Mutual information and capacity of a linear digital channel

Broomhead, D. S. and Sidorov, Nikita

2004

MIMS EPrint: **2006.20**

MUTUAL INFORMATION AND CAPACITY OF A LINEAR DIGITAL CHANNEL

D. S. BROOMHEAD AND NIKITA SIDOROV

ABSTRACT. In this paper we analyse a simple model of a digital communications channel. This model proves to be closely related to an iterated function system (IFS) related to the well-known Bernoulli convolution. We derive it from a randomly forced first-order ordinary differential equation. This allows the parameter of the Bernoulli convolution—the contraction rate, λ—to be related to the rate at which symbols are input to the channel. It is shown that for a channel with equiprobable binary inputs the mutual information between input and output distributions is the stationary measure of the complement of the overlap region of the IFS. We show that the mutual information is Hölder continuous with respect to λ and decreases hyper-exponentially as $\lambda \to 1$. We also study the case of non-equiprobable binary inputs and show that the maximum of the mutual information—the channel capacity—does not always correspond to equiprobable inputs.

INTRODUCTION

It is usual to model communications channels as linear filters [13]. Recently, a different approach to modelling *digital* communications channels has been proposed [2, 3, 4, 5] which exploits the discreteness of the symbol space. Here one lists the responses of the channel, in one sampling interval, to each of the possible input symbols. This approach combines the modelling of source and channel, which makes it possible to establish useful results about the properties of the channel without the need to assume a linear response.

These models are known as *iterated function systems* (IFS) (see, e.g., [1, 6, 11, 14] for a definition). In a nutshell, the dynamics of an IFS is given by random composition of a given collection of maps $\{f_i : \mathbf{X} \to \mathbf{X} \mid i \in \mathcal{S}\}$, where the selection is made according to some fixed probability distribution on a finite set $\mathcal{S}$ which we regard as the set of *inputs*. This naturally leads to a unique invariant stationary measure on $\mathbf{X}$ which in the present paper will be the Bernoulli convolution. The specific model we shall consider involves: $\mathcal{S} = \{\mathbf{0}, \mathbf{1}\}$ with probabilities p and $1 - p$, and two affine maps of the interval $\mathbf{X} = [0, 1]$

$$f_{\mathbf{0}}(x) = \lambda x,$$
$$f_{\mathbf{1}}(x) = \lambda x + 1 - \lambda.$$

where $\lambda \in (0, 1)$ is a parameter. (Physically, λ corresponds to the rate of transmission of data through the channel—see Section 1.) The *Bernoulli convolution*, $\mu_\lambda^{(p)}$, is the measure corresponding to the distribution of the random variable $(\lambda^{-1} - 1) \sum_{n=1}^{\infty} \varepsilon_n \lambda^n$

where the ε_n are i.i.d. random variables assuming the values 0 and 1 with probabilities p and $1-p$.

In this paper we use information theory to study the properties of this model. The main quantity of interest is the *mutual information* $R_\lambda^{(p)}$ [23, 18] between the distribution of symbols input to the channel and the distribution of outputs (which in the present paper will be identified with the states of the IFS). Roughly, the mutual information quantifies the degree to which knowledge of the channel output allows a correct inference to be made about the corresponding channel input. The *channel capacity* is then the maximum of this quantity over all possible input distributions.

In most of this paper we shall assume that the input symbols are equiprobable; in this case the mutual information is shown to be $R_\lambda^{(1/2)} =: R_\lambda = 2\mu_\lambda^{(1/2)}([0, 1-\lambda])$-bits if $\lambda \geq 1/2$ (Proposition 2.3). (Note that when $\lambda < 1/2$ there is a unique decoding which relates the output to the input symbol. In this case the mutual information is just $\log 2$ or 1-bit.) Our main results are as follows:

- The function $\lambda \mapsto R_\lambda$ is strictly decreasing on $(1/2, 1)$ (Lemma 3.1).
- The function $\lambda \mapsto R_\lambda$ is Hölder continuous at any $\lambda \in (1/2, 1)$ (Theorem 3.3).
- We study the asymptotic behaviour of $\lambda \mapsto R_\lambda$ as $\lambda \to 1/2$ and as $\lambda \to 1$ and prove that whereas in the former case it tends to unity at a linear rate (Proposition 3.5), in the latter it tends to zero at a rate which is faster than any exponential (Theorem 3.9). The importance of these results is that they show what happens physically as we try to increase the rate of transmission of information through the channel. Near to $\lambda = 1/2$ (where the output of the channel first becomes ambiguous) there is a linear rate in degradation. However, as the we approach $\lambda = 1$ (where the transmission rate is infinite) the degradation of information becomes catastrophic.

Section 4 is devoted to the case of general p. We show that:

- For any fixed $p \in (0, 1/2)$ there exists $\lambda \in ((\sqrt{5}-1)/2, 1)$ such that $R_\lambda^{(p)} > R_\lambda$ (Proposition 4.1). In other words at sufficiently large λ (surprisingly) the equiprobable input distribution does not maximise the mutual information.
- The function $p \mapsto R_\lambda^{(p)}$ is real analytic for $p \in (0, 1)$ and all $\lambda \in (1/2, 1)$ (Proposition 4.3).

