A constructive view on ergodic theorems

Bas Spitters
Radboud University
Nijmegen, the Netherlands

Abstract. Let $T$ be a positive $L_1$-$L_{\infty}$ contraction. We prove that the following statements are equivalent in constructive mathematics.

1. The projection in $L_2$ on the space $N := \text{cl}\{x-Tx : x \in L_2\}$ exists;
2. The sequence $(T^n)_{n \in \mathbb{N}}$ Cesàro-converges in the $L_2$ norm;
3. The sequence $(T^n)_{n \in \mathbb{N}}$ Cesàro-converges almost everywhere.

Thus, we find necessary and sufficient conditions for the Mean Ergodic Theorem and the Dunford-Schwartz Pointwise Ergodic Theorem.

1. Introduction

Bishop [4] (p.233) put forward the following problem connected with the question of finding a constructive interpretation of ergodic theorems.

Think of $X$ as the union of two equal tanks of fluid, and $T$ as a motion of the fluid, which is supposed to keep the fluid confined to the tank in which it has been placed. Imagine that there may be a small leak, which would in fact allow the fluid in the two tanks to mix, but that we are not able to decide whether a leak actually exists. Since the leak if it exists, is small, there will be little mixing between the tanks after unit time (that is, under the transformation $T$), but after a long time (that is, under the transformation $T^n$ for some large $n$) the mixing may be substantial.

He concluded that Birkhoff’s Ergodic Theorem is non-constructive.

Bishop proved, using so-called upcrossings, a version of the Chacon-Ornstein Theorem. This theorem is a generalization of Dunford and Schwartz’s version of the Pointwise Ergodic Theorem, a result which extends Birkhoff’s Ergodic Theorem. In Bishop’s ergodic theorem a limit is proved to exists in a constructively very weak sense. Bishop’s result is a so-called equal-hypothesis substitute for the ergodic theorem. Bishop considered finding an equal-conclusion substitute to be ‘an important open problem’, see [2] (p55). His student Nuber [8][9] found such an equal-conclusion substitute for Birkhoff’s Ergodic Theorem. His proof uses measure theoretic techniques and seems to work only for measure-preserving transformations. We use functional analytic techniques to give necessary and sufficient conditions for von Neumann’s Mean Ergodic Theorem and the Dunford and Schwartz version of the Pointwise Ergodic Theorem to hold.

In the context of Bishop’s constructive mathematics [3] we prove that for the Mean Ergodic Theorem to hold it is sufficient that the projection on the space of invariant functions exists. Conversely, from the convergence of the sequence in the conclusion of that theorem we obtain the projection. We also show that the Mean Ergodic Theorem is sufficient to prove the Dunford and Schwartz version of the Pointwise Ergodic Theorem, and again a converse is also true. The aim of this paper is to make these claims rigorous, see Theorem 16.

The paper is organized as follows. We first prove a Mean Ergodic Theorem. Then the Maximal Ergodic Theorem and Banach’s Principle are proved and used to prove the Pointwise Ergodic Theorem. Our presentation loosely follows that of Krengel [7] (p.65,p.159) and Dunford and Schwartz [5]. We use [3] as a general reference for constructive mathematics.

2. The mean ergodic theorem

The following definitions will be used throughout this chapter.
Let \( T \) be an operator on a Banach space \( \mathcal{X} \). Define the sum \( S_n := \sum_{k=0}^{n-1} T^k \) and the average \( A_n := \frac{1}{n} S_n \). Define the subspaces \( \mathcal{M} := \{ f \in \mathcal{X} : T \mathcal{F} \equiv f \} \) and \( \mathcal{N} := \text{cl} \{ x - T x : x \in \mathcal{X} \} \). An operator \( T \) is called a contraction if \( \|T x\| \leq \|x\| \) whenever \( x \in \mathcal{X} \).

When \( \mathcal{X} \) is a vector space and \( \mathcal{Y}, \mathcal{Z} \) are subspaces such that for every \( x \) in \( \mathcal{X} \) there exist unique \( y \) in \( \mathcal{Y} \) and \( z \) in \( \mathcal{Z} \) such that \( x = y + z \), we write \( \mathcal{X} = \mathcal{Y} \oplus \mathcal{Z} \).

**Theorem 1.** Let \( T \) be a contraction on a Banach space \( \mathcal{X} \). The sequence \( (A_n)_{n \in \mathbb{N}} \) converges if and only if \( \mathcal{X} = \mathcal{M} \oplus \mathcal{N} \), in which case \( \lim_{n \to \infty} A_n = P \mathcal{M} \), where \( P \mathcal{M} \) denotes the projection on \( \mathcal{M} \) parallel to \( \mathcal{N} \).

**Proof.** First suppose that \( \mathcal{X} = \mathcal{M} \oplus \mathcal{N} \). Let \( h = h_\mathcal{M} + h_\mathcal{N} \), where \( h_\mathcal{M} \in \mathcal{M} \) and \( h_\mathcal{N} \in \mathcal{N} \). We claim that \( A_n h \) converges to \( h \). First consider \( f = g - T g \) for some \( g \in \mathcal{X} \); then the sequence \( A_n f = \frac{1}{n} (g - T^n g) \) converges to 0. When \( f \in \mathcal{N} \), then there exists \( g \in \mathcal{X} \) such that \( \| f - (g - T g) \| < \varepsilon \), so for all \( n \in \mathbb{N} \), \( \| A_n f (g - T g) \| < \varepsilon \). Hence for large \( n \), \( \| A_n f \| < 2 \varepsilon \). Consequently, the sequence \( A_n f \) converges to 0 whenever \( f \in \mathcal{N} \) and, using the notation above, \( A_n h \) converges to \( h \).

