Communications in Information and Systems
Volume 19, Number 2, 147–165, 2019

Rate-distortion functions of non-stationary Markoff chains and their block-independent approximations

Mukul Agarwal

It is proved that the limit of the normalized rate-distortion functions of block independent approximations of a time-homogeneous, finite-state irreducible, aperiodic Markoff chain is independent of the initial distribution of the Markoff chain and thus, is also equal to the rate-distortion function of the Markoff chain.

1. Introduction

Consider a random source which evolves on a finite set. It follows from existing literature, see for example [1] and [2] (Pages 491-500, in particular, see Definition (9.8.3) and Theorem 9.8.3 for achievability), that the limit of the normalized rate-distortion functions of block-independent approximations of a stationary, ergodic source is equal to the rate-distortion function of the source. Specializing this theorem to time-homogeneous, irreducible, aperiodic Markoff chains, it follows that the limit of rate-distortion functions of block-independent approximations of an irreducible, aperiodic Markoff chain which starts in the stationary distribution is equal to the rate-distortion function of this Markoff chain. In what follows, all Markoff chains considered will be time homogeneous, and that, the Markoff chain is time homogeneous will not be stated explicitly again. It is known that the rate-distortion function of an irreducible, aperiodic Markoff chain is independent of its initial distribution (follows from [3]). In this paper, it will be proved that the limit of the normalized rate-distortion functions of block-independent approximations of an irreducible, aperiodic Markoff chain is independent of its initial distribution. It follows, then, that the rate-distortion function of an irreducible, aperiodic Markoff chain and the limit of the normalized rate-distortion functions of its block independent approximations are equal and these functions are independent of the initial distribution of the Markoff chain.

The intuition behind these results is that for an irreducible, aperiodic, finite state Markoff chain, the distribution of the Markoff chain at a certain time tends to the stationary distribution at a rate independent of the initial distribution, and the evolution of the Markoff chain for the finite amount of time when the distribution gets 'close enough' to the stationary distribution can be 'neglected' for the purpose of calculation of the rate-distortion function. Then, a limiting argument so that the distribution of the Markoff chain tends to the stationary distribution does the job. In the proofs, we also use the fact that the rate of convergence is, in fact, exponential in the time elapsed.

Literature on rate-distortion theory is vast. The seminal works are [1] and [4]. A work for rate-distortion theory for random processes is [5]. Much of the classical point-to-point literature on rate distortion theory gets subsumed under the books [2] and [3]. Another reference is [6]. The reader is refered to these three books and references therein for the literature on rate-distortion theory. In particular, the reader is referred to [3] because non-stationary sources are dealt with in great detail in this book, and the concern here is with a non-stationary process, albeit, a non-stationary Markoff chain. For understanding Markoff chains, the reader is referred to [7], [8], and [9], and for understanding convergence rates of Markoff chains, the reader is referred to [10].

2. Notation and definitions

$\mathbb{X}$ and $\mathbb{Y}$ denote the source input and source reproduction spaces respectively. Both are assumed to be finite sets. Asume that $\mathbb{X} = \mathbb{Y}$. Assume that the cardinality of $\mathbb{X}$ is greater than or equal to 2. $d : \mathbb{X} \times \mathbb{Y} \to [0, \infty)$ is the single-letter distortion measure. Assume that $d(x, x) = 0 \ \forall x \in \mathbb{X}$ and that $d(x, y) > 0$ if $x \neq y$. Denote

$$D_{\max} \triangleq \max_{x \in \mathbb{X}, y \in \mathbb{Y}} d(x, y), D_{\min} \triangleq \min_{\{x \in \mathbb{X}, y \in \mathbb{Y} | d(x,y) > 0\}} d(x, y) \tag{1}$$

In what follows, the distortion levels will be assumed to be strictly greater than 0. For $x^n \in \mathbb{X}^n, y^n \in \mathbb{Y}^n$, the n-letter rate-distortion measure is defined additively:

$$d^n(x^n, y^n) \triangleq \sum_{i=1}^{n} d(x^n(i), y^n(i)) \tag{2}$$

where $x^n(i)$ denotes the i^{th} component of x^n and likewise for y^n.

Let $X_1, X_2, \ldots,$ be a Markoff chain with transition probability matrix P, where each X_i is a random-variable on $\mathbb{X}$. For $x, x' \in \mathbb{X}$, $p_{xx'}$ denotes the probability that the Markoff chain is in state x' at time $t+1$ given that it is in state x at time t. Let $p_{xx'}$ be independent of t, that is, the Markoff chain is time homogeneous. Assume that the Markoff chain is irreducible, aperiodic. This implies that it has a stationary distribution, henceforth denoted by π, which will be reserved exclusively for the stationary distribution. In order to specify the Markoff chain completely, we need to specify its initial distribution. If $X_1 \sim \pi'$ denote the Markoff chain $(X_1, X_2, \ldots)$ by $X_{[\pi',P]}$. Recall that P is the transition probability matrix of the Markoff chain. $X_{[\pi',P]}$ will be called the Markoff $X_{[\pi',P]}$ chain. $X^n_{[\pi',P]}$ will denote $(X_1, X_2, \ldots, X_n)$.

The above mentioned assumptions that $\mathbb{X} = \mathbb{Y}$, $d(x, x) = 0$ and $d(x, y) > 0$ is $x \neq y$, that the distortion levels are strictly greater than zero, and that, the Markoff chain is irreducible, aperiodic, will be made throughtout this paper and will not be re-stated.

A rate R source-code is a sequence:
$< e^n, f^n >_1^\infty$, where $e^n : \mathbb{X}^n \to \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$ and $f^n : \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \to \mathbb{Y}^n$.

We say that rate R is achievable for source-coding the Markoff $X_{[\pi',P]}$ source within distortion-level D under the expected distortion criterion if there exists a rate R source code $< e^n, f^n >_1^\infty$ such that

$$\limsup_{n\to\infty} E\left[\frac{1}{n} d^n(X^n_{\pi'}, f^n(e^n(X^n_{[\pi',P]})))\right] \leq D \tag{3}$$

The infimum of all achievable rates is the rate-distortion function $R^E_{X_{[\pi',P]}}(D)$.

