Induced Transformations of Recurrent A.M.S. Dynamical Systems[*]

Sheng Huang[†] and Mikael Skoglund[‡]

Communication Theory Lab

School of Electrical Engineering

KTH Royal Institute of Technology

Stockholm, Sweden

First edition 10 October, 2013; revised 30 October, 2013 and 26 March, 2014

Abstract

This note proves that an induced transformation with respect to a finite measure set of a recurrent asymptotically mean stationary dynamical system with a sigma-finite measure is asymptotically mean stationary. Consequently, the Shannon–McMillan–Breiman Theorem, as well as the Shannon–McMillan Theorem, holds for all reduced processes of any finite-state recurrent asymptotically mean stationary random process.

As a byproduct, a ratio ergodic theorem for asymptotically mean stationary dynamical systems is presented.

Keywords: Induced transformation; asymptotically mean stationary; dynamical system; Shannon–McMillan–Breiman theorem; ratio ergodic theorem

MSC 2010 subject classifications: Primary 37A05; secondary 37A30, 94A15, 60G10, 37A50

[*]This work was funded in part by the Swedish Research Council.

[†]sheng.huang@ee.kth.se

[‡]skoglund@ee.kth.se

1 Introduction

1.1 Asymptotically Mean Stationary

A dynamical system $(\Omega, \mathscr{F}, \mu, T)$ with a finite measure, e.g. probability measure, is said to be asymptotically mean stationary[1] (a.m.s.) [GK80] if the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}B\right)$$

exists for all $B \in \mathscr{F}$. As proved in [GK80], a system being a.m.s. is a necessary and sufficient condition to ensure that

$$\frac{1}{n} \sum_{i=0}^{n-1} fT^{i}$$

converges μ-almost everywhere (μ-a.e.) on Ω for every bounded $\mathscr{F}$-measurable real-valued function, say f. Let

$$\overline{\mu}(B) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}B\right), \forall\, B \in \mathscr{F},$$

$$\text{and } \overline{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} fT^{i}(x), \forall\, x \in \Omega.$$

Then, by the Vitali–Hahn–Saks Theorem, it is easily seen that $\overline{\mu}$ is a finite measure on $(\Omega, \mathscr{F})$, and $\overline{f}$ is $\mathscr{F}$-measurable. Moreover, $(\Omega, \mathscr{F}, \overline{\mu}, T)$ is invariant, in other words, T is a measure preserving transformation on $(\Omega, \mathscr{F}, \overline{\mu})$, i.e.

$$\overline{\mu}(B) = \overline{\mu}\left(T^{-1}B\right), \forall\, B \in \mathscr{F},$$

and $\overline{f}$ is T-invariant a.e., i.e.

$$\overline{f} = \overline{f}T \text{ a.e.},$$

with respect to both μ and $\overline{\mu}$. In fact, $\overline{f}$ is simply the conditional expectation $E_{\overline{\mu}}(f|\mathscr{I})$, where $\mathscr{I} \subseteq \mathscr{F}$ is the σ-algebra of T-invariant sets ($B \in \mathscr{F}$ is said to be T-invariant if

[1] Perhaps it is better to replace "stationary" with "invariant," because a stationary measure defined in [GK80] is usually called an invariant measure in the language of ergodic theory. However, in order to be consistent, we will follow existing literature and use the terminology "asymptotically mean stationary," while the reader can read it as "asymptotically mean invariant" if preferred.

$B = T^{-1}B$). Therefore, if $(\Omega, \mathscr{F}, \mu, T)$ is ergodic, i.e.

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } \mu(\Omega - B) = 0, \forall\, B \in \mathscr{F},$$

then $\overline{f} = E_{\overline{\mu}}(f|\mathscr{I})$ equals to a constant a.e. with respect to both μ and $\overline{\mu}$.

We emphasize that the definition (cited from [GK80]) of the a.m.s. property given above is only valid for finite measures. In order to address dynamical systems with non-finite measures, in particular those with σ-finite measures, we generalise the definition as follows.

Definition 1.1. A dynamical system $(\Omega, \mathscr{F}, \mu, T)$ is said to be *asymptotically mean stationary* (*a.m.s.*) if there exists a measure $\overline{\mu}$ on $(\Omega, \mathscr{F})$ satisfying:

1. For any $B \in \mathscr{F}$ of finite measure, i.e. $\mu(B) < \infty$,

$$\overline{\mu}(B) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}B);$$

2. For any T-invariant set $B \in \mathscr{F}$, $\overline{\mu}(B) = \mu(B)$.

Such a measure $\overline{\mu}$ is named the *invariant mean*[2] of μ.

The following proposition clearly explains why the terminology "asymptotically mean stationary" and "invariant mean" are suggested.