Let \( f \in \mathcal{X} \); then there exist \( f_\mathcal{M} \in \mathcal{M} \) and \( f_\mathcal{N} \in \mathcal{N} \) such that \( f = f_\mathcal{M} + f_\mathcal{N} \). Consequently, \( A_n f = f_\mathcal{M} + A_n f_\mathcal{N} \) which converges to \( f_\mathcal{M} \) when \( n \) tends to \( \infty \).

Now suppose that the sequence \( (A_n)_{n \in \mathbb{N}} \) converges to an operator \( P \). The equalities \( T P = P = PT \) follow easily from the definition of the sequence \( (A_n)_{n \in \mathbb{N}} \). Consequently, \( P = 0 \) on \( \mathcal{N} \) and

\[
P^2 = \lim_{n \to \infty} A_n P^{(TP = P)} = \lim_{n \to \infty} P = P.
\]

If \( z \in \mathcal{M} \cap \mathcal{N} \) and \( \varepsilon > 0 \), then there exists \( u \in \mathcal{X} \) such that \( \| z - (u - T u) \| < \varepsilon \). Hence for all \( n \in \mathbb{N} \), \( \| A_n (z - (u - T u)) \| < \varepsilon \). Because \( A_n (u - T u) \) converges to 0 and for all \( n \in \mathbb{N} \), \( A_n z = z \) we see that \( \| z \| \leq \varepsilon \). Consequently, \( \mathcal{M} \cap \mathcal{N} = \{ 0 \} \).

To see that \((I - P)x \in \mathcal{N} \) observe that

\[
(I - T)(\frac{n-1}{n}I + \frac{n-2}{n}T + \ldots + \frac{1}{n}T^{n-2}) = I - A_n
\]

for all \( n \) in \( \mathbb{N} \). Define

\[
y_n := (\frac{n-1}{n}I + \frac{n-2}{n}T + \ldots + \frac{1}{n}T^{n-2})x;
\]

then \((I - T)y_n \to (I - P)x \). Consequently, \( \mathcal{X} = \mathcal{M} \oplus \mathcal{N} \). \( \square \)

Let \( \mathcal{H} \) be a Hilbert space and let \( T \) be an operator on \( \mathcal{H} \). Let \( x \in \mathcal{H} \). There exists a vector \( x^* \) such that \( \langle T y, x \rangle = \langle x, x^* \rangle \) if and only if the functional \( y \mapsto \langle T y, x \rangle \) is normable. This follows from the Riesz representation theorem [3] (p.419). If such a vector exists we will denote it by \( T^* x \) even if the adjoint is not totally defined\(^1\).

**Theorem 2.** [Mean Ergodic Theorem] Let \( T \) be a contraction on a Hilbert space \( \mathcal{H} \). Then the sequence \( (A_n)_{n \in \mathbb{N}} \) converges if and only if \( \mathcal{N} \) is located; in this case the sequence \( (A_n)_{n \in \mathbb{N}} \) converges to the orthogonal projection \( P_\mathcal{M} \) on \( \mathcal{M} \).

**Proof.** We first prove that \( \mathcal{M} \) and \( \mathcal{N} \) are orthogonal. Suppose that \( x \in \mathcal{M} \), i.e. \( T x = x \). We claim that the map \( y \mapsto \langle T y, x \rangle \) is normable. Since \( |\langle T y, x \rangle| \leq \| x \| \| y \| \) whenever \( y \in \mathcal{H} \), \( \| y \| \) is an upper bound on the norm. On the other hand this upper bound is attained at \( x \). It follows that \( x \in \text{Dom}T^* \) and \( \| T^* x \| = \| x \| \). Now,

\[
\| T^* x - x \|^2 = \langle T^* x - x, T^* x - x \rangle
= \| T^* x \|^2 + \| x \|^2 - \langle x, T^* x \rangle - \langle T^* x, x \rangle
= \| T^* x \|^2 + \| x \|^2 - \langle T x, x \rangle - \langle x, T x \rangle
= \| T^* x \|^2 + \| x \|^2 - 2 \langle x, x \rangle = \| T^* x \|^2 - \| x \|^2 = 0.
\]

\(^1\) Classically, the adjoint of an operator is always totally defined. Constructively this is not the case.
Consequently, $T^*x = x$ and so
\[ \langle x, (I - T)y \rangle = \langle (I - T)^*x, y \rangle = 0 \]
for all $y$ in $\mathcal{H}$. We see that $\mathcal{M}$ and $\mathcal{N}$ are orthogonal.

Suppose that the sequence $(A_n)_{n \in \mathbb{N}}$ converges. Theorem 1 shows that $\mathcal{H} = \mathcal{M} \oplus \mathcal{N}$. Because $\mathcal{M}$ and $\mathcal{N}$ are also orthogonal, $\mathcal{M}$ and $\mathcal{N}$ are located.

Conversely, suppose that $\mathcal{N}$ is located. We know that $\mathcal{M} \subset \mathcal{N}^\perp$. We will prove that $\mathcal{N}^\perp \subset \mathcal{M}$. Let $x \in \mathcal{N}^\perp$. Then $\langle (I - T)y, x \rangle = 0$ whenever $y \in \mathcal{H}$. It follows that $\langle y, x \rangle = (T y, x)$, i.e. $x = T^*x$. By a similar argument as above we see that $T x = x$. We conclude that $\mathcal{M} = \mathcal{N}^\perp$, so by Theorem 1 the sequence $(A_n)_{n \in \mathbb{N}}$ converges.