1. THE BASIC MODEL

1.1. **Derivation.** The goal of this section is to explain the relevance of Bernoulli convolutions to the digital channel problem by deriving the IFS from a continuous time model of the channel. We use the simplest recursive linear channel model:

$$(1.1) \qquad \frac{dx}{dt} = -\gamma x + s(t),$$

where $x = x(t)$ represents both the state of the channel and its output at some time t, $\gamma > 0$ characterises the damping of the channel, and $s(t)$ represents the input.

Assume that the input to the channel consists of symbols taken from a binary alphabet, $\{\mathbf{0}, \mathbf{1}\}$, and write $\Sigma = \prod_1^\infty \{\mathbf{0}, \mathbf{1}\}$. We take this as a model of the signal source. It will be assumed that each symbol choice is made independently of all the others. This is incorporated by working with the product measure $\nu^{(p)} = \prod_1^\infty \{p, 1-p\}$ on Σ. (Here p is the probability of choosing $\mathbf{0}$.) Often we shall assume that a coding has been employed which maximises the entropy of the source; in which case $p = 1/2$.

We represent physical pulses corresponding to the input symbols by two compactly supported functions $s_\mathbf{0} : \mathbb{R} \to \mathbb{R}$ and $s_\mathbf{1} : \mathbb{R} \to \mathbb{R}$:

$$(1.2) \qquad s_{\mathbf{0},\mathbf{1}}(t) = \begin{cases} \varsigma_{\mathbf{0},\mathbf{1}}(t), & t \in [0, \tau) \\ 0, & \text{otherwise,} \end{cases}$$

where τ is the clock period (the inverse of the symbol input rate). The quantities $\varsigma_\mathbf{0}$ and $\varsigma_\mathbf{1}$ are bounded functions so that the integrals

$$\int_0^\tau e^{\gamma t} \varsigma_{\mathbf{0},\mathbf{1}}(t) \, dt$$

are finite. For any $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n, \ldots) \in \Sigma$, the corresponding input function $s(t)$ can then be written as

$$(1.3) \qquad s(t) = \sum_{n=1}^\infty s_{\varepsilon_n}(t - n\tau).$$

The IFS channel model corresponding to equation (1.1), given an input signal in the form of equation (1.3), is found by integrating over a clock period; for $x_n = x(n\tau)$ and given the next input symbol, the state of the channel one clock period later is:

$$(1.4) \qquad x_{n+1} = \lambda x_n + b_{\varepsilon_{n+1}}, \quad \text{where} \quad \varepsilon_{n+1} \in \{\mathbf{0}, \mathbf{1}\}.$$

Here $\lambda = e^{-\gamma\tau}$ and

$$b_{\varepsilon_{n+1}} = e^{-\gamma\tau} \int_0^\tau e^{\gamma t} \varsigma_{\varepsilon_{n+1}}(t) \, dt.$$

Note this provides an interpretation of λ in terms of τ.

We can reduce equation (1.4) to a canonical form. Assume that $b_\mathbf{0} \neq b_\mathbf{1}$; then, by shifting the origin, $x \mapsto x + b_\mathbf{0}/(1 - \lambda)$, and rescaling $x \mapsto x(b_\mathbf{1} - b_\mathbf{0})/(1 - \lambda)$, we indeed obtain the following family of maps on the interval $[0, 1]$

$$(1.5) \qquad \begin{aligned} f_\mathbf{0}(x) &= \lambda x, \\ f_\mathbf{1}(x) &= \lambda x + 1 - \lambda. \end{aligned}$$

Recall that the Bernoulli convolution $\mu_\lambda^{(p)}$ is defined as the invariant measure of the IFS given by (1.5)—see, e.g., [21]. A more direct definition is given below.

1.2. General properties.

In this section we shall summarise some salient properties of the Bernoulli convolution (see [21, 27] for more details). It is relevant to our application because its existence as a unique stationary measure is a consequence of $\gamma > 0$ (equivalently $0 \leq \lambda < 1$) which constitutes a stable channel. The measure, $\mu_\lambda^{(p)}$, is supported on a compact set $\mathcal{A}_\lambda \subseteq [0, 1]$ (see, e.g., [6]). The set $\mathcal{A}_\lambda$ satisfies the usual fixed point condition

$$(1.6) \qquad \mathcal{A}_\lambda = f_\mathbf{0}(\mathcal{A}_\lambda) \cup f_\mathbf{1}(\mathcal{A}_\lambda),$$

and the stationarity of the measure implies

$$(1.7) \qquad \mu_\lambda^{(p)}(E) = p\,\mu_\lambda^{(p)} \circ f_\mathbf{0}^{-1}(E) + (1 - p)\mu_\lambda^{(p)} \circ f_\mathbf{1}^{-1}(E)$$

for any measurable subset, E, of $\mathcal{A}_\lambda$. We use the superscript to indicate the dependence of the measure on input distribution. When $p = 1/2$, we shall drop the superscript.

