $p \in (2,\infty)$ or $p \in (1,2)$. Indeed, when $p \in (2,\infty)$, (*) is precisely the result desired, and, when $p \in (1,2)$, (*) gives the desired result after one divides through by $(p-1)^p$ and reverses the roles of x and y.

I begin the verification of (*) by checking that

$$(**) \qquad p^{2-p}\,(p-1)^{p-1} \begin{cases} > 1 & \text{if} \quad p \in (2,\infty) \\ < 1 & \text{if} \quad p \in (1,2). \end{cases}$$

To this end, set $f(p) = (p-1)\log(p-1) - (p-2)\log p$ for $p \in (1,\infty)$. Then f is strictly convex on $(1,2)$ and strictly concave on $(2,\infty)$. Thus, $f \restriction (1,2)$ cannot achieve a maximum and therefore, since $\lim_{p\searrow 1} f(p) = 0 = f(2)$, $f < 0$ on $(1,2)$. Similarly, $f \restriction (2,\infty)$ cannot achieve a minimum, and therefore, since $f(2) = 0$ while $\lim_{p\nearrow\infty} f(p) = \infty$, we have that $f > 0$ on $(2,\infty)$.

Next, observe that proving (*) comes down to checking that

$$\Phi(s) \equiv p^{2-p}\,(p-1)^{p-1}\,(1-ps) - (1-s)^p + (p-1)^p\, s^p \begin{cases} \ge 0 & \text{if } p \in (2,\infty) \\ \le 0 & \text{if } p \in (1,2), \end{cases}$$

for $s \in [0,1]$. To this end, note that, by (**), $\Phi(0) > 0$ when $p \in (2,\infty)$ and $\Phi(0) < 0$ when $p \in (1,2)$. Also, for $s \in (0,1)$,

$$\Phi'(s) = p\Big[(p-1)^p\, s^{p-1} + (1-s)^{p-1} - p^{2-p}\,(p-1)^{p-1}\Big]$$

and

$$\Phi''(s) = p(p-1)\Big[(p-1)^p\, s^{p-2} - (1-s)^{p-2}\Big].$$

In particular, we see that $\Phi\big(\frac{1}{p}\big) = \Phi'\big(\frac{1}{p}\big) = 0$. In addition, depending on whether $p \in (2,\infty)$ or $p \in (1,2)$: $\lim_{s\searrow 0}\Phi''(s)$ is negative or positive, Φ'' is strictly increasing or decreasing on $(0,1)$, and $\lim_{s\nearrow 1}\Phi''(1)$ is positive or negative. Hence, there exists a unique $t = t_p \in (0,1)$ with the property that

$$\Phi'' \restriction (0,t)\begin{cases} < 0 & \text{if } p \in (2,\infty) \\ > 0 & \text{if } p \in (1,2) \end{cases} \quad \text{and} \quad \Phi'' \restriction (t,1)\begin{cases} > 0 & \text{if } p \in (2,\infty) \\ < 0 & \text{if } p \in (1,2) \end{cases}.$$

Moreover, from the equation $\Phi''(t) = 0$, it is easy to see that $t \in \Big(0, \frac{1}{p}\Big)$.

Now suppose that $p \in (2,\infty)$ and consider Φ on each of the intervals $\big[\frac{1}{p},1\big]$, $\big[t,\frac{1}{p}\big]$, and $\big[0,t\big]$ separately. Because both Φ and Φ' vanish at $\frac{1}{p}$ while $\Phi'' > 0$ on $\big(\frac{1}{p},1\big)$, it is clear that $\Phi > 0$ on $\big(\frac{1}{p},1\big]$. Next, because $\Phi'\big(\frac{1}{p}\big) = 0$ and $\Phi'' \restriction \big(t,\frac{1}{p}\big) > 0$, we know that Φ is strictly decreasing on $\big(t,\frac{1}{p}\big)$ and therefore that $\Phi \restriction \big[t,\frac{1}{p}\big) > \Phi\big(\frac{1}{p}\big) = 0$. Finally, because $\Phi'' \restriction (0,t) < 0$ while $\Phi(0) \wedge \Phi(t) \geq 0$, we also know that $\Phi \restriction (0,t) > 0$. The argument when $p \in (1,2)$ is similar; only this time all the signs are reversed. $\square$