The block-independent approximation (henceforth shortened to BIA) $X^T_{[\pi',P]}$ source is a sequence of random vectors $(S_1, S_2, \ldots, S_n, \ldots)$, where S_i are independent, and $\forall i$, $S_i \sim X^T_{[\pi',P]}$. To simplify notation, we will sometimes denote $(S_1, S_2, \ldots)$ by S. S^n will denote $(S_1, S_2, \ldots, S_n)$. Note that BIA $X^T_{[\pi',P]}$ source is an i.i.d. vector source and will also be called the vector i.i.d. $X^T_{[\pi',P]}$ source. Since the BIA $X^T_{[\pi',P]}$ source is an i.i.d vector source, the rate-distortion function for it is defined in exactly the same way as for an i.i.d. source. The details are as follows: The source input space for the BIA $X^T_{[\pi',P]}$ source is $\mathbb{X}^T$ and the source reproduction space is $\mathbb{Y}^T$. Denote these by $\mathbb{S}$ and $\mathbb{T}$ respectively. A generic point in $\mathbb{S}$ is a T-length sequence s. The i^{th} component of s is denoted by $s(i)$. A generic point in $\mathbb{T}$ is a T-length sequence t. The i^{th} component of t is denoted by $t(i)$. The single letter distortion measure is denoted by d' and is defined as $d'(s, t) \triangleq \sum_{j=1}^T d(s(j), t(j))$. For $s^n \in \mathbb{S}^n$, $t^n \in \mathbb{T}^n$, the n-letter distortion measure d'^n is defined additively: $d'^n(s^n, t^n) \triangleq \sum_{i=1}^n d'(s^n(i), t^n(i))$. Note that s can be thought of as either a scalar in $\mathbb{S}$ or a T dimensional vector in

$\mathbb{X}^T$. With this identification, $d' = d^T$ and d'^n can be thought of as d^{nT}. A rate R source code is a sequence $< e^n, f^n >_1^\infty$, where $e^n : \mathbb{S}^n \to \{1, 2, \dots, 2^{\lfloor nR \rfloor}\}$ and $f^n : \{1, 2, \dots, 2^{\lfloor nR \rfloor}\} \to \mathbb{T}^n$. We say that rate R is achievable for source-coding the BIA $X^T_{[\pi', P]}$ source within distortion-level D under the expected distortion criterion if there exist a sequence of rate R source codes $< e^n, f^n >_1^\infty$ such that

$$\limsup_{n \to \infty} E\left[\frac{1}{n} d'^n(S^n, f^n(e^n(S^n)))\right] \leq D \tag{4}$$

The infimum of all achievable rates corresponding to a given distortion level D is the operational rate-distortion function at that distortion level, henceforth denoted by $R^E_{X^T_{[\pi', P]}}(D)$. The normalized rate-distortion function at block-length T and distortion level D is defined as

$$\frac{1}{T} R^E_{X^T_{[\pi', P]}}(TD) \tag{5}$$

and the limit is

$$\lim_{T \to \infty} \frac{1}{T} R^E_{X^T_{[\pi', P]}}(TD) \tag{6}$$

The theorems in this paper prove the equality of $R^E_{X_{[\pi', P]}}(D)$ and (6), and that these functions do not depend on π'. The statements of these theorems are stated in Section 3. Before that, we carry out a discussion on the rate-distortion function of a non-stationary Markoff chain.

3. Main result

Theorem 1 (Rate distortion function of a Markoff chain does not depend on the initial distribution)**.** $R^E_{X_{[\pi', P]}}(D) = R^E_{X_{[\pi, P]}}(D)$ *where* π *is the stationary distribution and* π' *is an arbitrary probability distribution on* $\mathbb{X}$

Proof. Follows from [3], Chapter 12, in particular, Section 12.8, by noting that a Markoff source which does not start in the stationary distribution is an AMS source. An independent proof tailored for Markoff chains can be found in Appendix A. □

Theorem 2 (The normalized rate-distortion functions of BIA of a Markoff chain is independent of the initial distribution)**.** *For* $D > 0$*,*

$$\lim_{T \to \infty} \frac{1}{T} R^E_{X^T_{[\pi', P]}}(TD) \text{ exists, and is independent of } \pi' \tag{7}$$

This theorem will be proved in Section 5.

Theorem 3 (The normalized rate-distortion function of BIA of a Markoff chain is equal to the rate-distortion function of the Markoff chain and both are independent of the initial distribution)**.**

(8)
$$R^E_{X_{[\pi',P]}}(D) = R^E_{X_{[\pi,P]}}(D) = \lim_{T\to\infty}\frac{1}{T}R^E_{X^T_{[\pi,P]}}(TD) = \lim_{T\to\infty}\frac{1}{T}R^E_{X^T_{[\pi',P]}}(TD)$$

where π is the stationary distribution and π' is an arbitrary distribution on $\mathbb{X}$.

Proof. Follows from Theorems 1, 2 and [2], Pages 490-500. □

In order to prove Theorem 2, we need more notation and this is the subject of the next section. The theorem is proved in the section following the next. In particular, $J_\tau(X^T_{[\pi',P]})$ is defined in order to make precise, the intuition stated in Section 1, and is the random variable corresponding to the Markoff chain after 'removing' the first τ times slots, and thus, as $\tau \to \infty$, for $T > \tau$, the initial distribution of the Markoff chain $J_{[\tau]}(X^T_{\pi',P})$ tends to the stationary distribution of the Markoff chain.

4. Further notation

The information-theoretic rate-distortion function of the vector i.i.d. $X^T_{[\pi',P]}$ source is denoted and defined as

$$R^I_{X^T_{[\pi',P]}}(D) \triangleq \inf_{\mathbb{W}} I(X^T;Y^T) \tag{9}$$

where $X^T \sim X^T_{[\pi',P]}$ and $\mathbb{W}$ is the set of $W : \mathbb{S} \to \mathbb{P}(\mathbb{T})$ (that is, W is a conditional PMF from $\mathbb{S}$ to $\mathbb{T}$) defined as

$$\mathbb{W} \triangleq \left\{ W \;\middle|\; \sum_{s\in\mathbb{S}, y\in\mathbb{T}} p_{X^T_{[\pi',P]}}(s)W(t|s)d'(s,t) \le D \right\} \tag{10}$$

where $p_{X^T_{[\pi',P]}}$ denotes the distribution corresponding to $X^T_{[\pi',P]}$. Note that this is the usual definition of the information-theoretic rate-distortion function for an i.i.d. source; just that the source under consideration is vector i.i.d. Recall, also, the rate-distortion function, that the information-theoretic rate-distortion function for an i.i.d. source is equal to the operational rate-distortion function of the source (requires some assumptions which are met in this paper because the input and reproduction spaces are finite).