A more explicit formula for the measure can be given as follows. Let π_λ be the map from Σ onto $\mathcal{A}_\lambda$ defined as the following projection[1]:

$$(1.8) \qquad \pi_\lambda(\varepsilon_1, \varepsilon_2, \ldots) := (\lambda^{-1} - 1) \sum_{n=1}^{\infty} \varepsilon_n \lambda^n.$$

This map is constructed by *backward iteration* [6]; since $f_\mathbf{0}$ and $f_\mathbf{1}$ are both contractions, the sequence $f_{\varepsilon_1}(x), f_{\varepsilon_1} \circ f_{\varepsilon_2}(x), f_{\varepsilon_1} \circ f_{\varepsilon_2} \circ f_{\varepsilon_3}(x) \ldots$ converges to a point in $\mathcal{A}_\lambda$ which depends only on the sequence $\varepsilon_1, \varepsilon_2, \ldots$ Then, in terms of the product measure on Σ,

$$(1.9) \qquad \mu_\lambda^{(p)}(E) = \nu^{(p)}(\pi_\lambda^{-1} E)$$

for any measurable subset $E \subseteq \mathcal{A}_\lambda$. The map π_λ can be thought of as a way of addressing points in $\mathcal{A}_\lambda$. The fact that it is a surjection means that every point in $\mathcal{A}_\lambda$ corresponds to at least one infinitely long sequence of binary inputs to the channel.

There are two cases of interest:

1.2.1. $0 < \lambda < 1/2$. In this case it is easy to see that $\mathcal{A}_\lambda$ is homeomorphic to the middle thirds Cantor set (and, indeed, is the middle thirds Cantor set when $\lambda = 1/3$) and π_λ is injective. From the point of view of digital channels this is significant since every point in $\mathcal{A}_\lambda$ can be associated uniquely with an input sequence. Thus the channel output at any time is an unambiguous representation of all the symbols that have been input. In practise, this property is of limited utility because the increase in measurement precision required to resolve the symbol transmitted n steps earlier scales exponentially with n. However, since the sets $f_\mathbf{0}(\mathcal{A}_\lambda)$ and $f_\mathbf{1}(\mathcal{A}_\lambda)$ are disjoint, a much easier task is to establish the value of the last symbol transmitted by determining which of these sets contains the current channel state.

1.2.2. $1/2 \le \lambda < 1$. In this case we have a Bernoulli convolution with overlap; π_λ is not an injection[2]. In this parameter range $\mathcal{A}_\lambda = [0,1]$ and the critical set $f_\mathbf{0}(\mathcal{A}_\lambda) \cap f_\mathbf{1}(\mathcal{A}_\lambda)$ is the closed interval $[1-\lambda, \lambda] =: E_\mathsf{A}$. From the point of view of signal processing, E_A is the set of states of the channel which do not provide information about the last symbol transmitted.

Much of the interest in Bernoulli convolutions has been focused on the "fine structure" of the stationary measure μ_λ for $\lambda \in (1/2, 1)$. The *law of pure types* asserts that for any given λ it is either absolutely continuous or purely singular [15]. This interest dates back to original seminal work of P. Erdős [7] who showed that there exist values of λ in this range for which μ_λ is (surprisingly) singular (see Section 5 for more detail). More recently, B. Solomyak [26] showed that these values are exceptional in the sense that μ_λ is absolutely continuous for almost all values of $\lambda \in (1/2, 1)$ (see also the survey paper [21]).

---

[1]Note the abuse of notation in the summation on the right hand side of equation (1.8) where the symbols $\varepsilon_n \in \{\mathbf{0}, \mathbf{1}\}$ are interpreted as having numerical values.

[2]Actually, the second author has recently shown [24, 25] that for any $\lambda > 1/2$, the preimage $\pi_\lambda^{-1}\{x\}$ has the cardinality of the continuum for a.e. $x \in (0,1)$. In the opposite direction P. Glendinning and the second author [12] have shown in particular that if $\lambda < (\sqrt{5} - 1)/2 \approx 0.618\ldots$, then there always exist x such that $\pi_\lambda^{-1}\{x\}$ is a single sequence.

2. SOME INFORMATION THEORY

2.1. **Input and output.** The source for the channel is assumed to be modelled by Σ equipped with the product measure $\nu^{(p)}$ and the Bernoulli shift $\sigma : \Sigma \to \Sigma$. Assuming that the input to the channel is generated by an orbit of points in Σ: $\boldsymbol{\varepsilon}, \sigma\boldsymbol{\varepsilon}, \sigma^2\boldsymbol{\varepsilon}, \ldots$, the input at time n is taken to be the first component of the nth point on the orbit, $(\sigma^{n-1}\boldsymbol{\varepsilon})_1 = \varepsilon_n$. The model given by equation (1.1) with the input function in the form of equation (1.3) is of this form.

Each output of the channel is assumed to be classified using no memory of any previous output[3]. To do this, we form a disjoint partition of the possible output values consisting of: the ambiguous set, $E_{\mathsf{A}} = f_{\mathbf{0}}(\mathcal{A}_\lambda) \cap f_{\mathbf{1}}(\mathcal{A}_\lambda)$; and the sets $E_{\mathbf{0}} = f_{\mathbf{0}}(\mathcal{A}_\lambda) \backslash E_{\mathsf{A}}$ and $E_{\mathbf{1}} = f_{\mathbf{1}}(\mathcal{A}_\lambda) \backslash E_{\mathsf{A}}$ of output values which can be attributed unambiguously to specific input symbols. Specifically, for the Bernoulli convolution, if $1/2 \leq \lambda < 1$, $E_{\mathsf{A}} = [1 - \lambda, \lambda]$, $E_{\mathbf{0}} = [0, 1 - \lambda)$ and $E_{\mathbf{1}} = (\lambda, 1]$. The goal of this section is to derive the explicit form of the mutual information for our model.