Lemma 6.3.4. *Again let $p \in (1,\infty)$ be given, and define $u : E \times F \longrightarrow \mathbb{R}$ as in Lemma 6.3.3. In addition, define the functions v and w on $E^2 \setminus \{0,0\}$ by*

$$v(x,y) = p\big(\|y\|_E + \|x\|_E\big)^{p-2}\big(\|y\|_E + (2-p)\|x\|_E\big)$$

and

$$w(x,y) = p(1-p)\big(\|y\|_E + \|x\|_E\big)^{p-2}\|x\|_E.$$

Then, for $(x,y) \in E^2$ and $(k,h) \in E^2$ satisfying

$$\min_{t\in[0,1]}\big(\|y+th\|_E \wedge \|x+tk\|_E\big) > 0 \quad \text{and} \quad \|h\|_E \leq \|k\|_E,$$

one has

$$u(x+k,y+h) - u(x,y) \leq v(x,y)\,\mathfrak{Re}\left(\tfrac{y}{\|y\|_F},h\right)_F + w(x,y)\,\mathfrak{Re}\left(\tfrac{x}{\|x\|_E},k\right)_E$$

when $p \in [2,\infty)$ and

$$(p-1)\big[u(x+k,y+h) - u(x,y)\big] \leq -w(y,x)\,\mathfrak{Re}\left(\tfrac{y}{\|y\|_E},h\right)_E - v(y,x)\,\mathfrak{Re}\left(\tfrac{x}{\|x\|_E},k\right)_E$$

when $p \in (1,2]$.

PROOF: Set

$$\begin{aligned}\Phi(t) &= \Phi\big(t;(x,k),(y,h)\big)\\ &\equiv \big(\|y+th\|_E-(p-1)\|x+tk\|_E\big)\big(\|x+tk\|_E+\|y+th\|_E\big)^{p-1},\end{aligned}$$

and observe that

$$u\big(x+tk,y+th\big)=\begin{cases}\Phi\big(t;(x,k),(y,h)\big) & \text{if} \quad p\in[2,\infty)\\ -(p-1)^{-1}\,\Phi\big(t;(y,h),(x,k)\big) & \text{if} \quad p\in(1,2).\end{cases}$$

Hence, it suffices for us to check that

$$\Phi'(t)=v(x+tk,y+th)\mathfrak{Re}\left(\tfrac{y+th}{\|y+th\|_E},h\right)_E+w(x+tk,y+th)\mathfrak{Re}\left(\tfrac{x+tk}{\|x+tk\|_E},k\right)_E$$

and prove that

$$\Phi''\big(t;(x,k),(y,h)\big)\begin{cases}\le 0 & \text{if } p\in[2,\infty) \text{ and } \|h\|_E\le\|k\|_E\\ \ge 0 & \text{if } p\in(1,2] \quad\text{and } \|h\|_E\ge\|k\|_E.\end{cases}$$

To prove the preceding, set $x(t)=x+tk$, $y(t)=y+th$, $\Psi(t)=\|x(t)\|_E+\|y(t)\|_E$, $a(t)=\frac{\mathfrak{Re}\left(x(t),k\right)_E}{\|x(t)\|_E}$, and $b(t)=\frac{\mathfrak{Re}\left(y(t),h\right)_E}{\|y(t)\|_E}$. One then has that

$$\begin{aligned}\Phi'(t)&=p\Psi(t)^{p-2}\left[(1-p)\|x(t)\|_E\,a(t)+\Big(\|y(t)\|_E+(2-p)\|x(t)\|_E\Big)\,b(t)\right]\\ &=p\Big[(1-p)\Psi(t)^{p-2}\,\|x(t)\|_E\,\big(a(t)+b(t)\big)+\Psi(t)^{p-1}\,b(t)\Big].\end{aligned}$$

In particular, the first expression establishes the required form for $\Phi'(t)$. In addition, from the second expression, we see that

$$\begin{aligned}-\frac{\Phi''(t)}{p}&=(p-1)(p-2)\,\Psi(t)^{p-3}\,\|x(t)\|_E\,\big(a(t)+b(t)\big)^2\\ &\qquad+(p-1)\Psi(t)^{p-2}\left[a(t)\big(a(t)+b(t)\big)+\tfrac{\|x(t)\|_E}{\|y(t)\|_E}\,b_\perp(t)^2+a_\perp(t)^2\right]\\ &\qquad\quad-\Psi(t)^{p-2}\left[(p-1)\big(a(t)+b(t)\big)\,b(t)+\Psi(t)\,\tfrac{b_\perp(t)^2}{\|y(t)\|_E}\right]\\ &=(p-1)(p-2)\,\Psi(t)^{p-3}\,\|x(t)\|_E\,\big(a(t)+b(t)\big)^2\\ &\qquad+(p-1)\Psi(t)^{p-2}\big(\|k\|_E^2-\|h\|_E^2\big)+(p-2)\Psi(t)^{p-1}\,\tfrac{b_\perp(t)^2}{\|y(t)\|_E},\end{aligned}$$