Proposition 1.2. *Let $(\Omega, \mathscr{F}, \mu, T)$ be a.m.s. and $\overline{\mu}$ be an invariant mean of μ. If μ is σ-finite, then $(\Omega, \mathscr{F}, \overline{\mu}, T)$ is invariant.*

Proof. For any $B \in \mathscr{F}$, if $\mu(B) < \infty$ obviously $\overline{\mu}(B) = \overline{\mu}\left(T^{-n}B\right)$ for any positive integer n. If $\mu(B) = \infty$, then there exists a countable partition $\{B_i : i \in \mathbb{N}^+\}$, with $\mu(B_i) < \infty$, of B since μ is σ-finite. Moreover, $\left\{T^{-1}B_i\right\}$ is a countable partition of $T^{-1}B$, and $\overline{\mu}(B_i) = \overline{\mu}\left(T^{-1}B_i\right) < \infty$ for all feasible i. As a consequence,

$$\overline{\mu}\left(T^{-1}B\right) = \sum_{i=1}^{\infty} \overline{\mu}\left(T^{-1}B_i\right) = \sum_{i=1}^{\infty} \overline{\mu}(B_i) = \overline{\mu}(B).$$

Hence, $(\Omega, \mathscr{F}, \overline{\mu}, T)$ is invariant. □

[2] In [GK80], the term "stationary mean" is used instead of "invariant mean" for a finite measure μ.

Remark 1. Obviously, an invariant system $(\Omega, \mathscr{F}, m, T)$ is a.m.s. with m being the invariant mean of itself. Actually, if μ in Definition 1.1 is finite, then the second requirement in the definition is redundant, because the fact $\overline{\mu}(B) = \mu(B)$ for any T-invariant set B can be deduced from the first requirement. Therefore, Definition 1.1 covers the original definition from [GK80] as a special case. However, for a non-finite measure, the second condition is crucial.

Example 1.3. Let $\mathbb{R}^+ = (0, +\infty)$, $\mathscr{B}$ be the Borel σ-algebra on $\mathbb{R}^+$, μ be the Lebesgue measure on $(\mathbb{R}^+, \mathscr{B})$, and $T(x) = x^2, \forall\, x \in \mathbb{R}^+$. For set function $\lambda : \mathscr{B} \to \mathbb{R}$ given by:

1. For all $B \in \mathscr{B}$ with $\mu(B) < \infty$, $\lambda(B) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}B\right);$
2. For all $B \in \mathscr{B}$ with $\mu(B) = \infty$, $\lambda(B) = \sum_{i=1}^{\infty} \lambda(B_i)$, where $\mu(B_i) < \infty$ and the B_i's form a countable partition of B.

It is easy to verify that λ is well-defined. In exact terms, for any measurable set B with $\mu(B) = \infty$ and any two countable partitions $\{B_i'\}$ and $\{B_i''\}$, where $\mu(B_i') < \infty$ and $\mu(B_i'') < \infty$, of B,

$$\sum_{i=1}^{\infty} \lambda(B_i') = \sum_{i=1}^{\infty} \lambda(B_i'').$$

In addition, one can also prove that λ is a finite, hence σ-finite, measure over $(\mathbb{R}^+, \mathscr{B})$, since $\lambda(\mathbb{R}^+) = 1$. However, λ is not an invariant mean of μ, because $[1, +\infty)$ is a T-invariant set while

$$\mu([1, +\infty)) = \infty \neq 0 = \lambda([1, +\infty)).$$

From this one sees that $(\mathbb{R}^+, \mathscr{B}, \mu, T)$ is not a.m.s.. To prove this by contradiction, suppose $\overline{\mu}$ is an invariant mean of μ, then

$$\begin{aligned}
\overline{\mu}([1, +\infty)) &= \sum_{j=1}^{\infty} \overline{\mu}([j, j+1)) \\
&= \sum_{j=1}^{\infty} \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}[j, j+1)\right) = \sum_{j=1}^{\infty} 0 = 0 \\
&\neq \infty = \mu([1, +\infty)).
\end{aligned}$$

1.2 Induced Transformations

For an invariant system $(\Omega, \mathscr{F}, m, T)$ with a finite measure m, Poincaré's Recurrence Theorem guarantees that

$$m\left(B - \bigcap_{i=0}^{\infty}\bigcup_{j=i}^{\infty} T^{-j}B\right) = 0, \forall\, B \in \mathscr{F}. \tag{1}$$

As a consequence, for any $A \in \mathscr{F}$ $(m(A) > 0)$, one can define a new transformation T_A on $(A_0, \mathscr{A}, m|_{\mathscr{A}})$, where $A_0 = A \cap \bigcap_{i=0}^{\infty}\bigcup_{j=i}^{\infty} T^{-j}A$ and $\mathscr{A} = \{A_0 \cap B | B \in \mathscr{F}\}$, such that

$$T_A(x) = T^{\psi_A^{(1)}(x)}(x), \forall\, x \in A_0,$$

where

$$\psi_A^{(1)}(x) = \min\left\{i \in \mathbb{N}^+ | T^i(x) \in A_0\right\}$$

is the first return time function. Consequently, $(A_0, \mathscr{A}, m|_{\mathscr{A}}, T_A)$ forms a new dynamical system. Such a transformation T_A is called an induced transformation of $(\Omega, \mathscr{F}, m, T)$ (with respect to A) [Kak43].