2.2. **Mutual Information.** The mutual information between the input distribution and output distribution can be defined as the difference between the entropy of the source and the entropy of the source conditioned on the output:

$$(2.1) \qquad R_\lambda^{(p)} = H_S^{(p)} - H^{(p)}(\text{in}|\text{out}).$$

The two extremes in the relation between input and output are when:

a: The output is independent of the input. In this case the mutual information is zero because $H^{(p)}(\text{in}|\text{out}) = H_S^{(p)}$.

b: The input determines the output exactly. In this case $H^{(p)}(\text{in}|\text{out}) = 0$ and the mutual information is precisely the information contained in the input.

Of the two terms in equation (2.1), the entropy of the binary source is well known:

$$H_S^{(p)} = -(p \log p + (1 - p) \log(1 - p)).$$

The remaining term, the entropy of the source conditioned on the output, is obtained from the following result (whose proof we leave to the reader as a simple exercise):

Proposition 2.1. *The entropy—measured in bits per input symbol—of the source conditioned on the output is*

$$H^{(p)}(\text{in}|\text{out}) = \mu_\lambda^{(p)}(E_{\mathsf{A}}) H^{(p)}(\text{in}|E_{\mathsf{A}}),$$

where the quantity

$$H^{(p)}(\text{in}|E_{\mathsf{A}}) = -\sum_{\varepsilon=\mathbf{0},\mathbf{1}} \mu_\lambda^{(p)}(\varepsilon|E_{\mathsf{A}}) \log \mu_\lambda^{(p)}(\varepsilon|E_{\mathsf{A}})$$

is the entropy of the source conditioned on the output being ambiguous.

The conditional probabilities can be expressed as:

$$\mu_\lambda^{(p)}(\mathbf{0}|E_{\mathsf{A}}) = \frac{p - \mu_\lambda^{(p)}(E_{\mathbf{0}})}{\mu_\lambda^{(p)}(E_{\mathsf{A}})}, \quad \mu_\lambda^{(p)}(\mathbf{1}|E_{\mathsf{A}}) = \frac{1 - p - \mu_\lambda^{(p)}(E_{\mathbf{1}})}{\mu_\lambda^{(p)}(E_{\mathsf{A}})}.$$

There is an explicit formula for the measure of $E_{\mathbf{0}}$

---

[3]In contrast, it is always possible to use consecutive pairs of outputs to make an unambiguous characterisation the latest input symbol. This follows because the channel is an all-pole filter; that is, it can be inverted, given x_n and x_{n-1}, by forming the combination $x_n - \lambda x_{n-1}$ (see equations (1.5)).

Lemma 2.2. *We have*

(2.2) $$\mu_\lambda^{(p)}(E_\mathbf{0}) = p\nu^{(p)}(\mathfrak{A}_\lambda),$$

where

(2.3) $$\mathfrak{A}_\lambda = \left\{ (\varepsilon_1, \varepsilon_2, \ldots) \in \Sigma : \sum_{n=1}^{\infty} \varepsilon_n \lambda^n \leq 1 \right\}.$$

Proof. By equation (1.9)

$$
\begin{aligned}
\pi_\lambda^{-1}(E_\mathbf{0}) &= \left\{ (\varepsilon_1, \varepsilon_2, \ldots) \in \Sigma : (\lambda^{-1} - 1) \sum_{n=1}^{\infty} \varepsilon_n \lambda^n \leq 1 - \lambda \right\} \\
&= \left\{ (\varepsilon_1, \varepsilon_2, \ldots) \in \Sigma : \sum_{n=1}^{\infty} \varepsilon_n \lambda^{n-1} \leq 1 \right\} \\
&= \left\{ (\varepsilon_1, \varepsilon_2, \ldots) \in \Sigma : \varepsilon_1 + \sum_{n=1}^{\infty} \varepsilon_{n+1} \lambda^n \leq 1 \right\}.
\end{aligned}
$$

Apart from the singleton $(\mathbf{1}\dot{\mathbf{0}})$ which has zero measure, the elements of this set all have the form $(\mathbf{0}, \varepsilon_2, \ldots)$, and now the claim follows from the definition of $\nu^{(p)}$. $\square$

Thus, using the terminology of the paper [17], we shall be studying the *subexpansions* of unity in base $\beta = \lambda^{-1}$.

We can solve the channel capacity problem completely if $\mu_\lambda^{(p)}(E_\mathsf{A}) = 0$, that is, $\lambda \leq 1/2$. In this case the conditional entropy vanishes—see Proposition 2.1—and the mutual information is simply the entropy of the source. It follows that the maximum of the mutual information—the channel capacity—occurs when $p = 1/2$ and is one bit per symbol. When $\lambda > 1/2$, $\mu_\lambda^{(p)}(E_\mathsf{A}) \neq 0$ and the problem becomes more complicated. Our first goal is to study the equiprobable case $p = 1/2$. Here, in view of $\mu_\lambda(E_\mathbf{0}) = \mu_\lambda(E_\mathbf{1})$, it follows from Proposition 2.1 that

$$H(\mathrm{in}|E_\mathsf{A}) = \log 2.$$

The mutual information is therefore related directly to the stationary measure by the following

Proposition 2.3. *The mutual information—measured in bits per input symbol—of the channel specified above with uniform input distribution is*

$$R_\lambda^{(1/2)} =: R_\lambda = \mu_\lambda(E_\mathbf{0} \cup E_\mathbf{1}).$$

That is,

(2.4) $$R_\lambda = 2\mu_\lambda([0, 1 - \lambda]).$$

The first expression given in Proposition 2.3 has a clear, intuitive appeal. Each symbol transmitted either puts the channel into an ambiguous state (in which case the channel state provides no information about the input) or it does not. In the latter case—which occurs with probability $\mu_\lambda(E_\mathbf{0} \cup E_\mathbf{1})$—one bit of information is conveyed.