Let $s \in \mathbb{S}$. Denote by V_τ the projection transformation. $V_\tau(s) \triangleq (s(\tau+1), s(\tau+2), \ldots s(T))$. Fix s. Denote $\mathbb{A} \triangleq \{t \in \mathbb{S} \mid V_\tau(t) = V_\tau(s)\}$. Under the distribution induced by $X^T_{[\pi',P]}$, the probability of the set $\mathbb{A}$ is

$$\pi'^{(\tau)}(s(\tau+1)) \prod_{i=\tau+1}^{T-1} p_{s(i)s(i+1)} \tag{11}$$

for some distribution $\pi'^{(\tau)}$ on $\mathbb{X}$ which satisfies $\pi'^{(\tau)}(x) \to \pi(x)$ as $\tau \to \infty$ $\forall x \in \mathbb{X}$. Note further, that if $\pi' = \pi$, $\pi'^{(\tau)} = \pi$. For $x \in \mathbb{X}$, denote $\pi'^{(\tau)}(x) = \pi(x) + \delta^{(\tau)}(x)$ where $\delta^{(\tau)}(x) \to 0$ as $\tau \to \infty$. $\delta^{(\tau)}(x)$ may be negative.

Denote by $J_\tau(X^T_{[\pi',P]})$, the probability distribution on $\mathbb{X}^{T-\tau}$ which causes the probability of a sequence $r \in \mathbb{X}^{T-\tau}$ to be

$$\pi'^{(\tau)}(r(1)) \prod_{i=1}^{T-1} p_{r(i)r(i+1)} \tag{12}$$

Note that $J_\tau(X^T_{[\pi',P]})$ is the marginal of $X^T_{[\pi',P]}$ on the last $T-\tau$ dimensions. An i.i.d. source can be formed from $J_\tau(X^T_{[\pi',P]})$ by taking a sequence of independent random vectors, each distributed as $J_\tau(X^T_{[\pi',P]})$. This will be called the vector i.i.d. $J_\tau(X^T_{[\pi',P]})$ source. The rate-distortion function for the vector i.i.d. $J_\tau(X^T_{[\pi',P]})$ source in defined in the same way as the rate-distortion function for the vector i.i.d. $X^T_{[\pi',P]}$ source: For $T-\tau$ length sequences, the single-letter distortion measure is defined as $d''(p,q) = \sum_{i=1}^{T-\tau} d(p(i), q(i))$ where $p \in \mathbb{X}^{T-\tau}$, $q \in \mathbb{Y}^{T-\tau}$. The n-letter rate-distortion measure is defined additively: $d''^n(p^n, q^n) = \sum_{i=1}^{n} d''(p^n(i), q^n(i))$ where $p^n \in (\mathbb{X}^{T-\tau})^n$ and $q^n \in (\mathbb{Y}^{T-\tau})^n$. A sequence of rate R source codes is a sequence $< e^n, f^n >_1^\infty$, where $e^n : (\mathbb{X}^{T-\tau})^n \to \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$ and $f^n : \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \to (\mathbb{Y}^{T-\tau})^n$. The rate-distortion functions for i.i.d. $J_\tau(X^T_{[\pi',P]})$ source when the distortion measure is d" is defined analogously as for the i.i.d. $X^T_{[\pi',P]}$ vector source; the details are omitted. Denote the operational rate-distortion function for the vector i.i.d. $J_\tau(X^T_{[\pi'',P]})$ source by $R^E_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and denote the information-theoretic rate-distortion function for the same source by $R^I_{J_\tau(X^T_{[\pi',P]})}(\cdot)$.

For the same reason as that stated before regarding d', $d'' = d^{T-\tau}$ and d''^n can be thought of as $d^{n(T-\tau)}$.

5. Proof of the Theorem 2

Before we prove the theorem note the following, which is a trivial consequence of the definition of convexity, see for example, [11]:

Lemma 1. *Let* $f : [0, \infty) \to [0, \infty)$ *be a convex* $\cup$ *non-increasing function. Let* $f(0) = K$*. Let* $0 < a < a'$*. Then,*

$$f(a) - f(a') \leq \frac{K}{a}(a' - a) \tag{13}$$

Proof.

$$\begin{aligned} &\text{Definition of convexity} \\ &\Rightarrow \left(1 - \frac{a}{a'}\right) f(0) + \frac{a}{a'} f(a') \geq f(a) \Rightarrow \frac{f(a') - f(a)}{a' - a} \geq \frac{f(a) - f(0)}{a - 0} \\ &\Rightarrow \frac{f(a) - f(a')}{a' - a} \leq \frac{f(0) - f(a)}{a} \Rightarrow f(a) - f(a') \leq \frac{K}{a}(a' - a) \end{aligned} \tag{14}$$

□

This lemma will be used crucially in the proof of the theorem, which follows below.

Proof of Theorem 2:

Proof. By the rate-distortion theorem, $R^E_{X^T_{[\pi',P]}}(TD) = R^I_{X^T_{[\pi',P]}}(TD)$. Comparing definitions with [2], Page 491,

$$\frac{1}{T} R^I_{X^T_{[\pi,P]}}(TD) \ \ \text{(notation in this document)} = R_T(D) \ \text{(notation in [2])} \tag{15}$$

In order to charify the above, what we mean is that our notation (left hand side of the above equation) is different from the notation used in [2]. The notation used in [2] is the right hand side of the above equation. We make this correspondence because we want to use results from [2] in what follows. By Theorem 9.8.1 in [2], it follows that

$$\lim_{T\to\infty} \frac{1}{T} R^E_{X^T_{[\pi,P]}}(TD) \text{ exists} \tag{16}$$

(16) will be used crucially towards the end of the proof.

The intuition stated in Section 1 is made precise here, in the following three steps:

1. Bound the difference between $R^E_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and $R^E_{J_\tau(X^T_{[\pi,P]})}(\cdot)$.
2. Relate $R^E_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and $R^E_{X^T_{[\pi',P]}}(\cdot)$.
3. Use these relations to prove the desired result by computing various bounds.