Let $(X, \mu)$ be a measure space. A measure-preserving transformation of $X$ is a partial function from a full set to a full set such that for all integrable sets $A$, $\tau(A)$ is integrable and $\mu(\tau(A)) = \mu(A)$. If $\tau$ is a measure-preserving transformation, then $T_\tau f := f \circ \tau$ is a contraction on $L_2$. This shows that our result generalizes the possibly more familiar formulation of the Theorem.

Bishop and Bridges [3] (problem 46, p.395) give the following version of the Mean Ergodic Theorem. Let $T$ be a unitary operator on a Hilbert space $H$; then for all $x \in H$ the sequence $(A_n x)_{n \in \mathbb{N}}$ converges if and only if the sequence $(\|A_n x\|)_{n \in \mathbb{N}}$ converges.

3. Maximal Ergodic Theorems

Let $(X, \mu)$ be a measure space. An operator $T$ on $L_1(\mu)$ is an $L_1$-$L_\infty$ contraction if $T$ is a contraction on $L_1$ that contracts the $L_\infty$-norm on $L_1 \cap L_\infty$ — that is, $\|f\|_1 \leq \|T f\|_1$ for all real numbers $m$, $\|T f\|_1 \leq m$ whenever $f \in L_1$ and $|f| \leq m$. An operator $T$ on an ordered vector space is positive if $T f \geq 0$ whenever $f \geq 0$. When $\tau$ is a measure-preserving transformation, then $T_\tau$ is a positive $L_1$-$L_\infty$ contraction.

Let $T$ be a positive $L_1$-$L_\infty$ contraction. Define the operator $M_n$ by
\[
M_n f := \sup_{k \leq n} A_k f
\]
for all $n$ in $\mathbb{N}$. Garcia’s proof [7] (p.8) of the following Theorem 3 is constructive. Note however that we do not make any claims about $M_\infty$. This operator is defined classically as $\sup_{k \in \mathbb{N}} A_k$, but constructively $M_\infty f$ may not be a measurable function for all $f \in L_1$, i.e. we may not be able to find simple functions approximating $M_\infty f$.

In the constructive theory of measure spaces it is not always possible to compute the measure of the set $\{f < \alpha\} := \{x : f(x) < \alpha\}$. However, we can compute the measure for all but countably many $\alpha$, such $\alpha$ are called admissible, see [3] for details.

**Theorem 3.** [Hopf’s Maximal Ergodic Theorem] Let $T$ be a positive contraction on $L_1(\mu)$. Let $n$ be a natural number. If $\alpha \geq 0$ is admissible for $M_n f$, then
\[
\int_{\{M_n f \geq \alpha\}} f \geq 0.
\]

**Corollary 4.** [Wiener [7] (p.51)] Let $T$ be a positive $L_1$-$L_\infty$ contraction. Let $f \in L_1$, $n \in \mathbb{N}$ and $\alpha > 0$ be admissible for $M_n f$. Then for all $n$,
\[
\mu[M_n f \geq \alpha] \leq \frac{1}{\alpha} \int_{\{M_n f \geq \alpha\}} f.
\]

The following theorem is sometimes called the little Riesz theorem. The proof we give here is an adaptation of [7] (Lemma 1.7.4).

2. When $p = 2$, this order-theoretic definition differs from the definition of a positive operator on a Hilbert space.
Proposition 5. If $T$ is a positive $L_1$-$L_\infty$ contraction, then $T$ can be uniquely extended to an $L_p$ contraction for all $p \geq 1$.

Proof. Because for all $p \geq 1$, $L_1 \cap L_\infty$ is a dense subset of $L_p$, it is enough to prove that $\|Tf\|_p \leq \|f\|_p$ for all $f \in L_1 \cap L_\infty$. To achieve this goal we will prove that

$$Tf \leq T(f^p)^{\frac{1}{p}}$$

(1)

for all positive simple functions $f$ and $p > 1$.

Assume for a moment that we have done so and let $f$ be a positive simple function and $p > 1$. Then $(Tf)^p \leq T(f^p)$, so $\|(Tf)^p\|_1 \leq \|T(f^p)\|_1 \leq \|f^p\|_1$ and thus $\|Tf\|_p \leq \|f\|_p$. Observe that this inequality trivially holds for $p = 1$. Consequently, $\|Tf\|_p > \|f\|_p$ is impossible, i.e. $\|Tf\|_p \leq \|f\|_p$ holds for all $p \geq 1$, even if we are unable to decide $p = 1$ or $p > 1$. It follows that for all $p \geq 1$, $T$ is a positive $L_p$-contraction on the positive simple functions, and consequently also on the simple functions. The simple functions are dense and $T$ is a contraction, so the operator $T$ restricted to the simple functions can be uniquely extended to $L_p$. This extension agrees with $T$ on $L_1 \cap L_p$ for all $p \geq 1$.

We will now prove (1) for a positive simple function $f$ and $p > 1$. We will assume that $Y := [f > 0]$ is integrable. Since the simple functions $f$ with this property are also dense in $L_p^+$, we do not lose generality. We note that $f = f_{XY}$. We define $q := (1 - \frac{1}{p})^{-1}$ as usual. For all real numbers $a, b \geq 0$:

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

It follows that for all real numbers $c, d > 0$:

$$\frac{f}{c} \frac{\chi_Y}{d} \leq \frac{f^p}{c^p} + \frac{\chi_Y^q}{d^q}$$

Consequently,

$$T(f \cdot \chi_Y) \leq T(f^p) \frac{cd}{c^p} + T(\chi_Y^q) \frac{dc}{d^q} \quad a.e.$$

(2)

Let $F$ be a full set on which (2) holds for rational $c$ and $d$ and thus by continuity for all $c, d > 0$. Compute $M, m \in \mathbb{R}^+$ such that $M \geq f \geq m \chi_Y > 0$. Then $f \leq m^{1-p}f^p$. Let $F' \subset F$ be a full set such that for all $x \in F'$

$$f(x) \leq M, \quad (Tf)(x) \leq m^{1-p}(Tf^p)(x), \quad (Tf)(x) \leq M(T\chi_Y)(x) \quad \text{and} \quad (T\chi_Y)^{\frac{1}{p}}(x) \leq 1.$$

Fix $x \in F'$.