where $a_\perp(t)=\sqrt{\|k\|_E^2-a(t)^2}$ and $b_\perp(t)=\sqrt{\|h\|_E^2-b(t)^2}$. Hence the required properties of $\Phi''(t)$ have also been established. □

PROOF OF THEOREM 6.3.1: Set $K_n = X_n - X_{n-1}$ and $H_n = Y_n - Y_{n-1}$ for $n \in \mathbb{Z}^+$. I will assume that there is an $\epsilon > 0$ with the property that

$$\big\|X_0(\omega) - \operatorname{span}\{K_n(\omega) : n \in \mathbb{Z}^+\}\big\|_E \geq \epsilon$$

and

$$\big\|Y_0(\omega) - \operatorname{span}\{H_n(\omega) : n \in \mathbb{Z}^+\}\big\|_E \geq \epsilon$$

for all $\omega \in \Omega$. Indeed, if this is not already the case, then we can replace E by $\mathbb{R} \times E$ (or, when E is complex, $\mathbb{C} \times E$) and $X_n(\omega)$ and $Y_n(\omega)$, respectively, by

$$X_n^{(\epsilon)}(\omega) \equiv \big(\epsilon, X_n(\omega)\big) \quad \text{and} \quad Y_n^{(\epsilon)}(\omega) \equiv \big(\epsilon, Y_n(\omega)\big),$$

for each $n \in \mathbb{N}$. Clearly, (6.3.2) for each $X_n^{(\epsilon)}$ and $Y_n^{(\epsilon)}$ implies (6.3.2) for X_n and Y_n after one lets $\epsilon \searrow 0$. Finally, because there is nothing to do when the right-hand side of (6.3.2) is infinite, let $p \in (1, \infty)$ be given, and assume that $X_n \in L^p(\mathbb{P}; E)$ for each $n \in \mathbb{N}$. In particular, if u is the function defined in Lemma 6.3.3 and v and w are those defined in Lemma 6.3.4, then

$$u(X_n, Y_n) \in L^1(\mathbb{P}; \mathbb{R}) \quad \text{and} \quad v(X_n, Y_n),\, w(X_n, Y_n) \in L^{p'}(\mathbb{P}; \mathbb{R})$$

for all $n \in \mathbb{N}$, where $p' = \frac{p}{p-1}$ is the Hölder conjugate of p.

Note that, by Lemma 6.3.3, it suffices for us to show that $A_n \equiv \mathbb{E}^{\mathbb{P}}\big[u\big(X_n, Y_n\big)\big] \leq 0$, $n \in \mathbb{N}$. Since $u\big(X_0, Y_0\big) \leq 0$ $\mathbb{P}$-almost surely, there is no question that $A_0 \leq 0$. Next, assume that $A_n \leq 0$, and, depending on whether $p \in [2, \infty)$ or $p \in (1, 2]$, use the appropriate part of Lemma 6.3.4 to see that

$$\begin{aligned} A_{n+1} \leq & \mathbb{E}^{\mathbb{P}}\left[v(X_n, Y_n)\mathfrak{Re}\left(\tfrac{Y_n}{\|Y_n\|_E}, H_{n+1}\right)_E\right] \\ & + \mathbb{E}^{\mathbb{P}}\left[w(X_n, Y_n)\mathfrak{Re}\left(\tfrac{X_n}{\|X_n\|_E}, K_{n+1}\right)_E\right] \end{aligned}$$

or

$$\begin{aligned} A_{n+1} \leq & -\mathbb{E}^{\mathbb{P}}\left[w(Y_n, X_n)\mathfrak{Re}\left(\tfrac{Y_n}{\|Y_n\|_E}, H_{n+1}\right)_E\right] \\ & - \mathbb{E}^{\mathbb{P}}\left[v(Y_n, X_n)\mathfrak{Re}\left(\tfrac{X_n}{\|X_n\|_E}, K_{n+1}\right)_E\right]. \end{aligned}$$