In Section 3 we investigate how this picture varies as a function of the rate at which symbols are transmitted. We shall state our results in terms of λ which is the natural parameter for the mathematical development. The interesting physical parameter
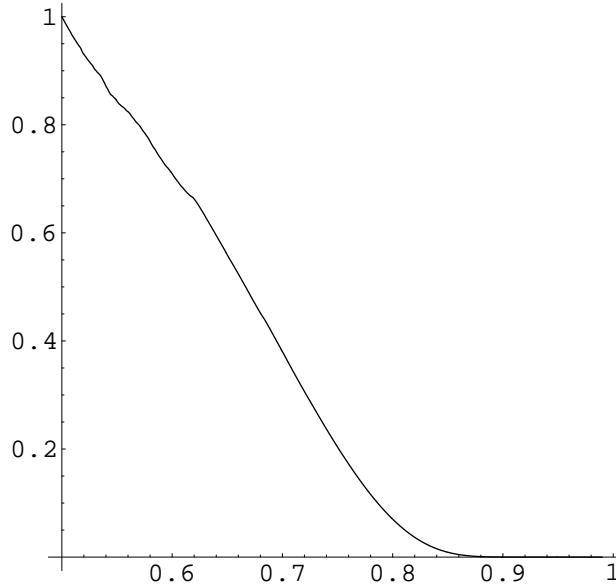
FIGURE 1. Plot of a numerical estimate of the mutual information, R_λ, (measured in bits per symbol) as a function of λ. Note that when $\lambda < 1/2$ (not shown), $R_\lambda = 1$.

could be argued to be the rate at which symbols are input to the channel, i.e., τ^{-1}. We note that $\tau^{-1} \propto |\log \lambda|^{-1}$, and in particular that λ tends to unity as τ^{-1} increases to $+\infty$. As well as R_λ, therefore, we shall also consider the quantity

$$(2.5) \qquad \widetilde{R}_\lambda := \frac{R_\lambda}{\tau} = \frac{\gamma R_\lambda}{|\log \lambda|}$$

which gives a scaled mutual information in units of information per unit time. For ease of reference, we shall refer to this quantity as the *information rate*.

3. MAIN RESULTS

In this section we study the behaviour of the mutual information and the information rate when $p = 1/2$. Our approach is based on reducing the problem to one of making estimates of the product measure, ν, on the symbol space Σ. By equations (2.4) and (2.2) we have

$$(3.1) \qquad R_\lambda = \nu(\mathfrak{A}_\lambda).$$

where $\mathfrak{A}_\lambda$ is the set defined in equation (2.3). As we remarked earlier, this is the set of the *subexpansions* of unity in base $\beta = \lambda^{-1}$.

3.1. **Monotonicity.** Figure 1 shows a numerical estimate of R_λ which suggests that R_λ is a decreasing function of the symbol input rate for $\lambda \in (1/2, 1)$ (see the Appendix to find out about numerical methods we used). The following assertion justifies this suggestion:

Lemma 3.1. *The function $\lambda \mapsto R_\lambda$ is strictly decreasing on $(1/2, 1)$.*

Proof. From the equations (2.3) and (3.1) it follows immediately that R_λ is nonincreasing. To prove the claim, it suffices to show that for any λ and any $\lambda' > \lambda$,

$$\nu(\mathfrak{A}_\lambda \setminus \mathfrak{A}_{\lambda'}) > 0.$$

We need some facts from the theory of β-*expansions.* Namely, let $(a_n)_1^\infty$ and $(a'_n)_1^\infty$ be the greedy expansions of 1 in bases $\beta = \lambda^{-1}$ and $\beta' = (\lambda')^{-1}$ respectively. More precisely, put $a_n = [\tau_\beta^{n-1}(1)]$, (where $\tau_\beta(x) = \beta x \bmod 1$ is the β-shift on the interval $[0,1)$) and similarly for λ'. It is known [20] that $(a_n)_1^\infty$ must be lexicographically greater than $(a'_n)_1^\infty$, i.e., there exists $n \geq 2$ such that $a_j \equiv a'_j$ for $1 \leq j \leq n-1$ and $a_n > a'_n$ (which means that $a_n = \mathbf{1}, a'_n = \mathbf{0}$).