If $T(f^p)(x) = 0$, then $T(f)(x) \leq m^{1-p}T(f^p)(x) = 0 = T(f^p)^{\frac{1}{p}}(x)$.

If $T(\chi_Y)(x) = 0$, then $T(f)(x) \leq MT(\chi_Y)(x) = 0 \leq T(f^p)^{\frac{1}{p}}(x)$.

If $T(f^p)(x) > 0$ and $T(\chi_Y)(x) > 0$, then we define $c := T(f^p)^{\frac{1}{p}}(x)$ and $d := T(\chi_Y)^{\frac{1}{p}}(x)$. The right hand side of (2) equals $cd\left(\frac{1}{p} + \frac{1}{q}\right) = cd$. Because $f = f\chi_Y$ we obtain:

$$T(f)(x) \leq T(f^p)^{\frac{1}{p}} T(\chi_Y)^{\frac{1}{p}}(x) \leq T(f^p)^{\frac{1}{p}}(x).$$

(3)

We conclude that in any case $T(f)(x) > T(f^p)^{\frac{1}{p}}(x)$ is impossible. It follows that (3) holds for all $x \in F'$. Consequently, (1) holds and we have thus completed the proof. 

From this point onwards we will assume that a positive $L_1$-$L_\infty$ contraction is extended to $L_p$ for all $p \geq 1$.

Theorem 6. [Dominated Ergodic Theorem] Let $T$ be a positive $L_1$-$L_\infty$ contraction. Then for all $n \in \mathbb{N}$, $p > 1$ and $f \in L_p$: $\|M_n f\|_p \leq \frac{p}{p-1} \|f\|_p$. 
Proof. Fix \( n \in \mathbb{N}, \ p > 1, \) and \( f \in L_p^+. \)
\[
\int (M_n f)^p d\mu = \int_0^{M_n f(s)} p \alpha^{p-1} d\alpha d\mu(s) \\
= p \int_0^\infty \alpha^{p-1} \chi_{M_n f \geq \alpha}(s) d\alpha d\mu(s) \\
\overset{\text{Fubini}}{=} p \int_0^\infty \alpha^{p-2} f(s) d\mu d\alpha \\
\overset{\text{Wiener}}{\leq} p \int_0^\infty \bigg( \alpha^{p-2} \chi_{M_n f \geq \alpha}(s) \bigg) f(s) d\mu d\alpha \\
\overset{\text{Fubini}}{=} p \int f(s) \int_0^\infty \alpha^{p-2} \chi_{M_n f \geq \alpha}(s) d\alpha d\mu(s) \\
= p \int f(s) \int_0^{M_n f(s)} \alpha^{p-2} d\alpha d\mu(s) \\
= p \int_0^\infty \bigg( \frac{f(M_n f)^{p-1}}{p-1} d\mu \bigg) \\
\leq \frac{p}{p-1} \|f\|_p \|M_n f\|_p^{p-1}. 
\]
This last inequality follows from Hölder’s inequality and the fact that \( M_n f \leq \sum_{k=0}^n A_k f, \) so that \( M_n f \in L_p. \) For all \( f \in L_p, \ M_n f \leq M_n |f| \) and thus \( \|M_n f\|_p \leq \|M_n |f|\|_p \leq \frac{p}{p-1} \|f\|_p. \)

A function \( f: \mathbb{R} \to \mathbb{R} \) is convex if
\[
f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y),
\]
whenever \( \lambda \in [0, 1] \) and \( x, y \in \mathbb{R}. \)

Theorem 7. A (total) convex function from \( \mathbb{R} \) to \( \mathbb{R} \) has an non-decreasing derivative which is defined in all but countably many points.

Proof. To prove this we will use Bishop’s profile theorem, a constructive substitute for the classical lemma stating that every non-decreasing real function is continuous in all but countably many points. Like Bishop we define the set \( \mathcal{E} \) to be the set of piecewise linear functions
\[
h_{xy}(z) := \min \{z, y\} - \min \{z, x\} \\
y - x
\]
whenever \( x < y. \) These functions are 0 when \( z \leq x \) and 1 when \( x \geq y. \) We define
\[
\Lambda(h_{xy}) := \frac{f(y) - f(x)}{y - x}.
\]
Then \( \Lambda \) is increasing\(^3\) on the set \( \mathcal{E} \) with the order inherited from the the functions from \( \mathbb{R} \) to \( \mathbb{R} \) — that is, \( (\mathcal{E}, \Lambda) \) is a profile. The profile theorem ensures that all but countably many points are smooth. In the present case this means that the function is differentiable at these points. \( \square \)

Lemma 8. [Jensen’s inequality] Let \( (X, \mu) \) be a finite measure space and \( \phi: \mathbb{R} \to \mathbb{R} \) be a convex function. If \( f, \phi \circ f \) are in \( L_1, \) then
\[
\phi \left( \frac{1}{\mu(X)} \int f \right) \mu(X) \leq \int \phi \circ f.
\]