The first step in the proof is to come up with a bound for the difference between $R^E_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and $R^E_{J_\tau(X^T_{[\pi,P]})}(\cdot)$. To this end, we first do the same for $R^I_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and $R^I_{J_\tau(X^T_{[\pi,P]})}(\cdot)$. To this end, denote the distribution corresponding to $J_\tau(X^T_{[\pi',P]})$ on $\mathbb{X}^{T-\tau}$ by Q', and the distribution corresponding to $J_\tau(X^T_{[\pi,P]})$ by Q. The l^1 distance between Q' and Q,

(17)

$$\begin{aligned}
l^1(Q',Q) &\triangleq \sum_{x^{t-\tau}\in\mathbb{X}^{T-\tau}} \left|Q'(x^{T-\tau}) - Q(x^{T-\tau})\right| \\
&= \sum_{x^{t-\tau}\in\mathbb{X}^{t-\tau}} |\pi'^{(\tau)}(x^{t-\tau}(1) - \pi^{(\tau)}(x^{t-\tau}(1)| \prod_{i=1}^{T-\tau-1} p_{x^{t-\tau}(i)x^{t-\tau}(i+1)} \\
&= \sum_{x\in\mathbb{X}} |\delta^{(\tau)}(x)| \times 1 \\
&\triangleq \delta^{(\tau)}
\end{aligned}$$

In the above calculation, we have used the fact that if $\pi' = \pi$, $\pi'^{(\tau)} = \pi$.

Condition (Z) stated in [12] holds based on the assumptions we have made, Lemma 2 in [12] can be applied, and it follows that for τ sufficiently large (reasoning stated below after a few lines) and any $T > \tau$,

$$\left|\frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D) - \frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi,P]})}((T-\tau)D)\right| \leq \frac{1}{T-\tau}\frac{7d^*}{\tilde{d}}\delta^{(\tau)} \log\left(\frac{|\mathbb{X}|^{T-\tau}|\mathbb{Y}|^{T-\tau}}{\delta^{(\tau)}}\right) \tag{18}$$

In (18), $\delta^{(\tau)} \log \frac{1}{\delta^{(\tau)}}$ is defined as zero if $\delta^{(\tau)}$ is zero. $|\mathbb{X}^{T-\tau}|$ and $|\mathbb{Y}^{T-\tau}|$ denote the cardinalities of the input and output spaces on which the random source $J_\tau(X^T_{[\pi',P]})$ is defined. d^* is defined as

$$d^* \triangleq \max_{x^{T-\tau}\in\mathbb{X}^{T-\tau}, y^{T-\tau}\in\mathbb{Y}^{T-\tau}} d''(x^{T-\tau}, y^{T-\tau}) = (T-\tau)D_{\max} \tag{19}$$

and $\tilde{d}$ is defined as

$$\tilde{d} \triangleq \min_{\{x^{T-\tau}\in\mathbb{X}^{T-\tau}, y^{T-\tau}\in\mathbb{Y}^{T-\tau} \mid d''(x^{T-\tau},y^{T-\tau})>0\}} d''(x^{T-\tau}, y^{T-\tau}) = D_{\min} \tag{20}$$

It follows from (18) that

$$\begin{aligned} &\left|\frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D) - \frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi,P]})}((T-\tau)D)\right| \\ &\leq 7\frac{D_{\max}}{D_{\min}} \delta^{(\tau)} \log\left(\frac{|\mathbb{X}|^{T-\tau}|\mathbb{Y}|^{T-\tau}}{\delta^{(\tau)}}\right) \\ &= 7\frac{D_{\max}}{D_{\min}} \left[\delta^{(\tau)} \log\frac{1}{\delta^{(\tau)}} + (T-\tau)\log(|\mathbb{X}||\mathbb{Y}|)\delta^{(\tau)}\right] \end{aligned} \tag{21}$$

From Fact 3 and Fact 4 in [10], it follows, by noting that $\delta^{(\tau)} = \sum_{x\in\mathbb{X}} |\pi'^{(\tau)}(x) - \pi(x)| \leq |\mathbb{X}| \max_{x\in\mathbb{X}} |\pi'^{(\tau)}(x) - \pi(x)|$, that for sufficiently large τ,

$$\delta^{(\tau)} \leq |\mathbb{X}| C \tau^{J-1} \lambda_*^{\tau-J+1} \tag{22}$$

for some $\lambda_* < 1$, some constant J and some constant C. We will take $\tau = \sqrt{T}$. Also, we have stated above that (18) holds, but that this requires τ to be sufficiently large. This is because by Lemma 2 in [12], we need τ large enough so that

$$\delta^{(\tau)} \leq \frac{D_{\min}}{4D_{\max}(T-\tau)} \tag{23}$$

which is possible for T sufficiently large and $\tau = \sqrt{T}$, considering the fact that $\delta^{(\tau)} \to 0$ as $\tau \to \infty$ exponentially fast in τ by (22) and that, the polynomial factors in τ in the expression (22) for $\delta^{(\tau)}$ do not matter (see, for example, [11]), and this is another reason why we require τ to be sufficiently large. It follows, then, from (21), (22), and the equality of the information-theoretic and operational rate-distortion functions for i.i.d. sources, that with $\tau = \sqrt{T}$,

$$\begin{aligned} &\left|\frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D) - \frac{1}{T-\tau} R^I_{J_\tau(X^T_{[\pi,P]})}((T-\tau)D)\right| \\ &\triangleq \alpha_T \to 0 \text{ as } T \to \infty \text{ for some } \alpha_T \end{aligned} \tag{24}$$

where $\alpha_T \to 0$ as $T \to \infty$ follows for the same reason of exponential fall of $\delta^{(\tau)}$ with τ as stated above. The bound (24) will be used crucially later, towards the end of the proof.

Next step is to relate $R^E_{J_\tau(X^T_{[\pi',P]})}(\cdot)$ and $R^E_{X^T_{[\pi',P]}}(\cdot)$. We will argue the following:

$$R^E_{X^T_{[\pi',P]}}((T-\tau)D + \tau D_{\max}) \le R^E_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D) \tag{25}$$

and

$$R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \le R^E_{X^T_{[\pi',P]}}(TD) \tag{26}$$

Very rough idea to prove (25) is the following: Given a sequence of rate R source codes for the vector i.i.d. $J_\tau(X^T_{[\pi',P]})$ source, we can use the same sequence of rate R source-codes for the vector i.i.d. $X^T_{[\pi',P]}$ source by not coding the time-slots which were not projected onto when defining $J_\tau(X^T_{[\pi',P]})$. These banished slots will incur a maximum distortion of $\tau D_{\max}$ per symbol of $X^T_{[\pi',P]}$. (25) follows. See Appendix B for precise argument.