Consider the cylinder $C = [\varepsilon_1 = a_1, \ldots, \varepsilon_n = a_n] \subset \Sigma$. By the definition of the greedy expansion (as the largest possible one),

$$\sum_{j=1}^n a_j (\lambda')^j > 1,$$

whence $C \cap \mathfrak{A}_{\lambda'} = \emptyset$; let us show that $\nu(C \cap \mathfrak{A}_\lambda) > 0$. Without loss of generality we may always assume that $(a_{n+1}, a_{n+2}, \ldots) \neq (\mathbf{0}, \mathbf{0}, \ldots)$ – it suffices to "adjust" the greedy expansion in such a way that there will be no zero tail [20]. Then $\kappa := 1 - \sum_1^n a_j \lambda^j > 0$. Hence $[\varepsilon_1 = a_1, \ldots, \varepsilon_n = a_n, \varepsilon_{n+1} = \cdots = \varepsilon_{n+L} = \mathbf{0}] \subset \mathfrak{A}_\lambda$ for L sufficiently large; it suffices to take

$$L = \left[\frac{\log \kappa (1 - \lambda)}{\log \lambda} \right].$$

$\square$

3.2. Hölder continuity.

From the formula (2.4) and the continuity of μ_λ it follows immediately that the function $\lambda \mapsto R_\lambda$ is also continuous (as a composition of two continuous functions). Returning to Figure 1, one might wonder whether R_λ is in fact Lipschitz. In particular, it appears that something happens for values of λ in the region $(1/2, g)$, where $g := (\sqrt{5} - 1)/2 \approx 0.618$, i.e., the smaller golden ratio[4]. The plots of $\widetilde{R}_\lambda$ shown in Figures 2 and 3 also suggest that something complicated going on in this region (see Section 5 for precise open questions). It turns out, however, that R_λ—and, therefore, $\widetilde{R}_\lambda$—is Hölder continuous (but not necessarily Lipschitz) on the whole interval $(1/2, 1)$. To prove this, we need some preliminaries.

Let

$$\Gamma(\lambda) := \left\{ (\varepsilon_n)_1^\infty \in \Sigma : \sum_{n=1}^\infty \varepsilon_n \lambda^n = 1 \right\}.$$

The set $\Gamma(\lambda)$ has been intensively studied in the 1990's by Hungarian mathematicians led by P. Erdős [8, 9, 10], see also [16, 17]. Note that $\Gamma(\lambda) \neq \emptyset$ for all λ (use the greedy algorithm described above). In some cases $\operatorname{card} \Gamma(\lambda) = 1$; actually, it was shown in [9] that for any $\ell \in \mathbb{N} \cup \aleph_0$ there exists λ such that $\operatorname{card} \Gamma(\lambda) = \ell$.

Let $\Gamma_n(\lambda)$ denote the set of $\mathbf{0}$–$\mathbf{1}$ words of length n that can be extended to sequences from $\Gamma(\lambda)$. As is easy to see, since $\pi_\lambda(\Sigma) = [0, 1]$,

$$(3.2) \qquad \Gamma_n(\lambda) = \left\{ (\varepsilon_1 \ldots \varepsilon_n) \in \Sigma_n : 1 - \frac{\lambda^{n+1}}{1 - \lambda} \leq \sum_{j=1}^n \varepsilon_j \lambda^j \leq 1 \right\}.$$

---

[4]It is worth mentioning [10, Theorem 3] that if $\lambda \in (g, 1)$, then $\pi_\lambda^{-1}\{x\}$ has the cardinality of the continuum for *every* $x \in (0, 1)$.

Finally, let

$$\gamma_n(\lambda) := \#\Gamma_n(\lambda)$$

and

$$\gamma(\lambda) = \limsup_{n\to\infty} \gamma_n(\lambda)^{1/n} \in [1,2].$$

The quantity $\gamma(\lambda)$ indicates "how many" λ-expansions of unity there are. Below we will show that $\gamma(\lambda) < 2$ for each $\lambda \in (1/2, 1)$.

Proposition 3.2. *For any* $1/2 < \lambda < \lambda' < 1$,

$$(3.3) \qquad R_\lambda - R_{\lambda'} = O((\lambda' - \lambda)^\alpha), \quad \alpha = \frac{\log 2 - \log\gamma(\lambda)}{2\log 1/\lambda}.$$

Proof. Obviously, the claim is about "close" λ and λ', so without loss of generality one may take the subsequence $\lambda' = \lambda'_n = \lambda + (1-\lambda)^2\lambda^{2n-1}$ starting with some n. Indeed, $\widetilde\lambda - \lambda \asymp \lambda^{2n}$ for any $\widetilde\lambda \in (\lambda'_n, \lambda'_{n-1})$ and it suffices to note that the estimate in (3.3) depends only on logarithms.