Proof. Let \( x \) be in the domain of \( \phi'. \) Then for all real numbers \( y, \)
\[
\phi(y) \geq \phi'(x)(y - x) + \phi(x),
\]
and hence
\[
\phi(f(t)) \geq \phi'(x)(f(t) - x) + \phi(x)
\]

\(^3\) This is most easily seen by a geometric argument considering the graph of a convex function.
for all \( t \in \text{Dom} f \). By integrating we obtain

\[
\int \phi \circ f \geq \phi'(x) \left( \left( \int f \right) - x \mu(X) \right) + \phi(x) \mu(X).
\]

If we take a sequence \((x_n)_{n \in \mathbb{N}}\) in \( \text{Dom} \phi' \) tending to \( \frac{1}{\mu(X)} \int f \), we obtain the inequality above. \( \square \)

The \( L_p \) Ergodic Theorem can be proved classically for finite measure spaces using the pointwise ergodic theorem \([14]\). Another proof \([7]\) (Thm. 2.1.2) uses a non-constructive compactness argument. Our proof works not only for finite measure spaces, but also for \( \sigma \)-finite measure spaces. We make some preparations.

Let \( \mu \) be a finite measure. If \( 1 \leq p < q \) and \( f \in L_q \), then \(|f|^p\) is measurable and bounded by \(|f|^q \vee 1\), hence \(|f|^p\) is integrable and by taking \( \phi(x) = x^q \) in Jensen’s inequality we obtain

\[
\|f\|_p \leq \|f\|_q \mu(X)^{1 - \frac{1}{q}}.
\]

We see that \( L_q \subset L_p \).

Let \( \mu \) be \( \sigma \)-finite. If \( p \geq q \geq 1 \) and \( f \leq M \), then

\[
\|f\|_p^p = \int |f|^p = \int |f|^q |f|^{p - q} \leq M^{p - q} \|f\|_q^q.
\]

Consequently,

\[
\|f\|_p \leq M^{1 - \frac{q}{p}} \|f\|_q^{\frac{q}{p}}.
\]

**Theorem 9.** [\( L^p \) Ergodic Theorem] Let \( p, q \geq 1 \) and \((X, \mu)\) be a finite measure space or let \( p > 1, q \geq 1 \) and \((X, \mu)\) a \( \sigma \)-finite measure space. Let \( T \) be a positive \( L_1-L_\infty \) contraction. If the sequence \((A_n)_{n \in \mathbb{N}}\) converges in \( L_q \), then it converges in \( L_p \).

**Proof.** Let \( f \in L_p \) and choose a simple \( g \) such that \( \|f - g\|_p < \frac{\varepsilon}{4} \). Let \( M \) be a bound for \( g \). For all \( n, m \in \mathbb{N} \):

\[
\|A_n f - A_m f\|_p \leq \|A_n f - A_n g\|_p + \|A_n g - A_m g\|_p + \|A_m g - A_m f\|_p \\
\leq \frac{\varepsilon}{4} + \|A_n g - A_m g\|_p + \frac{\varepsilon}{4}.
\]

If we show that \( \|A_n g - A_m g\|_p \to 0 \), when \( m, n \to \infty \), then \((A_n f)_{n \in \mathbb{N}}\) is a Cauchy sequence in \( L_p \). That \( \|A_n g - A_m g\|_p \to 0 \) follows from the fact that the sequence \((A_n g)_{n \in \mathbb{N}}\) converges in \( L_q \) and the inequalities \((4)\) and \((5)\) for the case \( \mu \) is finite or \( \mu \) is \( \sigma \)-finite and \( p \geq q \). The case \( \mu \) is \( \sigma \)-finite and \( 1 < p \leq q \) is more difficult. We will now proceed to consider this case. If \( p \geq 1, h \in L_p, n, m, l, r \in \mathbb{N} \) and \( m = l n + r \), then

\[
\|A_m h\|_p - \|A_n h\|_p \leq \frac{1}{t} \| I + T^n + \ldots + T^{(l - 1)n} A_n h \|_p - \| A_n h \|_p \\
+ \frac{1}{m} \| T^n (I + \ldots + T^{r - 1}) h \|_p \\
\leq 0 + \frac{r}{m} \| h \|_p \leq \frac{r}{m} \| h \|_p.
\]

It follows that the sequence \((\|A_n h\|_p)_{n \in \mathbb{N}}\) is essentially decreasing, that is for each \( n \in \mathbb{N} \) and each \( \varepsilon > 0 \), there exists \( N \in \mathbb{N} \) such that \( \|A_m h\|_p \leq \|A_n h\|_p + \varepsilon \) for all \( m \geq N \).

Let \( g \in L_q \cap L_{p'} \). Let \( \tilde{g} \) be the limit of \( A_n g \) in \( L_q \). The function \( \tilde{g} \) is in \( L_{p'} \), because \( T \) contracts the \( L_{p'} \)-norm. We will prove that \( \lim_{n \to \infty} A_n g = \tilde{g} \) in \( L_p \). By looking at \( g - \tilde{g} \) we may assume that \( \tilde{g} = 0 \).