Very rough idea to prove (26) is the following: Consider a two-dimensional random vector (A, B) on some space and the i.i.d. source got by taking i.i.d. copies of (A, B). Consider a distortion measure which is additive over the two dimensions. Consider, also, the i.i.d. source formed by taking identical copies of A. Then, for a given distortion level, the rate-distortion function of the vector i.i.d. (A, B) source is greater than or equal to the rate-distortion function of the i.i.d. A source. This is stated more rigorously in Appendix B. Note that $J_\tau(X^T_{[\pi,P]})$ is a projection of $X^T_{[\pi,P]}$ onto certain dimensions and the distortion measure over these dimensions is additive. (26) follows from this.

Next, we get to Step 3. In what follows, $\tau = \sqrt{T}$. Assuming $TD > \tau D_{\max}$ (which holds for T sufficiently large if $\tau = \sqrt{T}$), by replacing D in (25) by

$$\frac{TD - \tau D_{\max}}{T - \tau} \tag{27}$$

and by (26), it follows that

$$R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \le R^E_{X^T_{[\pi',P]}}(TD) \le R^E_{J_\tau(X^T_{[\pi',P]})}(TD - \tau D_{\max}) \tag{28}$$

It follows from (28) by rearranging, that

$$0 \le R^E_{X^T_{[\pi',P]}}(TD) - R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \tag{29}$$

$$\leq R^E_{J_\tau(X^T_{[\pi',P]})}(TD - \tau D_{\max}) - R^E_{J_\tau(X^T_{[\pi',P]})}(TD)$$

By noting that $R^E_{J_\tau(X^T_{[\pi',P]})}(D)$ is a non-increasing, convex $\cup$ function of D which is upper bounded by $(T-\tau)\log|\mathbb{X}|$ at $D=0$, it follows, for T sufficiently large, by Lemma 1 that

(30)
$$R^E_{J_\tau(X^T_{[\pi',P]})}(TD - \tau D_{\max}) - R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \leq \tau D_{\max} \log|\mathbb{X}| \frac{T-\tau}{TD - \tau D_{\max}}$$

From (30) and (29), it follows that

$$\lim_{T\to\infty} \left| \frac{1}{T} R^E_{X^T_{[\pi',P]}}(TD) - \frac{1}{T} R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \right| = 0 \tag{31}$$

Note further, by noting that $R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \leq (T-\tau)\log|\mathbb{X}|$, that

(32)
$$\lim_{T\to\infty} \left| \frac{1}{T} R^E_{J_\tau(X^T_{([\pi',P]})}(TD) - \frac{1}{T-\tau} R^E_{J_\tau(X^T_{[\pi',P]})}(TD) \right| \leq \lim_{T\to\infty} \frac{\tau}{T(T-\tau)}(T-\tau)\log|\mathbb{X}| = 0$$

Also, by noting that $R^E_{J_\tau(X^T_{[\pi',P]})}(D)$ is a non-increasing, convex $\cup$ function of D which is upper bounded by $(T-\tau)\log|\mathbb{X}|$, it follows by use of Lemma 1 that

(33)
$$\lim_{T\to\infty} \left| \frac{1}{T-\tau} R^E_{J_\tau(X^T_{[\pi',P]})}(TD) - \frac{1}{T-\tau} R^E_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D) \right| \leq \lim_{T\to\infty} \frac{1}{T-\tau} \frac{(T-\tau)\log|\mathbb{X}|}{(T-\tau)D} \tau D \to 0 \text{ as } T\to\infty$$

It follows, then, from (31), (32), (33), by use of the triangle inequality, and by noting that

$$\lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n + \lim_{n\to\infty} c_n = \lim_{n\to\infty} (a_n + b_n + c_n) \tag{34}$$

if the three limits on the left hand side exist (follows from definitions, see for example [11]), that

$$\lim_{T\to\infty}\left|\frac{1}{T}R^E_{X^T_{[\pi',P]}}(TD)-\frac{1}{T-\tau}R^E_{J_\tau(X^T_{[\pi',P]})}((T-\tau)D)\right|=0 \tag{35}$$

From (24) and (35), it follows by the use of triangle inequality, that for T sufficiently large,

$$\left|\frac{1}{T}R^E_{X^T_{[\pi',P]}}(TD)-\frac{1}{T-\tau}R^E_{J_\tau(X^T_{[\pi,P]})}((T-\tau)D)\right|\le\alpha_T+\kappa_T \tag{36}$$

for some $\kappa_T\to 0$ as $T\to\infty$.

The above equation holds for $\pi'=\pi$ too, that is, for T sufficiently large,

$$\left|\frac{1}{T}R^E_{X^T_{[\pi,P]}}(TD)-\frac{1}{T-\tau}R^E_{J_\tau(X^T_{[\pi,P]})}((T-\tau)D)\right|\le\alpha_T+\eta_T \tag{37}$$

for some $\eta_T\to 0$ as $T\to\infty$.

From (36) and (37), by use of the triangle inequality, it follows, that for T sufficiently large,

$$\left|\frac{1}{T}R^E_{X^T_{[\pi',P]}}(TD)-\frac{1}{T}R^E_{X^T_{[\pi,P]}}(TD)\right|\le 2\alpha_T+\eta_T+\kappa_T \tag{38}$$

From (38) and (16), and by noting that $\alpha_T,\kappa_T,\eta_T\to 0$ as $T\to\infty$, it follows that

$$\lim_{T\to\infty}\frac{1}{T}R^E_{X^T_{[\pi',P]}}(TD)\text{ exists and is independent of }\pi' \tag{39}$$

This finishes the proof. □

The assumptions $\mathbb{X}=\mathbb{Y}$, $d(x,x)=0$, $d(x,y)>0$ if $x\neq y$ which have been made are not necessary, and can be replaced by weaker assumptions. Nothing is lost in terms of idea of the proof by making these assumptions, and making these assumptions prevents one from thinking of pathological cases; for these reasons they have been made.