Assume $(\varepsilon_j)_1^\infty \in \mathfrak{A}_\lambda \setminus \mathfrak{A}_{\lambda'}$. Our goal is to show that $(\varepsilon_j)_1^n \in \Gamma_n(\lambda)$ unless $\varepsilon_{n+1} = \cdots = \varepsilon_{2n} = 1$. Since $(\varepsilon_j)_1^\infty \notin \mathfrak{A}_{\lambda'}$, we have $\sum_1^\infty \varepsilon_j(\lambda')^j > 1$. Hence

$$\sum_1^\infty \varepsilon_j(\lambda + (1-\lambda)^2\lambda^{2n-1})^j \sim \sum_1^\infty \varepsilon_j\lambda^j + (1-\lambda)^2\lambda^{2n-1}\sum_1^{n'} j\varepsilon_j\lambda^j \geq 1$$

(all other binomial terms are negligible). Thus,

$$\sum_1^n \varepsilon_j\lambda^j \geq 1 - (1-\lambda)^2\lambda^{2n-1}\sum_1^\infty j\varepsilon_j\lambda^j - \sum_{n+1}^\infty \varepsilon_j\lambda^j$$

$$\geq 1 - (1-\lambda)^2\lambda^{2n-1}\cdot\frac{\lambda}{(1-\lambda)^2} - \sum_{n+1}^\infty \varepsilon_j\lambda^j$$

$$= 1 - \lambda^{2n} - \sum_{n+1}^\infty \varepsilon_j\lambda^j.$$

Since $\varepsilon_{n+1}\ldots\varepsilon_{2n} \neq 1\ldots 1$,

$$\sum_{n+1}^\infty \varepsilon_j\lambda^j \leq \frac{\lambda^{n+1}}{1-\lambda} - \lambda^{2n},$$

whence

$$\sum_1^n \varepsilon_j\lambda^j \geq 1 - \frac{\lambda^{n+1}}{1-\lambda},$$

which by (3.2) implies $(\varepsilon_j)_1^n \in \Gamma_n(\lambda)$. Finally,

$$\nu(\mathfrak{A}_\lambda \setminus \mathfrak{A}_{\lambda'}) = \nu((\mathfrak{A}_\lambda \setminus \mathfrak{A}_{\lambda'}) \cap [\varepsilon_{n+1} = \cdots = \varepsilon_{2n} = 1])$$

$$+ \nu((\mathfrak{A}_\lambda \setminus \mathfrak{A}_{\lambda'}) \cap [\varepsilon_{n+1}\ldots\varepsilon_{2n} \neq 1\ldots 1])$$

$$\leq 2^{-n} + \frac{\gamma_n(\lambda)}{2^n} \leq 2\cdot\frac{\gamma_n(\lambda)}{2^n}.$$

Now the claim follows from the fact that $\lambda' - \lambda \asymp \lambda^{2n}$. $\square$

Theorem 3.3. *The function* $\lambda \mapsto R_\lambda$ *is Hölder continuous at any* $\lambda \in (1/2, 1)$ *with the exponent* $\frac{(2\lambda-1)\log 2}{2\log 1/\lambda}$.
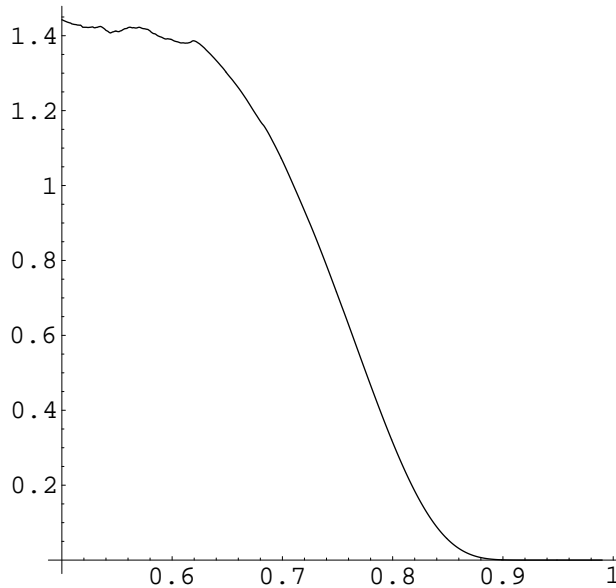
FIGURE 2. Plot of a numerical estimate of the information rate, $\widetilde{R}_\lambda$, (measured in bits per second and assuming $\gamma = 1$) as a function of λ. Note that when $\lambda < 1/2$ (not shown), $\widetilde{R}_\lambda = \log \lambda^{-1}$.

Proof. Let m denote the minimal j such that $\lambda + \lambda^2 + \cdots + \lambda^j > 1$. Then a sequence from $\Gamma(\lambda)$ cannot begin with m unities, whence

$$\gamma(\lambda) \leq 2^{\frac{2^{m-1}}{2^m}} < 2.$$

Finally, by (3.3),

$$\alpha \geq 2^{-m-1} \frac{\log 2}{\log 1/\lambda} > \frac{(2\lambda - 1)\log 2}{2\log 1/\lambda}.$$

$\square$

Corollary 3.4. *The function $\lambda \mapsto \widetilde{R}_\lambda$ is Hölder continuous at any $\lambda \in (1/2, 1)$.*

Proof. It suffices to apply (2.5). $\square$

3.3. **Asymptotic behaviour.** If the mutual information is a strictly decreasing function of the symbol input rate when $\lambda > 1/2$, what about the information rate? Is it possible to increase the information transmitted by increasing the symbol input rate? The numerical estimate of $\widetilde{R}_\lambda$ (see Figure 2 and Figure 3), suggests that this is not possible in any absolute sense, but that locally it may be possible because $\widetilde{R}_\lambda$ appears not to be monotone over the whole of $(1/2, 1)$. The numerical work does suggest a decrease in the immediate neighbourhood of $\lambda = 1/2$. Moreover, for large enough values of λ, both R_λ and $\widetilde{R}_\lambda$ appear to decrease catastrophically (which will be shown rigidly below). The non-monotone region appears to be associated with the interval of λ in which singular measures are known to exist[5].