On the other hand, for an arbitrary a.m.s. dynamical system $(\Omega, \mathscr{F}, \mu, T)$, the situation (of defining the concept of induced transformation) becomes delicate, because (1) is not necessarily valid even for a finite measure μ, unless $\mu \ll \overline{\mu}$ [Gra09, Theorem 7.4]. Thus, there could be some $A \in \mathscr{F}$ of positive measure, such that T_A is not defined on any non-empty subset of A. To avoid a situation of this sort, we shall focus on dynamical systems for which (1) holds.

Definition 1.4. A dynamical system $(\Omega, \mathscr{F}, \mu, T)$ is said to be *recurrent* (*conservative*) if

$$\mu\left(B - \bigcap_{i=0}^{\infty}\bigcup_{j=i}^{\infty} T^{-j}B\right) = 0, \forall\, B \in \mathscr{F}.$$

Remark 2. There are several equivalent definitions of recurrence (conservativeness). Please refer to [Gra09, Chapter 7.4] and [Aar97] for more details.

It is well-known that, for a recurrent invariant system $(\Omega, \mathscr{F}, m, T)$ with m being σ-finite, $(A_0, \mathscr{A}, m|_{\mathscr{A}}, T_A)$ with $0 < m(A) < \infty$ is invariant. Unfortunately, the available proof of this result relies heavily on the invariance assumption. In other words, for more general systems, e.g. a.m.s. systems, the case is not yet settled. Thus, the solo purpose of this note is to prove that, if $(\Omega, \mathscr{F}, \mu, T)$ is a recurrent a.m.s. dynamical system with μ being σ-finite, then $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ is also a.m.s. for all $0 < \mu(A) < \infty$. At the same time, a connection between the invariant mean of $\mu|_{\mathscr{A}}$ and μ is established (see Theorem 2.1 and Theorem 3.4).

As a direct conclusion of this assertion, we have that the Shannon–McMillan–Breiman Theorem, as well as the Shannon–McMillan Theorem, holds for all reduced processes of any finite-state recurrent a.m.s. random process (see Section 4).

2 Finite Measure μ

We first prove the assertion for dynamical systems equipped with finite measures.

To facilitate our discussion, we designate 1_A as the indicator function of a set $A \subseteq \Omega$. To be precise, $1_A(x) = \begin{cases} 1; & \text{if } x \in A; \\ 0; & \text{if } x \in \Omega - A. \end{cases}$

Theorem 2.1. *For a recurrent a.m.s. dynamical system $(\Omega, \mathscr{F}, \mu, T)$ with a finite measure μ and any $A \in \mathscr{F}$ with $\mu(A) > 0$, $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ is a.m.s.. Moreover, the invariant mean $\overline{\mu|_{\mathscr{A}}}$ of $\mu|_{\mathscr{A}}$ admits*

$$\overline{\mu|_{\mathscr{A}}}(B) = \int_A \frac{\overline{1}_B}{\overline{1}_A} d\mu, \forall\, B \in \mathscr{A}.$$

Remark 3. The integral in Theorem 2.1 implicitly implies that $\overline{1}_A \neq 0$ μ-a.e. on A, as we will prove later (see Lemma 2.3). Besides, as mentioned, $\overline{1}_A = E_{\overline{\mu}}(1_A|\mathscr{I})$ and $\overline{1}_B = E_{\overline{\mu}}(1_B|\mathscr{I})$, where $\mathscr{I}$ is the σ-algebra of T-invariant sets, μ-a.e. and $\overline{\mu}$-a.e. on Ω. Therefore,

$$\overline{\mu|_{\mathscr{A}}}(B) = \int_A \frac{E_{\overline{\mu}}(1_B|\mathscr{I})}{E_{\overline{\mu}}(1_A|\mathscr{I})} d\mu, \forall\, B \in \mathscr{A}.$$

To prove Theorem 2.1, a couple of supporting lemmas are required.

Lemma 2.2. *Let $(\Omega, \mathscr{F}, \mu, T)$ (μ is not necessarily finite) be an arbitrary dynamical* *system. For any* $A \subseteq \Omega$ *and* $x \in \Omega$ *for which the limit* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 1_A T^i(x)$ *exists, let*

$$O = \left\{\omega \in \Omega | \overline{1}_A(\omega) = 0\right\}.$$

We have that the limit $\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} 1_{A-O} T^i(x)$ *exists and* $\overline{1}_{A-O}(x) = \overline{1}_A(x)$.