But, since $v(X_n, Y_n)\frac{Y_n}{\|Y_n\|_E}$ is $\mathcal{F}_n$-measurable, $\mathbb{E}^{\mathbb{P}}[H_{n+1}|\mathcal{F}_n] = 0$, and therefore (cf. Exercise 5.1.18)

$$\mathbb{E}^{\mathbb{P}}\left[v(X_n, Y_n)\mathfrak{Re}\left(\tfrac{Y_n}{\|Y_n\|_E}, H_{n+1}\right)_E\right] = 0.$$

Since the same reasoning shows that each of the other terms on the right-hand side vanishes, we have now proved that $A_{n+1} \leq 0$. □

As an immediate consequence of Theorem (6.3.2), we have the following answer to the question raised at the beginning of this section.

Corollary 6.3.5. *Suppose that $(X_n, \mathcal{F}_n, \mathbb{P})$ is a martingale with values in a separable (real or complex) Hilbert space E. Further, let F be a second separable, complex Hilbert space, and suppose that $\{\boldsymbol{\sigma}_n : n \geq 0\}$ is a sequence of $\mathrm{Hom}(E; F)$-valued random variables with the properties that $\boldsymbol{\sigma}_0$ is constant, $\boldsymbol{\sigma}_n$ is $\mathcal{F}_n$-measurable for $n \geq 1$, and $\|\boldsymbol{\sigma}_n\|_{\mathrm{op}} \leq \sigma < \infty$ (a.s., $\mathbb{P}$) for some constant $\sigma < \infty$ and all $n \in \mathbb{N}$. If $\|Y_0\|_F \leq \sigma \|X_0\|_E$ and $Y_n - Y_{n-1} = \boldsymbol{\sigma}_{n-1}(X_n - X_{n-1})$ for $n \geq 1$, then $(Y_n, \mathcal{F}_n, \mathbb{P})$ is an F-valued martingale and, for each $p \in (1, \infty)$, (cf. (6.3.2))*

$$\|Y_n\|_{L^p(\mathbb{P};F)} \leq \sigma B_p \|X_n\|_{L^p(\mathbb{P};E)}, \quad n \in \mathbb{N}.$$

§ **6.3.2. Burkholder's Inequality.** In many applications, the most useful form of Burkholder's result is as a generalization to $p \neq 2$ of the obvious equality

$$\mathbb{E}^{\mathbb{P}}\big[|X_n - X_0|^2\big] = \mathbb{E}^{\mathbb{P}}\left[\sum_{m=1}^{n} |X_m - X_{m-1}|^2\right].$$

This is the form of his inequality which is best known, as such, is called **Burkholder's inequality**. Notice that his inequality can be viewed as a vast generalization of Khinchine's Inequality (2.3.27), although it applies only when $p \in (1, \infty)$.

Theorem 6.3.6 (**Burkholder's Inequality**). *Let $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$ and $\big\{\mathcal{F}_n : n \in \mathbb{N}\big\}$ be as in Theorem 6.3.1, and let $\big(X_n, \mathcal{F}_n, \mathbb{P}\big)$ be a martingale with values in the separable Hilbert space E. Then, for each $p \in (1, \infty)$,*

$$\begin{aligned} &\frac{1}{B_p} \sup_{n \in \mathbb{N}} \big\|X_n - X_0\big\|_{L^p(\mathbb{P};E)} \\ &\qquad \leq \mathbb{E}^{\mathbb{P}}\left[\left(\sum_{1}^{\infty} \big\|X_n - X_{n-1}\big\|_E^2\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} \\ &\qquad\qquad \leq B_p \sup_{n \in \mathbb{N}} \big\|X_n - X_0\big\|_{L^p(\mathbb{P};E)}, \end{aligned} \tag{6.3.7}$$

with B_p as in (6.3.2).

Proof: Let $F = \ell^2(\mathbb{N}; E)$ be the separable Hilbert space of sequences

$$y = \big(x_0, \ldots, x_n, \ldots\big) \in E^{\mathbb{N}}$$

satisfying

$$\|y\|_F \equiv \left(\sum_{0}^{\infty} \|x_n\|_E^2\right)^{\frac{1}{2}} < \infty;$$

and define

$$Y_n(\omega) = \big(X_0(\omega), X_1(\omega) - X_0(\omega), \dots, X_n(\omega) - X_{n-1}(\omega), 0, 0, \dots\big) \in F$$

for $\omega \in \Omega$ and $n \in \mathbb{N}$. Obviously, $\big(Y_n, \mathcal{F}_n, \mathbb{P}\big)$ is an F-valued martingale. Moreover,

$$\|X_0\|_E = \|Y_0\|_F \quad \text{and} \quad \|X_n - X_{n-1}\|_E = \|Y_n - Y_{n-1}\|_F, \quad n \in \mathbb{N};$$

and therefore the right hand side of (6.3.7) is implied by (6.3.2) while the left hand side also follows from (6.3.2) when the roles of the X_n and Y_n's are reversed. □