6. Discussion and research directions

6.1. A note on the definition of the distortion incurred by a source-code

To be entirely correct, the distortion produced by a source-code for a Markoff source should be defined as follows: Let n be the block-length. Denote $U_i \triangleq$

$X_{(i-1)n+1}^{in}$. Each U_i is thus, a random vector of length n. Let $< e^n, f^n >_1^\infty$ be a source to code the source $X_{[\pi',P]}$. When the block length is n, we would like to use the source-code successively over all intervals of time of block-length n. Thus, it is more logical to define the distortion as:

$$\limsup_{n\to\infty} \sup_{i\in\mathbb{N}} E\left[\frac{1}{n} d^n(U_i, f^n(e^n(U_i)))\right] \tag{40}$$

and correspondingly define the rate-distortion function. The author believes (but it still requires a proof) that this does not change the theorems stated in this paper for the same reason as why the statements in this paper are true, that is, because, a Markoff chain converges to the stationary distribution at a rate which is independent of the initial distribution of the Markoff chain (and further, this rate of convergence is exponential as we shall see later, though this fact may not be needed in proving that this change of definition of distortion does not make a difference to the statements of the theorem). Note that if $\pi' = \pi$, the stationary distribution, the sup in the above definition can be removed since the distribution of $X_{(i-1)n+1}$ is independent of i.

6.2. Order m time homogeneous Markoff chains

In an order m time homogeneous Markoff chain, the distribution of X_i may depend on $X_{i-1}, X_{i-2}, \dots, X_{i-m}$ in a way which is independent of i. Consider an order m Markoff chain $X_{[\pi',P]} = (X_1, X_2, X_3, \dots)$ where now, the initial distribution is the distribution of $(X_1, X_2, \dots, X_m)$ and the transition matrix P is a transition matrix giving the distribution of X_i given $(X_{i-1}, X_{i-2}, \dots, X_{i-m})$. Define $Y_i = X_{(i-1)m+1}^{im}$. Then, $Y_{[\gamma',Q]} = (Y_1, Y_2, Y_3, \dots)$ is a Markoff chain where γ' and Q are defined in terms of π' and P respectively. Assume that $Y_{[\mu,Q]} = (Y_1, Y_2, Y_3, \dots)$ is irreducible, aperiodic. Then, Theorems 1, 2, 3 hold for the Markoff chain $Y_{[\gamma',Q]}$ and its block-independent approximations. Based on this, the author conjectures that Theorems 1, 2, 3 should also be true for the Markoff chain $X_{[\pi',P]}$ and its block independent approximations. The idea here is that $(X_1, X_2, \dots, X_{im} + r)$ where $r < m$ is the same as $(Y_1, Y_2, \dots, Y_m, X_{im+1}, X_{im+2}, \dots, X_{im+r})$ and the length of the sequence $(X_{im+1}, X_{im+2}, \dots, X_{im+r})$, that is, r is much less than the length of the sequence $(X_1, X_2, \dots X_{im}$ as i gets larger and larger. In order to prove the result rigorously, an 'interpolation' argument so that $(X_{im+1}, X_{im+2}, \dots, X_{im+r})$ can be incorporated after $(Y_1, Y_2, \dots, Y_m)$, is needed. Many of the ideas used in this paper would be likely needed to find such an interpolation argument. This is a conjecture at this point and requires proof; however, the author would be surprised, for the above reasons, if Theorems 1, 2, 3 do not hold for order m Markoff chains which satisfy the irreducibility, aperiodicity condition, as stated above.

6.3. ψ-mixing sources or a variant?

A set of sources to which this result may be generalizable with the proof technique used is ψ-mixing sources or close variants, appropriately defined. See [13], [14] and [15] for mixing of sources and [14], [15], in particular, for results on ψ-mixing sources. The main property (among others) that made ψ-mixing sources amenable to the result in [15] is the decomposition in Lemma 1 in [15], wherein, a stationary ψ-mixing source is written as a convex combination of an i.i.d. distribution and another general distribution where the i.i.d. distribution dominates as memory is lost with time. Precisely, the equation is Equation (19) in [15]:

$$\Pr(X_{t+\tau+1}^{t+\tau+T} \in \mathbb{B} | X_1^t \in \mathbb{A}) = (1-\lambda_\tau) P_T(\mathbb{B}) + \lambda_\tau P'_{t,\tau,T,\mathbb{A}}(\mathbb{B}) \tag{41}$$

where $\lambda_\tau \to 0$ as $\tau \to \infty$. This lemma, though, required stationarity. If a variant of (41) would hold for non-stationary sources, then, there is a possibility that the result in this paper be generalized to such sources. Irreducible, aperiodic Markoff chains statisfy this property, with $P_T(\mathbb{B})$ taken as the stationary distribution, and P' is some distribution depending on the initial distribution of the Markoff chain. An important bound in proving Theorem 2 in this paper is the l^1 distance between Q and Q', see (17). This result will hold for sources which satisfy (41) or a variant. Similarly, proving (25) and (26) in the proof of Theorem 2 or similar equations may also be possible. The rest of the proof of Theorem 2 is bounding various differences of ‘close by’ rate-distortion functions and this may be possible too. This is just an idea at this point and needs to be studied carefully to see if any of this is at all possible.

6.4. Other extensions

In addition to the extensions discussed in the previous two sub-sections, it would be worthwhile trying to generalize the theorem in this paper to ergodic sources to the extent possible; this would not only make the result general, but also shed light on the ‘internal workings’ of rate-distortion theory. Further, it would be worthwhile trying to prove this result using existing literature, in particular, see if it follows directly from some result, for example, in [3]; this would help with generalization and insight into the 'internal workings' of rate-distortion theory, too. It would also be worthwhile to explore whether the theorems in this paper hold for AMS sources, the definition for which can be found in [3]. Theory of large deviations might also have a role to play in the extensions to more general ergodic sources.

6.5. Recapitulation

In this paper, it was proved that the limit of the normalized rate-distortion functions of block independent approximations of an irreducible, aperiodic Markoff chain is independent of the initial distribution and is equal to the rate-distortion function of the Markoff chain. Various extensions have been discussed.

Acknowledgements

The author is infinitely grateful to Prof. Robert Gray for tens of hours of his time spent in insightful discussions with the author. The author also thanks one of the reviewers for pointing out an error in the proof of Theorem 2 (which has since been corrected), and to both the reviewers for their comments which significantly helped improve the presentation in the paper.