Proof. By definition,

$$\frac{1}{n}\sum_{i=0}^{n-1} 1_A T^i(x) = \frac{1}{n}\sum_{i=0}^{n-1} 1_{A-O} T^i(x) + \frac{1}{n}\sum_{i=0}^{n-1} 1_{A\cap O} T^i(x).$$

If $T^i(x) \notin A \cap O$ for all $i \in \mathbb{N}$, then $\frac{1}{n}\sum_{i=0}^{n-1} 1_{A\cap O} T^i(x)$ constantly equals to 0. Otherwise, $T^{i_0}(x) \in A \cap O$ for some i_0. Let $k = \min\{i \in \mathbb{N} | T^i(x) \in A \cap O\}$ and $y = T^k(x)$. Then, for all $n > k$, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} 1_{A\cap O} T^i(x) =& \frac{1}{n}\sum_{i=k}^{n-1} 1_{A\cap O} T^i(x) \\
=& \frac{1}{n}\sum_{i=0}^{n-k-1} 1_{A\cap O} T^i(y) \\
\leq& \frac{1}{n}\sum_{i=0}^{n-k-1} 1_A T^i(y) \\
\to& 0, n \to \infty,
\end{aligned}$$

since $y \in O$. Therefore, $\frac{1}{n}\sum_{i=0}^{n-1} 1_{A-O} T^i(x) \to \overline{1}_A(x)$, $n \to \infty$. □

Lemma 2.3. *In Theorem 2.1, we have that*

$$\overline{1}_A \neq 0 \ a.e. \ on \ A,$$

with respect to both μ *and its invariant mean* $\overline{\mu}$.

Proof. Let $O = \left\{\omega \in \Omega | \overline{1}_A(\omega) = 0\right\}$. We get

$$\begin{aligned}
\overline{\mu}(A) =& \int_\Omega \overline{1}_A d\mu && (2)\\
=& \int_\Omega \overline{1}_{A-O} d\mu && (3)\\
=& \overline{\mu}(A - O). && (4)
\end{aligned}$$

where (2) and (4) are due to the fact that $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s. [Gra09, Corollary 7.9], and (3) follows from Lemma 2.2. Consequently, $\overline{\mu}(A \cap O) = 0$. Since $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s. and recurrent, we have that $\mu \ll \overline{\mu}$ by [Gra09, Theorem 7.4]. Therefore, $\mu(A \cap O) = 0$. $\square$

Proof of Theorem 2.1. For any $x \in A_0$ and positive integer n, let

$$\psi_A^{(n)}(x) = \sum_{i=0}^{n-1} \psi_A^{(1)}(T_A^i(x)).$$

It is easy to see that $\psi_A^{(n)}$ (the nth return time function) is well-defined since the system is recurrent. For any $B \in \mathscr{A}$, we have that

$$\begin{aligned}
\frac{1}{n}\sum_{i=0}^{n-1} \mu\left(T_A^{-i}B\right) &= \int_{A_0} \frac{1}{n}\sum_{i=0}^{n-1} 1_B T_A^i(\omega) d\mu(\omega) \\
&= \int_{A} \frac{1}{n}\sum_{i=0}^{n-1} 1_B T_A^i(\omega) d\mu(\omega) \qquad (5)\\
&= \int_{A} \frac{1}{n}\sum_{i=0}^{\psi_A^{(n)}(\omega)-1} 1_B T^i(\omega) d\mu(\omega) \\
&= \int_{A} \frac{\psi_A^{(n)}(\omega)}{n} \frac{1}{\psi_A^{(n)}(\omega)}\sum_{i=0}^{\psi_A^{(n)}(\omega)-1} 1_B T^i(\omega) d\mu(\omega),
\end{aligned}$$

where (5) follows because $\mu(A - A_0) = 0$ since the system is recurrent. Due to the fact that $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s., it follows that

$$\begin{aligned}
\frac{n}{\psi_A^{(n)}(\omega)} = \frac{1}{\psi_A^{(n)}(\omega)} &\sum_{i=0}^{\psi_A^{(n)}(\omega)-1} 1_A T^i(\omega) \to \overline{1}_A(\omega) \ \mu\text{-a.e. and} \\
\frac{1}{\psi_A^{(n)}(\omega)} &\sum_{i=0}^{\psi_A^{(n)}(\omega)-1} 1_B T^i(\omega) \to \overline{1}_B(\omega) \ \mu\text{-a.e.}
\end{aligned}$$

as $n \to \infty$. Let $O = \{\omega \in \Omega | \overline{1}_A(\omega) = 0\}$. We conclude that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu\left(T_A^{-i}B\right) \\
&= \lim_{n\to\infty} \int_{A-O} \frac{\psi_A^{(n)}(\omega)}{n} \frac{1}{\psi_A^{(n)}(\omega)}\sum_{i=0}^{\psi_A^{(n)}(\omega)-1} 1_B T^i(\omega) d\mu(\omega) \qquad (6)\\
&= \int_{A-O} \frac{\overline{1}_B}{\overline{1}_A} d\mu = \int_{A} \frac{\overline{1}_B}{\overline{1}_A} d\mu, \qquad (7)
\end{aligned}$$

where (6) is due to the fact that $\mu(A \cap O) = 0$ by Lemma 2.3 and (7) follows from the Dominated Convergence Theorem [Rud86]. The theorem is established. □

Corollary 2.4. *If* $(\Omega, \mathscr{F}, \mu, T)$ *in Theorem 2.1 is ergodic, then*

$$\overline{\mu|_{\mathscr{A}}}(B) = \frac{\mu(A)\overline{\mu}(B)}{\overline{\mu}(A)}, \forall\, B \in \mathscr{A}.$$