Exercises for § 6.3

EXERCISE 6.3.8. Because it arises repeatedly in the theory of stochastic integration, one of the most frequent applications of Burkholder's inequality is to situations in which E is a separable Hilbert space and $(X_n, \mathcal{F}_n, \mathbb{P})$ is an E-valued martingale for which one has an estimate of the form

$$K_p \equiv \sup_{n \in \mathbb{Z}^+} \left\| \mathbb{E}^{\mathbb{P}}\Big[\|X_n - X_{n-1}\|_E^{2p} \big| \mathcal{F}_{m-1}\Big]^{\frac{1}{2p}} \right\|_{L^\infty(\mathbb{P};\mathbb{R})} < \infty$$

for some $p \in [1, \infty)$. To see how such an estimate gets used, let F be a second separable Hilbert space and suppose that $\{\boldsymbol{\sigma}_n : n \in \mathbb{N}\}$ is a sequence of $\mathrm{Hom}(E;F)$-valued random variables with the properties that, for each $n \in \mathbb{N}$, $\boldsymbol{\sigma}_n$ is $\mathcal{F}_n$-measurable and $a_n \equiv \mathbb{E}^{\mathbb{P}}\big[\|\boldsymbol{\sigma}_n\|_{\mathrm{op}}^{2p}\big]^{\frac{1}{2p}} < \infty$. Set $Y_0 = 0$ and

$$Y_n = \sum_{m=1}^{n} \boldsymbol{\sigma}_{m-1}\big(X_m - X_{m-1}\big) \quad \text{for } n \in \mathbb{Z}^+,$$

and show that

$$\big\|Y_n\big\|_{L^{2p}(\mathbb{P};F)} \le (2p-1) n^{\frac{1}{2}} K_p \left(\frac{1}{n}\sum_{m=0}^{n-1} a_m^{2p}\right)^{\frac{1}{2p}}.$$

EXERCISE 6.3.9. Return to the setting in Exercise 5.2.37, and let $\lambda_{[0,1)}$ denote Lebesgue measure on $[0,1)$. Given $f \in L^2(\lambda_{[0,1)};\mathbb{C})$, show that, for each $p \in (1, \infty)$,

$$\begin{aligned}(p-1) \wedge \frac{1}{p-1} \big\|f - (f, \mathbf{1})_{L^2(\lambda_{[0,1)};\mathbb{C})}\big\|_{L^p([0,1);\mathbb{C})} &\\ \le \left(\int_{[0,1)} \left(\sum_{m=0}^{\infty} \big|\Delta_m(f)\big|^2\right)^{\frac{p}{2}} dt\right)^{\frac{1}{p}} &\\ \le (p-1) \vee \frac{1}{p-1} \big\|f - (f, \mathbf{1})_{L^2(\lambda_{[0,1)};\mathbb{C})}\big\|_{L^p(\lambda_{[0,1)};\mathbb{C})}.&\end{aligned}$$

For functions f with $(f, e_\ell)_{L^2(\lambda_{[0,1)};\mathbb{C})} = 0$ unless $\ell = \pm 2^m$ for some $m \in \mathbb{N}$, this estimate is a case of a famous theorem proved by Littlewood and Paley in order to generalize Parseval's identity to cover $p \neq 2$. Unfortunately, the argument here is far too weak to give their inequality for general f's.

EXERCISE 6.3.10. In connection with the preceding exercise, it is interesting to note that there is an orthonormal basis for $L^2\big(\lambda_{[0,1)};\mathbb{R}\big)$ which, as distinguished from the trigonometric functions, can be nearly completely understood in terms of martingale analysis. Namely, recall the Rademacher functions $\{R_n : n \in \mathbb{Z}^+\}$ introduced in § 1.1.2. Next, use $\mathfrak{F}$ to denote the set of all finite subsets F of $\mathbb{Z}^+$, and define the **Walsh function** W_F for $F \in \mathfrak{F}$ by

$$W_F = \begin{cases} \mathbf{1} & \text{if} \quad F = \emptyset \\ \prod_{m \in F} R_m & \text{if} \quad F \neq \emptyset. \end{cases}$$

Finally, set $A_0 = \emptyset$ and $A_n = \{1, \dots, n\}$ for $n \in \mathbb{Z}^+$.