Because the sequence \((\|A_n g\|_p)_{n \in \mathbb{N}}\) is essentially decreasing, it is enough to find for each \( \eta > 0 \) and \( k \in \mathbb{N} \), an \( n > k \) such that \( \|A_n g\|_p < \eta \). We will now proceed to find such \( n \). Let \( \beta = \frac{\eta}{2} \). Take an integrable set \( B_1 \) such that \( |g(X - B_1)|_p < \beta \). We recall that \( A_n g \to 0 \) in \( L_q \). We can thus compute \( n_1 \) such that \( \|A_n g\|_q < \beta \mu(B_1)^{\frac{p}{q} - 1} \). By \((4)\)

\[
\| (A_n g) \chi_{B_1} \|_p \leq \| (A_n g) \chi_{B_1} \|_q \mu(B_1)^{1 - \frac{p}{q}} < \beta.
\]
Now
\[ \|A_{n_1}g\|_p^p < \eta \quad \text{or} \quad \|A_{n_1}g\|_p^p > \eta - \beta. \]
In the former case we are done, so we may assume that \( \|A_{n_1}g\|_p^p > \eta - \beta \) and thus \( \|(A_{n_1}g)\chi_{X-B_1}\|_p^p > \eta - 2\beta \). We choose \( B_2 \supset B_1 \) such that \( \|(A_{n_1}g)\chi_{B_2-B_1}\|_p^p \geq \eta - 3\beta \). Compute \( n_2 > n_1 \) such that \( \|A_{n_2}g\|_p^p < \beta \). The sequence converges a.e. on the whole space.

The following principle, called Banach’s principle, will be used as follows: we prove that a sequence of operators converges almost everywhere (a.e.) on a dense set and then conclude that
\[ \forall \alpha, \exists N \in \mathbb{N} \text{ such that } \|A_ng\|_p^p < \eta. \]
That such an \( N \) exists we see as follows. Choose \( K \in \mathbb{N} \) such that
\[ \left( \frac{p}{p-1} \right)^p - 1 \|g\|_p^p + \beta \leq K(\eta - 3\beta). \]
For all \( N \leq K \), \( \|A_{n,N}g\|_p^p > \eta - \beta \) or \( \|A_{n,N}g\|_p^p < \eta \). Suppose that for all \( N \leq K \), \( \|A_{n,N}g\|_p^p > \eta - \beta \). Define \( B_0 = \emptyset \) and \( n_0 = 0 \). Remark that for all \( N \in \mathbb{N} \),
\[ M_{N}\|g\| \geq \sum_{i=0}^{N} M_{N}\|g\|\chi_{B_{i+1}-B_i} \]
\[ \geq \sum_{i=0}^{N} A_{n_i}\|g\|\chi_{B_{i+1}-B_i} \]
\[ \geq \sum_{i=0}^{N} A_{n_i}g\chi_{B_{i+1}-B_i}. \]
It follows from the Dominated Ergodic Theorem that for all \( N \leq K \),
\[ \left( \frac{p}{p-1} \right)^p \|g\|_p^p \geq \|M\|_p \|g\|_p^p \]
\[ \geq \| \sum_{i=0}^{N} (A_{n_i}g)\chi_{B_{i+1}-B_i}\|_p^p \]
\[ = \sum_{i=0}^{N} \| (A_{n_i}g)\chi_{B_{i+1}-B_i}\|_p^p \]
\[ \geq N(\eta - 3\beta) + \|g\|_p^p - \beta. \]
It follows that there exists \( N \leq K \) such that \( \|A_{n,N}g\|_p^p < \eta \). Consequently, \( A_ng \to \bar{g} \) in \( L_p \).

Finally, we observe that even if we do not know whether \( p \leq q \) or \( p > q \) we can show that the sequence \( A_n \) converges in \( L_p \). To see this we observe that either \( p < q + \varepsilon \) or \( p > q \), the latter case has been treated before. In the former case, the sequence also converges in \( L_{q+\varepsilon} \) as we proved before, and we can thus proceed as above.

If \( \mu \) is \( \sigma \)-finite, it is not true in general that if for some \( p > 1 \), the sequence \( A_n \) converges to 0 in \( L_p \), then the sequence \( A_n \) converges in \( L_1 \). To see this let \( \tau(x) = x + 1 \) on \( \mathbb{R} \) with Lebesgue measure, \( T = T_\tau \) and \( f = \chi_{[0,1]} \). For all \( p > 1 \), \( A_nf \) converges to 0 in \( L_p \), but the sequence does not converge in \( L_1 \).

4. The Pointwise Ergodic Theorem

The following principle, called Banach’s principle, will be used as follows: we prove that a sequence of operators converges almost everywhere (a.e.) on a dense set and then conclude that the sequence converges a.e. on the whole space.

The proof of the following theorem would be easier if we could prove constructively that \( M_\infty \) is measurable.

**Theorem 10.** [Banach’s Principle] Let \( (Y, \mu) \) be a measure space. Let \( (T_n)_{n \in \mathbb{N}} \) be a sequence of linear operators from a Banach space \( X \) to the space \( \mu \)-measurable real functions. Define for each \( n \in \mathbb{N} \), an operator \( \tilde{M}_n \) by \( \tilde{M}_n x = \sup_{k \leq n} |T_k x| \) for all \( x \in X \). Suppose that there exists a positive decreasing function \( C \) from \( \mathbb{R} \) to \( \mathbb{R} \) such that \( \lim_{\alpha \to \infty} C(\alpha) = 0 \) and for all \( x \) and \( n \):
\[ \mu |\tilde{M}_n x| \geq \alpha \|x\| < C(\alpha), \]
whenever $\alpha \| x \|$ is admissible for $M_n x$. Then the set of elements $x \in X$ for which the sequence $(T_n x)_{n \in \mathbb{N}}$ converges a.e. is closed.

**Proof.** Suppose that a sequence $z_n$ converges to $z$ in $X$ in norm and that there exists a sequence $(f_n)_{n \in \mathbb{N}}$ of measurable functions such that for all $m \in \mathbb{N}$, $T_m z_n \to f_n$ a.e. We will prove that the sequence $(T_m z)_{m \in \mathbb{N}}$ converges a.e.