Proof. If $(\Omega, \mathscr{F}, \mu, T)$ is ergodic, then $\overline{1}_A = \overline{\mu}(A)$ and $\overline{1}_B = \overline{\mu}(B)$ a.e. with respect to both μ and $\overline{\mu}$. The statement follows. □

Remark 4. By Corollary 2.4, the system $\left(A_0, \mathscr{A}, \frac{1}{\mu(A)}\mu|_{\mathscr{A}}, T_A\right)$ is a.m.s. and ergodic, and $\frac{1}{\mu(A)}\mu|_{\mathscr{A}}$ is a probability measure on $(A_0, \mathscr{A})$ with invariant mean $\frac{1}{\mu(A)}\overline{\mu|_{\mathscr{A}}} = \frac{1}{\overline{\mu}(A)}\overline{\mu}|_{\mathscr{A}}$.

For dynamical systems with finite measures, it is indeed quite natural to believe that an induced transformation of a recurrent a.m.s. system is also a.m.s., hinted by the fact that an induced transformation of an invariant system is invariant. However, as seen from the above, the proof for the case of a.m.s. systems does not follow naturally from the one for the invariant case [Aar97]. After all, the system is no longer invariant.

3 σ-finite Measure μ

In the previous section, the assumption that μ is finite is important, it comes into play in many places in our argument. This assumption supports the use of the Dominated Convergence Theorem in the proof of Theorem 2.1, and it is also a requirement to guarantee convergence (μ-a.e.) of the sample mean of a bounded measurable real-valued function. Consequently, if instead μ is not finite, our method proving Theorem 2.1 is not applicable. In this section, we will therefore prove our assertion for the case of a σ-finite measure based on a different approach, which involves the ratio ergodic theorem of [Hop70].

For convenience, we define $\mathbf{S}_n(f)$ to be the finite sum $\sum_{i=0}^{n-1} fT^i$, for some given transformation T, non-negative integer n and real-valued function f.

Theorem 3.1 (Ratio Ergodic Theorem for Invariant Systems[3])**.** *Let $(\Omega, \mathscr{F}, m, T)$ be an invariant dynamical system with m being σ-finite. For any $f, g \in L^1(m)$ such that $g \geq 0$ and $\int_\Omega g dm > 0$, there exists a function $h(f,g) : \Omega \to \mathbb{R}$, such that*

$$\lim_{n\to\infty} \frac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)} = h(f,g) \ m\text{-a.e. on } D = \left\{ \omega \in \Omega \,\middle|\, \sup_n \mathbf{S}_n(g)(\omega) = \infty \right\}.$$

Moreover, $h(f,g)$ is T-invariant m-a.e. on D, it is $\mathscr{I}$-measurable, where $\mathscr{I} \subseteq D \cap \mathscr{F}$ is the σ-algebra of T-invariant sets, and

$$\int_I f dm = \int_I h(f,g) g dm, \forall\, I \in \mathscr{I}.$$

To our knowledge, the first[4] general ergodic theorem for a.m.s. systems is the generalisation of Birkhoff's ergodic theorem [Bir31] presented in [GK80]. Coincidentally, there is a version of Hopf's ratio ergodic theorem for a.m.s. systems.

Theorem 3.2 (Ratio Ergodic Theorem for A.M.S. Systems)**.** *Given an a.m.s. dynamical system $(\Omega, \mathscr{F}, \mu, T)$ with μ being σ-finite, let $\overline{\mu}$ be the invariant mean of μ. For any $f, g \in L^1(\overline{\mu})$ such that $g \geq 0$ and $\int_\Omega g d\overline{\mu} > 0$, there exists a function $h(f,g) : \Omega \to \mathbb{R}$, such that*

$$\begin{cases} \lim\limits_{n\to\infty} \dfrac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)} = h(f,g); \\ h(f,g) = h(f,g)T \end{cases} \quad \text{a.e. on } D = \left\{ \omega \in \Omega \,\middle|\, \sup_n \mathbf{S}_n(g)(\omega) = \infty \right\}$$

with respect to both μ and $\overline{\mu}$. Moreover, if $(\Omega, \mathscr{F}, \mu, T)$ is ergodic, then either $\mu(D) = \overline{\mu}(D) = 0$ or

$$h(f,g) = \frac{\int_\Omega f d\overline{\mu}}{\int_\Omega g d\overline{\mu}} \ \mu\text{-a.e. and } \overline{\mu}\text{-a.e. on } \Omega.$$

Proof. By Theorem 3.1,

$$\lim_{n\to\infty} \frac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)} = h(f,g) \ \overline{\mu}\text{-a.e. on } D,$$

[3] Hopf's ratio ergodic theorem for invariant systems is often presented differently in the literature, in each instance with different and delicate details. Readers are kindly referred to the related literature ([Hop70, Ste36, KK97, Zwe04] and etc.) for more information.