Appendix A. Proof of Theorem 1

Proof. The proof given here, makes precise, the intuition stated in Section 1. Consider two Markoff chains $X_{[\pi',P]} = (X'_1, X'_2, \ldots)$ and $X_{[\pi'',P]} = (X''_1, X''_2, \ldots)$, where π' and π'' probability distributions on $\mathbb{X}$. Denote $(X'_1, X'_2, \ldots, X'_n)$ by X'^n and $(X''_1, X''_2, \ldots, X''_n)$ by X''^n. Let τ be an integer. Think of n as large and τ to be much smaller than n. Denote the distribution of X'_τ by μ' and the distribution of X''_τ by μ''. Denote, $\epsilon' = \sum_{x \in \mathbb{X}} |\mu'(x) - \pi(x)|$ and $\epsilon'' = \sum_{x \in \mathbb{X}} |\mu''(x) - \pi(x)|$, where π is the stationary distribution of the Markoff chain (note that both Markoff chains have the same transition probability matrix P). For every $\epsilon > 0$, $\exists \tau^*_\epsilon$ such that $\epsilon' < \frac{\epsilon}{2}$ and $\epsilon'' < \frac{\epsilon}{2} \forall \tau \geq \tau^*_\epsilon$. Let $< e^n, f^n >_1^\infty$ be a source-code. Let the block-length be n. Think of n large and $\tau^*_\epsilon << n$. Use (e^n, f^n) to code $K'^n \triangleq (X'_{\tau^*_\epsilon+1}, X'_{\tau^*_\epsilon+2}, \ldots, X'_{\tau^*_\epsilon+n})$ and $K''^n \triangleq (X''_{\tau^*_\epsilon+1}, X''_{\tau^*_\epsilon+2}, \ldots, X''_{\tau^*_\epsilon+n})$. Denote the distribution of $K'^n(\tau^* + 1)$ by μ_1 and the distribution of $K''^n(\tau^* + 1)$ by μ_2. Note that $\sum_{x \in \mathbb{X}} |\mu_1(x) - \pi(x)| < \frac{\epsilon}{2}$ and $\sum_{x \in \mathbb{X}} |\mu_2(x) - \pi(x)| < \frac{\epsilon}{2}$. By the triangle inequality, it follows that $\sum_{x \in \mathbb{X}} |\mu_1(x) - \mu_2(x)| < \epsilon$. Then,

$$(42) \qquad \left| E\left[\frac{1}{n} d^n(K'^n, f^n(e^n(K'^n)))\right] - E\left[\frac{1}{n} d^n(K''^n, f^n(e^n(K''^n)))\right] \right|$$

$$= \left| \sum_{x_1, x_2, \ldots, x_n \in \mathbb{X}} \mu_1(x_1) \prod_{i=2}^{n} p_{x_{i-1}x_i} d^n(x^n, f^n(e^n(x^n))) - \right.$$

$$\left. \sum_{x_1, x_2, \ldots, x_n \in \mathbb{X}} \mu_2(x_1) \prod_{i=2}^{n} p_{x_{i-1}x_i} d^n(x^n, f^n(e^n(x^n))) \right|$$

$$= \left| \sum_{x_1, x_2, \ldots, x_n \in \mathbb{X}} (\mu_1(x_1) - \mu_2(x_1)) \prod_{i=2}^{n} p_{x_{i-1}x_i} d^n(x^n, f^n(e^n(x^n))) \right|$$

$$\leq \left| \sum_{x_1, x_2, \ldots, x_n \in \mathbb{X}} |\mu_1(x_1) - \mu_2(x_1)| \prod_{i=2}^{n} p_{x_{i-1}x_i} d^n(x^n, f^n(e^n(x^n))) \right|$$

$$\leq D_{\max} \left| \sum_{x_1, x_2, \ldots, x_n \in \mathbb{X}} |\mu_1(x_1) - \mu_2(x_1)| \prod_{i=2}^{n} p_{x_{i-1}x_i} \right|$$

$$= D_{\max} \left| \sum_{x_1 \in \mathbb{X}} |\mu_1(x_1) - \mu_2(x_1)| \prod_{x_2, \ldots, x_n \in \mathbb{X}} p_{x_{i-1}x_i} \right|$$

$$= D_{\max} \left| \sum_{x_1 \in \mathbb{X}} |\mu_1(x_1) - \mu_2(x_1)| \times 1 \right|, \text{ since } x_1 \text{ is fixed in the product}$$

$$\leq D_{\max} \epsilon$$

For $\delta > 0$ (think of δ small), $D > 0$, let $< e'^n, f'^n >_1^\infty$ be a source-code with rate $\leq R^E_{[\pi', P]}(D) + \delta$ to code the $X_{[\pi', P]}$ source with distortion D. Construct a source code $< e''^n, f''^n >_1^\infty$ to code the $X_{[\pi'', P]}$ source as follows. When the block-length is $n + \tau^*_\epsilon$, code $X''_1, X''_2, \ldots, X''_{\tau^*_\epsilon}$ arbitrarily, and code $(X''_{\tau^*_\epsilon+1}, X''_{\tau^*_\epsilon+2}, \ldots, X''_{\tau^*_\epsilon+n})$ using (e'^n, f'^n). It follows, by calculation of the distortion achieved for $(X''_1, X''_2, \ldots, X''_{\tau^*+n})$ by use of this code, in the process, using (42), and by noting that $\frac{\tau^*}{\tau^*+n}$ is a decreasing function of n, that

$$(43) \qquad R^E_{X_{[\pi'', P]}} \left(D + \frac{\tau^*_\epsilon}{\tau^*_\epsilon + n} D_{\max} + \epsilon D_{\max} \right) \leq R^E_{X_{[\pi', P]}}(D) + \delta$$

ϵ can be made arbitrarily small, τ^*_ϵ will depend on ϵ and n can be made arbitrarily large. It follows that for every $\alpha > 0$, every $\delta > 0$, $R^E_{X_{[\pi'', P]}}(D + \alpha) \leq R^E_{X_{[\pi', P]}}(D) + \delta$. By the continuity of $R^E_{X_{[\pi', P]}}(D)$ in D, it follows that $\forall \delta > 0$ $R^E_{X_{[\pi'', P]}}(D) \leq R^E_{X_{[\pi', P]}}(D) + \delta$. It follows, then, since $\delta > 0$ can be arbitrarily small, that $R^E_{X_{[\pi'',P]}}(D) \leq R^E_{X_{[\pi',P]}}(D)$. By interchanging π' and π'', it follows that, $R^E_{X_{[\pi',P]}}(D) \leq R^E_{X_{[\pi'',P]}}(D)$ Thus, $R^E_{X_{[\pi',P]}}(D) = R^E_{X_{[\pi'',P]}}(D)$. □

Appendix B. Proofs of (25) and (26)

To prove (25):

Proof. Let $< e^n, f^n >_1^\infty$ be the source code for the i.i.d. vector $J_\tau(X^T_{[\pi,P]})$ source.