For natural numbers $a$ and $b$, $\omega \in Y$ and $x \in X$ put

$$\Delta_{a, b}(\omega, x) := \sup_{a \leq n, m \leq b} |T_n x(\omega) - T_m x(\omega)|.$$  

Then

$$|\Delta_{a, b}(\cdot, z) - \Delta_{a, b}(\cdot, z_n)| \leq |\Delta_{a, b}(\cdot, z - z_n)| \leq 2M_b(z - z_n).$$  

Let $\varepsilon > 0$. Choose a natural numbers $n$ such that $C(\varepsilon, \|z - z_n\|) < \varepsilon$ and choose a natural number $k$ and an integrable set $D \subset Y$ such that $\mu(D) < \varepsilon$ and for all $a \geq k$:

$$\{\omega: |T_a z_n(\omega) - f_n(\omega)| > \varepsilon \} \subset D.$$  

For all $b \geq a \geq k$,

$$\mu(\|\Delta_{a, b}(\cdot, z)\| > 2\varepsilon) \leq \mu(\|\Delta_{a, b}(\cdot, z_n)\| > \varepsilon) + \mu(\|\Delta_{a, b}(\cdot, \|z - z_n\|) > \varepsilon) \leq \mu(D) + \mu(2M_b(\|z - z_n\|) > \varepsilon) \leq \varepsilon + C(\varepsilon, \|z - z_n\|) \leq 2\varepsilon.$$  

Construct ascending sequences $(n_m)_{m \in \mathbb{N}}$ and $(k_m)_{m \in \mathbb{N}}$ of natural numbers such that

$$C(\varepsilon 2^{-m} \|z - z_{m+1}\|) < \varepsilon 2^{-m}$$  

and

$$\mu(\|T_{n_m} z - f_{n_m}\| > \varepsilon 2^{-m}) < \varepsilon 2^{-m}$$  

for all $n \geq k_m$. Put

$$E := \bigcup_{m \geq 1} \{\omega: |\Delta_{k_m, k_{m+1}}(\omega, z)| > \varepsilon 2^{-m}\}.$$  

The set $E$ is integrable and $\mu(E) \leq \sum_{m=1}^{\infty} 2\varepsilon 2^{-m} = 2\varepsilon$. For $\omega \in E$ and $n, m \geq k_1$:

$$|T_n z(\omega) - T_m z(\omega)| < \varepsilon.$$  

Consequently, the sequence $(T_m z)_{m \in \mathbb{N}}$ converges a.e. \(\square\)


When $q \geq 1$ and $T$ is a positive $L_1$-$L_\infty$ contraction, we define $M_q = \{f \in L_q; T f = f\}$ and $N_q = \text{cl}\{T f - f: f \in L_q\}$.

**Theorem 11.** [Pointwise Ergodic Theorem] Let $\mu$ be a $\sigma$-finite measure. Let $T$ be a positive $L_1$-$L_\infty$ contraction. If $L_q = M_q \oplus N_q$, for some $q \geq 1$, then for all $p \geq 1$, the sequence $(A_n f)_{n \in \mathbb{N}}$ converges a.e. for all $f \in L_p$.

**Proof.** Suppose that $L_q = M_q \oplus N_q$, for some $q \geq 1$. First let $p > 1$. Theorem 9 and Theorem 1 show that $L_p = M_p \oplus N_p$. Suppose that $f = g + T h - h$, where $g \in L_p$, $T g = g$ and $h \in L_\infty \cap L_p$. The set of these $f$ is dense in $L_p$. Because $T$ contracts the $L_\infty$-norm, $\lim_{n \to \infty} A_n f = g$ a.e. For all admissible $\alpha > 0$,

$$\alpha \mu[M_n f \geq \alpha] \leq \left( \int_{\{M_n f \geq \alpha\}} f \right)^{\frac{1}{p}} \leq \left( \int_{\{M_n f \geq \alpha\}} f^p \right)^{\frac{1}{p}} \mu[M_n f \geq \alpha]^{1 - \frac{1}{p}} \leq \| f \|_p \mu[M_n f \geq \alpha]^{1 - \frac{1}{p}}.$$
Consequently, \( \alpha \mu[M_n f \geq \alpha] \leq \|f\|_p \). Define \( \tilde{M}_n f = \sup_{k \leq n} |A_k f| \), for all \( f \in L_p \). Substituting \( \alpha = \beta \|f\|_p \) and observing that \( \tilde{M}_n f = M_n f \) when \( f \geq 0 \), we obtain
\[
\mu[\tilde{M}_n f \geq \beta \|f\|_p] \leq \beta^{-p}.
\]

For all \( f \in L_p \), \( \tilde{M}_n f \leq M_n |f| \), so inequality (6) holds for all \( f \in L_p \). Banach’s principle shows that the sequence \( (A_n f)_{n \in \mathbb{N}} \) converges a.e. for all \( f \) in \( L_p \).

Finally we remove the assumption that \( p \) is strictly greater than 1. Let \( p \geq 1 \). The set \( L_{p+1} \cap L_p \) is dense in \( L_p \), so one can apply Banach’s principle since inequality (6) above holds for all \( p \geq 1 \).

The two-step argument above is used because in general we do not have convergence in the \( L_1 \)-norm. The space \( L_1 \) has an awkward geometrical structure: it is not uniformly convex.

**Theorem 12.** Let \( p \geq 1 \). If the sequence \( (A_n f)_{n \in \mathbb{N}} \) converges a.e. for all \( f \) in \( L_p \), then \( \mathcal{N} = \text{cl}\{x-Tx: x \in L_2\} \) is located in \( L_2 \).