[4] An earlier ergodic theorem from [Hur44] works for systems that are not necessarily invariant. However, that result relies on some additional constraints which, to the best of our knowledge, hinder an extension to a.m.s. systems.

for some function $h(f,g):\Omega\to\mathbb{R}$. Let

$$h^* = \overline{\lim_{n\to\infty}}\frac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)},$$
$$h_* = \underline{\lim_{n\to\infty}}\frac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)},$$

and define $D_l^u = \{x\in D\,|h^*(x)\geq u, h_*(x)\leq l\}$ for all $l,u\in\mathbb{Q}$. Obviously, D_l^u is T-invariant. Thus,

$$\mu(D_l^u) = \overline{\mu}(D_l^u) = 0, \forall\, l<u,$$

because $h^* = h_*$ $\overline{\mu}$-a.e. on D by Theorem 3.1. Consequently,

$$\mu\left(\{x\in D|h^*(x)>h_*(x)\}\right) = \mu\left(\bigcup_{l<u} D_l^u\right) \leq \sum_{l<u}\mu(D_l^u) = 0.$$

At every point $x\in D$ where the limit $\lim_{n\to\infty}\frac{\mathbf{S}_n(f)(x)}{\mathbf{S}_n(g)(x)}$ exists, it is obvious that

$$\lim_{n\to\infty}\frac{\mathbf{S}_n(f)(x)}{\mathbf{S}_n(g)(x)} = \lim_{n\to\infty}\frac{\mathbf{S}_n(f)(Tx)}{\mathbf{S}_n(g)(Tx)}.$$

Therefore, $h(f,g) = h(f,g)T$ a.e. on D with respect to both μ and $\overline{\mu}$. The last statement is valid due to ergodicity. □

Remark 5. In Theorem 3.2, $\int_\Omega g d\overline{\mu}>0$ can be replaced by $\int_\Omega g d\mu > 0$ if the system is recurrent. This is because $\int_\Omega g d\mu > 0 \implies \int_\Omega g d\overline{\mu} > 0$ by Lemma 3.3.

Lemma 3.3. *Given an a.m.s. dynamical system $(\Omega,\mathscr{F},\mu,T)$ with μ being σ-finite, let $\overline{\mu}$ be the invariant mean of μ. If $(\Omega,\mathscr{F},\mu,T)$ is recurrent, then $\mu\ll\overline{\mu}$.*

Proof. For any $B\in\mathscr{F}$ such that $\overline{\mu}(B)=0$, let $B_\infty = \bigcap_{i=0}^{\infty}\bigcup_{j=i}^{\infty}T^{-j}B$. We have that

$$0 = \sum_{j=0}^{\infty}\overline{\mu}(B) = \sum_{j=0}^{\infty}\overline{\mu}\left(T^{-j}B\right) \geq \overline{\mu}\left(\bigcup_{j=0}^{\infty}T^{-j}B\right) \geq \overline{\mu}(B_\infty)\geq 0.$$

Therefore, $\mu(B_\infty) = \overline{\mu}(B_\infty) = 0$ since B_∞ is T-invariant. Thus, $\mu(B) = \mu(B - B_\infty)$. Moreover, $\mu(B-B_\infty)=0$ by the definition of recurrence. As a conclusion, $\mu(B)=0$. □

Remark 6. Whenever μ is finite, the converse of Lemma 3.3 is also valid [Gra09, Theorem 7.4]. However, it is not necessarily true for a non-finite measure μ.

Theorem 3.4. *For a recurrent a.m.s. dynamical system $(\Omega, \mathscr{F}, \mu, T)$ with μ being σ-finite and any $A \in \mathscr{F}$ with $0 < \mu(A) < \infty$, $(A_0, \mathscr{A}, \mu|_{\mathscr{A}}, T_A)$ is a.m.s.. In particular, the invariant mean $\overline{\mu|_{\mathscr{A}}}$ of $\mu|_{\mathscr{A}}$ satisfies*

$$\overline{\mu|_{\mathscr{A}}}(B) = \int_A h(1_B, 1_A) d\mu, \forall\, B \in \mathscr{A},$$

where $h(1_B, 1_A) : \Omega \to \mathbb{R}$ satisfies

$$h(1_B, 1_A) = \lim_{n\to\infty} \frac{\mathbf{S}_n(1_B)}{\mathbf{S}_n(1_A)} \text{ a.e. on } D = \left\{ \omega \in \Omega \,\middle|\, \sup_n \mathbf{S}_n(1_A)(\omega) = \infty \right\} \tag{8}$$

with respect to both μ and $\overline{\mu}$.

Proof. First of all,

$$\int_\Omega 1_A d\mu = \mu(A) > 0 \implies \int_\Omega 1_A d\overline{\mu} > 0$$

by Lemma 3.3. Furthermore, since $\mu(B) \leq \mu(A) < \infty$ for any $B \subseteq \mathscr{A}$, we have that

$$\int_\Omega 1_B d\overline{\mu} = \overline{\mu}(B) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}B\right) < \infty \text{ and}$$
$$\int_\Omega 1_A d\overline{\mu} = \overline{\mu}(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T^{-i}A\right) < \infty$$

by definition. Therefore, there exists a function $h(1_B, 1_A) : \Omega \to \mathbb{R}$ satisfying (8) based on Theorem 3.2. Moreover, we have that

$$\begin{aligned} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T_A^{-i}B\right) &= \int_{A_0} \frac{1}{n} \sum_{i=0}^{n-1} 1_B T_A^i(\omega) d\mu(\omega) \\ &= \int_{A_0} \frac{\mathbf{S}_{k_n}(1_B)}{\mathbf{S}_{k_n}(1_A)} d\mu(\omega), \qquad (\text{where } k_n(\omega) = \phi_A^{(n)}(\omega)). \end{aligned}$$