Note that $e^n : (\mathbb{X}^{T-\tau})^n \to \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\}$ and $f^n : \{1, 2, \ldots, 2^{\lfloor nR \rfloor}\} \to (\mathbb{Y}^{T-\tau})^n$.

Let $s^n \in \mathbb{S}^n$ be a realization of S^n, the n-blocklength vector i.i.d. $X^T_{[\pi,P]}$ source which needs to be coded.

$s^n = (s^n(1), s^n(2), \ldots s^n(n))$ where each $s^n(i) \in \mathbb{S}$: $s^n(i) = (s^n(i)(1), s^n(i)(2), \ldots, s^n(i)(T))$.

Recall the projection operator, $J_\tau(s^n(i)) = (s^n(i)(\tau + 1), \ldots, s^n(i)(T))$.

Denote $J^n_\tau(s^n) = (J_\tau(s^n(1)), J_\tau(s^n(2)), \ldots, J_\tau(s^n(n)))$.

Then, $J^n_\tau(s^n)$ is an element of $(\mathbb{X}^{T-\tau})^n$. Denote $f^n(e^n(J^n_\tau(s^n))) = t'^n$.

Note that $t'^n = (t'^n(1), t'^n(2), \ldots, t'^n(n))$ where $t'^n(i) = (t'^n(i)(1), t'^n(i)(2), \ldots, t'^n(i)(T - \tau))$.

Fix a random $y \in \mathbb{Y}$. Define the extension transformation, $E_\tau(t'^n(i)) = (y, y, \ldots, y, t'^n(i)(1), t'^n(i)(2), \ldots, t'^n(i)(T-\tau))$, where the initial y's occur τ times.

Denote $E^n_\tau(t'^n) = (E_\tau(t'^n(1)), E_\tau(t'^n(2)), \ldots, E_\tau(t'^n(n)))$.

Note that $< e^n \circ J^n_\tau, E^n_\tau \circ f^n >_1^\infty$ is a rate R source code to code the i.i.d. vector $X^T_{[\pi,P]}$ source and that, $d'^n(s^n, E^n_\tau(f^n(e^n(J^n_\tau(s^n))))) \leq d''^n(s'^n, t'^n) + n\tau D_{\max}$. (25) follows. □

To prove (26):

Proof. Let (A, B) be a random vector on $\mathbb{A} \times \mathbb{B}$. Let $(A_1, B_1), (A_2, B_2), \ldots$ be a sequence where (A_i, B_i) are independent of each other and $(A_i, B_i) \sim (A, B)$. This sequence is the vector i.i.d (A, B) source. $\mathbb{A} \times \mathbb{B}$ is the source space. Let the source reproduction space be $\mathbb{A}' \times \mathbb{B}'$. $d_1 : \mathbb{A} \times \mathbb{A}' \to [0, \infty)$ is a distortion measure. $d_2 : \mathbb{B}' \times \mathbb{B}' \to [0, \infty)$ is a distortion measure. Assume that $\mathbb{A}, \mathbb{A}', \mathbb{B}, \mathbb{B}'$ are finite sets. Define: $d_0((a, b), (a', b')) \triangleq d_1(a, a') + d_2(b, b')$. d^n_1 and d^n_2, d^n_0 are respectively defined additively from d_1, d_2 and d_0. We can then define the rate-distortion functions for the i.i.d. A source and the i.i.d. (A, B) source, denoted, respectively, by $R^E_A(\cdot)$ and $R^E_{(A,B)}(\cdot)$. Then,

$$R^E_A(D) \leq R^E_{(A,B)}(D) \tag{44}$$

(44) is proved as follows: Given a code to code the i.i.d. (A, B) source, think of B_is as a source of common randomness, and use the obvious variant of the same code for coding the i.i.d. A source. Since the same code is used, (44) follows. Existence of a random code with a certain distortion implies the existence of a deterministic code with the same or lesser distortion. From this, (44) follows for deterministic codes. □

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, National Convention Record*, vol. 7, no. 4, pp. 142–163, March 1959.

[2] R. G. Gallager, *Information theory and reliable communication*. Wiley, January 1968.

[3] R. M. Gray, *Entropy and information theory*. Springer-Verlag, February 2011.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423 (Part 1) and pp. 623–656 (Part 2), July (Part 1) and October (Part 2) 1948. MR0026286

[5] A. N. Kolmogorov, "Theory of transmission of information," *Acad. R. P. Romine An. Romino-Soviet*, vol. 28, no. 1, pp. 5–33, 1959, translated in American Mathematical Society Translations, Series 2, Volume 33, 1963. MR0107572

[6] T. Berger, *Rate-distortion theory: mathematical basis for data compression*, ser. Prentice-Hall series in information and system sciences. Prentice Hall, October 1971. MR0408988

[7] B. V. Gnedenko, *The theory of probability*. Chelsea Publishing Company, New York, N. Y., 1962. MR0217824

[8] A. N. Shiryaev, *Probability*, 2nd ed. Springer, 1984. MR0737192

[9] W. Feller, *An introduction to probability theory and its applications, Volume 1*, 3rd ed. Wiley, 1968. MR0228020

[10] J. S. Rosenthal, "Convergence rates of Markov chains," *SIAM Review*, vol. 37, pp. 387–405, 1995. MR1355507

[11] V. A. Zorich, *Mathematical analysis, I and II*. Springer, March 2016. MR3445604

[12] H. Palaiyanur and A. Sahai, "On the uniform continuity of the rate-distortion function," in *2008 IEEE International Symposium on Information Theory*.

[13] Y. V. Prohorov and Y. A. Rozanov, *Probability theory: basic concepts, limit theorems, random processes*, 1st ed., ser. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berucksichtigung der Anwendungsgebiete, Band 157. Springer-Verlag, 1969.

[14] R. C. Bradley, "Basic properties of strong mixing conditions. a survey and some open questions," *Probability surveys*, vol. 2, pp. 107–144, 2005. MR2178042

[15] M. Agarwal, S. Mitter, and A. Sahai, "Layered black-box, behavioral interconnection perspective and applications to problems in communications, Part II: sources satisfying ψ-mixing criterion," *Communications in Information and Systems*, vol. 17, no. 4, 2017. MR3780974

Mukul Agarwal
Cambridge, MA
USA
E-mail address: magar@alum.mit.edu

Received December 7, 2012