**Proof.** Let \( f \) be an element of \( L_2 \). Without loss of generality we may assume that \( A_n f \to 0 \) a.e. We claim that the sequence \( (A_n f)_{n \in \mathbb{N}} \) converges weakly to 0. We may assume that \( f \) is a simple function. Since the sequence \( (A_n f)_{n \in \mathbb{N}} \) is bounded, the sequence \( (A_n f)g \) converges to 0 whenever \( g \) is a simple function. The inner product is continuous on \( L_2 \), so \( (A_n f, g) \to 0 \) for all \( f, g \in L_2 \) — that is, \( (A_n f)_{n \in \mathbb{N}} \) converges weakly to 0.

Define \( B_n := \left( \frac{2-n}{n} I + \frac{n-2}{n} T + \cdots + \frac{2-n}{n} T^{n-2} \right) \). By an argument similar to the proof of Theorem 1 we see that for each \( x \), \( (I-T)B_n x \) converges weakly to an element in the space \( \text{wcl}\{y-Ty: y \in L_2\} \) and thus
\[
L_2 = M \oplus \text{wcl}\{(I-T)x: x \in L_2\},
\]
where \( \text{wcl} \) denotes the weak closure. Since \( (I-T)B_n x = (I-A_n)x \) is a bounded sequence and the weak closure of a bounded convex inhabited subset coincides with its strong closure, by Lemma 5.2.4 in [11], we see that \( \mathcal{N} = \text{wcl}\{y-Ty: y \in L_2\} \). It follows that the display sum above is orthogonal and thus that \( \mathcal{N} \) is located.

**Lemma 13.** [6] Let \( E \) be a uniformly convex and uniformly smooth Banach space and \( C \) a bounded convex subset of \( E \). Then \( C \) is located if and only if \( \sup \{ f(z): z \in C \} \) exists for each normable linear functional \( f \) on \( E \).

**Lemma 14.** [3][6] For \( p > 1 \), the space \( L_p \) is uniformly convex and uniformly smooth and each normable functional may be represented by a functional \( f \mapsto \int f g \), for some \( g \) in \( L_q \), where \( \frac{1}{p} + \frac{1}{q} = 1 \).

**Lemma 15.** If \( \mathcal{N} \) is located in some \( L_p \), then both \( M \) and \( \mathcal{N} \) are located in all \( L_p \), where \( p > 1 \).

**Proof.** Locatability is a property of closed sets, so we may restrict to a set which is dense in all the \( L_p \)-spaces, for instance the bounded \( L_1 \) functions or the simple functions. Moreover, an inhabited set \( A \) is located if we can compute the distance \( \rho(x, A) \) for each \( x \). Let \( a \) be an element of \( A \). Then \( \rho(x, A) = \rho(x, A \cap B(x, \rho(x, a) + 1)) \), where \( B(x, r) \) denotes the ball around \( x \) with radius \( r \). Consequently, when considering located sets we may restrict to its bounded subsets.

We will now apply Lemma 14. Since the simple functions are dense in all the \( L_p \)-spaces, if \( \sup \{ z: z \in C \} \) exists for all \( f \) in \( L_p \), then this supremum exists for all \( f \) in set of the simple functions, and thus for all \( f \) in any \( L_{p'} \), where \( p' > 1 \). It follows that \( A \) is located in \( L_p \). In particular, if \( \mathcal{N} \) are located in \( L_p \), then it is located in \( L_2 \). Since these sets are orthogonal, they are thus both located in \( L_2 \). Consequently, they are both located in any \( L_{p'} \).

**Theorem 16.** Let \( \mu \) be a \( \sigma \)-finite measure and \( T \) a positive \( L_1 \)-\( L_\infty \) contraction. Denote by \( A_n \) the average \( \frac{1}{n} \sum_{i=0}^{n-1} T^i \). The following statements are equivalent:

1. The set \( \mathcal{N} = \text{cl}\{x-Tx: x \in L_2\} \) is located in \( L_2 \);
2. The sequence \((A_n)_{n \in \mathbb{N}}\) converges in \(L_2\);

3. For all \(p > 1\), the sequence \((A_n)_{n \in \mathbb{N}}\) converges in \(L_p\);

4. There is \(p > 1\) such that the sequence \((A_n)_{n \in \mathbb{N}}\) converges in \(L_p\);

5. For all \(p \geq 1\) and \(f \in L_p\), the sequence \((A_nf)_{n \in \mathbb{N}}\) converges a.e.;

6. There is \(p \geq 1\) such that for all \(f \in L_p\), the sequence \((A_nf)_{n \in \mathbb{N}}\) converges a.e.

For a finite measure we may replace \(p > 1\) by \(p \geq 1\).

**Proof.** (1) \(\Leftrightarrow\) (2) is Theorem 2. (4) \(\Rightarrow\) (3) is Theorem 9. (3) \(\Rightarrow\) (2) is trivial. (2) \(\Rightarrow\) (4) is trivial. (4) \(\Rightarrow\) (5) follows from Theorem 1 and Theorem 11. (5) \(\Rightarrow\) (6) is trivial. (6) \(\Rightarrow\) (1) follows from Theorem 12. \(\square\)

Some of these results have circulated in preprints for some time. They then appeared in my thesis [11]. The results have been used in [1] in the context of reverse mathematics.

I would like to thank Wim Veldman for his advice during the PhD-project. Finally, I would like to thank the referee for comments that helped to improve the presentation of the paper.

**Bibliography**