Obviously, $0 \leq h(1_B, 1_A) \leq 1$ μ-a.e. and $\overline{\mu}$-a.e. on D because $1_B \leq 1_A$, and $A_0 \subseteq D$ by the definitions of A_0 and D. Since $\mu(A_0) = \mu(A) < \infty$, the Dominated Convergence Theorem [Rud86] ensures that

$$\begin{aligned} \overline{\mu|_{\mathscr{A}}}(B) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu\left(T_A^{-i}B\right) \\ &= \int_{A_0} h(1_B, 1_A) d\mu \\ &= \int_A h(1_B, 1_A) d\mu. \end{aligned}$$

The statement is proved. ∎

Remark 7. In the proof of Theorem 3.4, the condition $\mu(A) < \infty$ cannot be dropped, since it ensures that $1_A \in L^1(\overline{\mu})$, i.e. $\overline{\mu}(A) < \infty$.

Corollary 3.5. *In Theorem 3.4, if $(\Omega, \mathscr{F}, \mu, T)$ is ergodic, then*

$$\overline{\mu|_{\mathscr{A}}}(B) = \frac{\mu(A)\overline{\mu}(B)}{\overline{\mu}(A)}, \forall\, B \in \mathscr{A}.$$

Proof. Since $\mu(D) \geq \mu(A_0) = \mu(A) > 0$ and $(\Omega, \mathscr{F}, \mu, T)$ is ergodic, we have that $\mu(\Omega - D) = 0$ and

$$h(1_B, 1_A) = \frac{\int_\Omega 1_B d\overline{\mu}}{\int_\Omega 1_A d\overline{\mu}} = \frac{\overline{\mu}(B)}{\overline{\mu}(A)}\ \mu\text{-a.e.}, \forall\, B \in \mathscr{A},$$

by Theorem 3.2. The conclusion follows. □

4 The Shannon–McMillan–Breiman Theorem

Let $(\Omega, \mathscr{F}, \mu, T)$ be a dynamical system with μ being a probability measure, and X be a random variable with a finite sample space $\mathscr{X}$ defined on $(\Omega, \mathscr{F}, \mu)$. [GK80, Corollary 4] shows that the Shannon–McMillan–Breiman Theorem (the Shannon–McMillan Theorem) holds for the process

$$\{X_i\}_{i=0}^{\infty} = \left\{X\left(T^i\right)\right\}_{i=0}^{\infty},$$

if $\{X_i\}_{i=0}^{\infty}$ is a.m.s., i.e. $(\Omega, \mathscr{F}, \mu, T)$ is a.m.s..

In addition to being a.m.s., assume that $(\Omega, \mathscr{F}, \mu, T)$ is also recurrent. Given a subset $\mathscr{Y} \subseteq \mathscr{X}$ of positive probability, i.e. $\Pr\{X \in \mathscr{Y}\} > 0$, the reduced process $\{Y_j\}_{j=0}^{\infty}$ with sub-state space $\mathscr{Y}$ is defined to be

$$\{Y_j\}_{j=0}^{\infty} = \left\{X_{i_j}\right\}_{j=0}^{\infty},$$

where

$$i_j = \begin{cases} \min\{i \geq 0 | X_i \in \mathscr{Y}\}; & j = 0; \\ \min\{i > i_{j-1} | X_i \in \mathscr{Y}\}; & j > 0. \end{cases}$$

It is of interest to know whether the Shannon–McMillan–Breiman Theorem (the Shannon–McMillan Theorem) holds also for $\{Y_j\}_{j=0}^{\infty}$. Let $A = X^{-1}(\mathscr{Y})$. It is easily seen that

$$\{Y_j\}_{j=0}^{\infty} = \left\{X\left(T_A^j\right)\right\}_{j=0}^{\infty}$$

is essentially a random process defined on

$$\left(A_0, A_0 \cap \mathscr{F}, \frac{1}{\mu(A)} \mu|_{A_0 \cap \mathscr{F}}, T_A \right),$$

which is a.m.s. by Theorem 2.1 (by Theorem 3.4 as well). As a conclusion, the Shannon–McMillan–Breiman Theorem (the Shannon–McMillan Theorem) holds for the reduced process $\{Y_j\}_{j=0}^{\infty}$.

Theorem 4.1. *The Shannon–McMillan–Breiman Theorem, as well as the Shannon–McMillan Theorem, holds for all reduced processes of any recurrent a.m.s. random process of finite states.*

Proof. The result follows from Theorem 2.1 (Theorem 3.4 as well) and [GK80]. □

5 Motivation

Writing this note was inspired by the authors' work on algebraic source coding theory, which lead us to look deeper into the use of ergodic theory to characterize certain existence and optimality results in the case of sources (i.e., random processes) with memory.