(**i**) For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the σ-algebra generated by the partition

$$\left\{ \left[\tfrac{k}{2^n}, \tfrac{k+1}{2^n}\right) : 0 \le k < 2^n \right\},$$

show that, for each $n \in \mathbb{Z}^+$, $\big\{W_F : F \subseteq A_n\big\}$ is an orthonormal basis for the subspace $L^2\big([0,1), \mathcal{F}_n, \lambda_{[0,1)};\mathbb{R}\big)$, and conclude from this that $\big\{W_F : F \in \mathfrak{F}\big\}$ forms an orthonormal basis for $L^2\big(\lambda_{[0,1)};\mathbb{R}\big)$.

(**ii**) Let $f \in L^1\big([0,1);\mathbb{R}\big)$ be given, and set

$$X_n^f = \sum_{F \subseteq A_n} \left(\int_{[0,1)} f(t)\, W_F(t)\, dt \right) W_F \quad \text{for } n \in \mathbb{N}.$$

Using the result in (**i**), show that $X_n^f = \mathbb{E}^{\lambda_{[0,1)}}[f|\mathcal{F}_n]$ and therefore that $\big(X_n^f, \mathcal{F}_n, \lambda_{[0,1)}\big)$ is a martingale. In particular, $X_n^f \longrightarrow f$ both (a.e., $\lambda_{[0,1)}$) as well as in $L^1\big(\lambda_{[0,1)};\mathbb{R}\big)$.

(**iii**) Show that for each $p \in (1,\infty)$ and $f \in L^1\big(\lambda_{[0,1)};\mathbb{R}\big)$ with mean value 0:

$$\begin{aligned} &(p-1) \wedge (p-1)^{-1} \|f\|_{L^p([0,1);\mathbb{R})} \\ &\qquad \le \left[\int_{[0,1)} \left(\sum_{n=1}^{\infty} \left[\sum_{F \subseteq A_n \setminus A_{n-1}} \left(\int_{[0,1)} f(s)\, W_F(s)\, ds \right) W_F(t) \right]^2 \right)^{\frac{p}{2}} dt \right]^{\frac{1}{p}} \\ &\qquad\qquad \le (p-1) \vee (p-1)^{-1} \|f\|_{L^p([0,1);\mathbb{R})}. \end{aligned}$$

EXERCISE 6.3.11. Although Burkholder's inequality is extremely useful, it does not give particularly good estimates in the case of martingales with bounded increments. For such martingales, the following line of reasoning, which was introduced by J. Azema in his thesis, is useful.

(**i**) For any $a \in \mathbb{R}$ and $x \in [-1,1]$, show that

$$e^{ax} \le \frac{1+x}{2}e^{a} + \frac{1-x}{2}e^{-a} = \cosh a + x \sinh a.$$

(**ii**) Suppose that $\{Y_1, \dots, Y_n\}$ are $[-1,1]$-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the property that, for each $1 \le m \le n$,

$$\mathbb{E}^{\mathbb{P}}\big[Y_{j_1} \cdots Y_{j_m}\big] = 0 \quad \text{for all } 1 \le j_1 < \cdots < j_m \le n.$$

Show that, for any $\{a_j\}_1^n \subseteq \mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}}\left[\exp\left(\sum_{j=1}^{n} a_j Y_j\right)\right] \le \prod_{j=1}^{n} \cosh a_j \le \exp\left(\frac{1}{2}\sum_{j=1}^{n} a_j^2\right),$$

and conclude that

$$P\left(\sum_{j=1}^{n} a_j Y_j \ge R\right) \le \exp\left(-\frac{R^2}{2\sum_{j=1}^{n} a_j^2}\right), \quad R \in [0,\infty).$$

(**iii**) Suppose that $(X_n, \mathcal{F}, \mathbb{P})$ is a bounded martingale with $X_0 \equiv 0$, and set $D_n \equiv \|X_n - X_{n-1}\|_{L^\infty(\mathbb{P})}$. Show that

$$P(X_n \ge R) \le \exp\left(-\frac{R^2}{2\sum_{j=1}^{n} D_j^2}\right), \quad R \in [0,\infty).$$