It is well-known that linear coding over finite fields is optimal for all Slepian–Wolf data compression scenarios [Eli55, SW73, Cov75, Csi82]. Unfortunately, the same conclusion for linear coding over finite rings (non-field rings in particular) can not be proved in a similar manner as in the case of coding over fields. As detailed in [HS12], the main reason lies on the simple fact that a non-field ring contains non-invertible element(s). Consequently, the size of the kernel of a linear encoder (linear mapping), say $f : \mathfrak{R}^n \to \mathfrak{R}$ (where $\mathfrak{R}$ is a finite ring), is often strictly larger than $|\mathfrak{R}|^{n-1}$. The authors have, however, managed to provide alternative techniques that enabled us to prove that linear coding over non-field rings can also be optimal for compressing correlated i.i.d. data sources [HS12, HS13b, HS13c], as well as irreducible Markovian sources [HS13a]. Compared to the available literature, we needed to dig deeper into understanding the behavior of the sources, in particular whether the *asymptotically equipartition property*[5] holds for any reduced process of an i.i.d., as well as irreducible Markov, random process [HS13a]. By

[5]The asymptotically equipartition property is in essence a weaker version of the Shannon–McMillan–Breiman Theorem.

the use of ergodic theory, and by treating a stochastic source/process as a dynamical system, we discovered that the Shannon–McMillan–Breiman Theorem holds for any dynamical system derived via an induced transformation of the original source/process, provided that the source/process is i.i.d. or irreducible Markovian. This property allowed us to study coding over non-field rings and prove several new results. As a consequence, the flexible structure of rings, e.g. non-prime characteristic, existence of zero divisors and etc., are turned into strict advantages. We showed that linear coding over rings strictly outperforms its field counterpart in some network information theory problems (see [HS12, HS13c, HS13a] for more details).

In order to investigate the algebraic source coding problem in more general settings, e.g. for a.m.s. sources/processes, it is crucial to investigate the ergodic behavior of the dynamical system modeling the source. In particular, we are interested in questions relating to the validity of the Shannon–McMillian–Breiman Theorem. In fact, Theorem 4.1 of this note is an important foundation for generalizing the results reported in [HS12, HS13b, HS13c, HS13a] to settings involving a.m.s. sources.

References

[Aar97] Jon Aaronson. *An Introduction to Infinite Ergodic Theory*. American Mathematical Society, Providence, R.I., 1997.

[Bir31] George D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America*, 17(12):656–660, December 1931.

[Cov75] Thomas M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Transactions on Information Theory*, 21(2):226–228, March 1975.

[Csi82] Imre Csiszár. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, 28(4):585–592, July 1982.

[Eli55] P. Elias. Coding for noisy channels. *IRE Convention Record*, 3:37–46, March 1955.

[GK80] Robert M. Gray and J. C. Kieffer. Asymptotically mean stationary measures. *The Annals of Probability*, 8(5):962–973, October 1980.

[Gra09] Robert M. Gray. *Probability, Random Processes, and Ergodic Properties*. Springer, 2nd edition, August 2009.

[Hop70] Eberhard Hopf. *Ergodentheorie (Ergebnisse der Mathematik und Ihrer Grenzgebiete / Zweiter Band) (German Edition)*. Springer, reprint der erstausgabe berlin 1937 edition edition, January 1970.

[HS12] Sheng Huang and Mikael Skoglund. *On Linear Coding over Finite Rings and Applications to Computing*. KTH Royal Institute of Technology, October 2012, http://people.kth.se/~sheng11.

[HS13a] Sheng Huang and Mikael Skoglund. *Encoding Irreducible Markovian Functions of Sources: An Application of Supremus Typicality*. KTH Royal Institute of Technology, May 2013, http://people.kth.se/~sheng11.

[HS13b] Sheng Huang and Mikael Skoglund. On achievability of linear source coding over finite rings. In *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1984–1988, 2013.

[HS13c] Sheng Huang and Mikael Skoglund. On existence of optimal linear encoders over non-field rings for data compression with application to computing. In *2013 IEEE Information Theory Workshop (ITW)*, 2013.

[Hur44] Witold Hurewicz. Ergodic theorem without invariant measure. *Annals of Mathematics*, 45(1):192–206, January 1944.

[Kak43] Shizuo Kakutani. Induced measure preserving transformations. *Proceedings of the Imperial Academy*, 19(10):635–641, 1943.

[KK97] Teturo Kamae and Michael Keane. A simple proof of the ratio ergodic theorem. *Osaka Journal of Mathematics*, 34(3):653–657, 1997.

[Rud86] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill Science/Engineering/Math, 3rd edition, May 1986.

[Ste36] W. Stepanoff. Sur une extension du théorème ergodique. *Compositio Mathematica*, 3:239–253, 1936.

[SW73] David Slepian and Jack K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, July 1973.

[Zwe04] Roland Zweimüller. Hopf's ratio ergodic theorem by inducing. *Colloquium Mathematicum*, 101(2):289–292, 2004.