---

[5]Recall that the maximal λ, for which μ_λ is known to be singular, is the root of $x^3 + x^2 = 1$, i.e., the reciprocal of the smallest Pisot number [7, 22] (see Section 5). The numerical value is $\lambda_{max} = 0.75488\ldots$
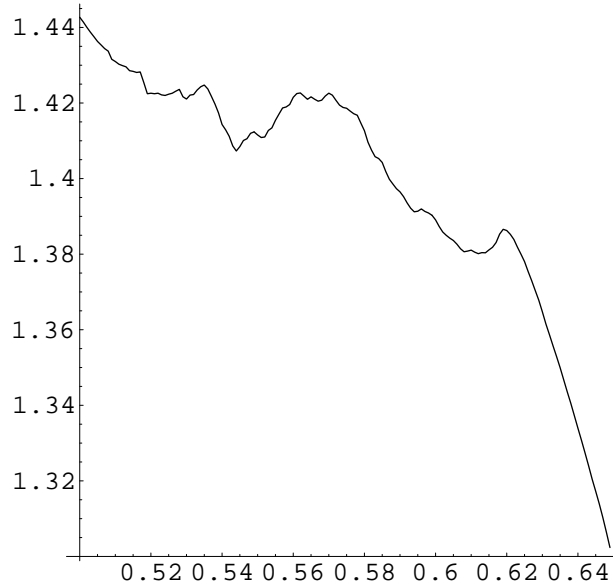
FIGURE 3. A detail of Figure 2, showing the irregular "plateau" region.

In the following, we shall look at asymptotic behaviour in the two limiting cases where numerically $\widetilde{R}_\lambda$ appears to be a decreasing function of symbol input rate. In the next section, we consider at the behaviour of R_λ and $\widetilde{R}_\lambda$ as $\lambda \to 1/2 + 0$, that is, near (and after) the transition to ambiguous transmission. In the subsequent sections, we investigate the behaviour of R_λ and $\widetilde{R}_\lambda$ as the symbol input rate tends to infinity—as $\lambda \to 1 - 0$. In both cases our goal is to estimate the product measure, ν, of the set $\mathfrak{A}_\lambda$. Our tool consists essentially of counting the number of $\mathbf{1}$'s that is consistent with the channel state being in the set $E_{\mathbf{0}}$. This can provide—through relation (2.3)—upper and lower bounds on the value of $\nu(\mathfrak{A}_\lambda)$.

3.3.1. *Asymptotic behaviour as $\lambda \to 1/2 + 0$.*

Proposition 3.5.
$$1 - R_\lambda \asymp 2\lambda - 1 \ \text{as} \ \lambda \to 1/2 + 0.$$

Proof. By (3.1), it suffices to estimate $1 - \nu(\mathfrak{A}_\lambda)$ as $\lambda \to 1/2 + 0$. Consider the minimal integer, m, and the maximal integer, r, which satisfy the following inequalities
$$\lambda + \lambda^2 + \cdots + \lambda^m > 1 \geq \lambda + \lambda^2 + \cdots + \lambda^r$$
or, equivalently
$$\lambda^{r+1} \geq 2\lambda - 1 > \lambda^{m+1}.$$
In other words,

(3.4) $\qquad m = m(\lambda) := \min \left\{ l \in \mathbb{N} : \lambda^{l+1} < 2\lambda - 1 \right\} = \left\lceil \dfrac{\log(2\lambda - 1)}{\log \lambda} \right\rceil + 1$

and
$$r = r(\lambda) := \max \left\{ l \in \mathbb{N} : \lambda^{l+1} \geq 2\lambda - 1 \right\} = m - 1.$$
Let $\mathcal{C}^{(q)}$ denote the set of all possible cylinders of length q in Σ, i.e.,
$$\mathcal{C}^{(q)} = \left\{ [\varepsilon_1 = i_1, \varepsilon_2 = i_2, \ldots, \varepsilon_q = i_q] : i_j \in \{\mathbf{0}, \mathbf{1}\} \right\}.$$

Obviously, $\mathcal{C}^{(q)}$ is a disjoint partition of Σ, whence

$$\nu(\mathfrak{A}_\lambda) = \sum_{C \in \mathcal{C}^{(m)}} \nu(\mathfrak{A}_\lambda \cap C).$$

The cylinder $[\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_m = \mathbf{1}]$ has an empty intersection with $\mathfrak{A}_\lambda$, because $\lambda + \lambda^2 + \cdots + \lambda^m > 1$. Since the measure of this cylinder is 2^{-m}, we have

$$\nu(\mathfrak{A}_\lambda) \le 1 - 2^{-m} \sim 1 - (2\lambda - 1) = 2(1 - \lambda),$$

whence $1 - \nu(\mathfrak{A}_\lambda) \ge \text{const} \cdot (2\lambda - 1)$.

For a lower bound, we observe that the cylinder $[\varepsilon_1 = \mathbf{0}]$ together with any cylinder of the form $[\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_{l-1} = \mathbf{1}, \varepsilon_l = \mathbf{0}]$ lies in $\mathfrak{A}_\lambda$ for any $l \le r$. Therefore,

$$\nu(\mathfrak{A}_\lambda) \ge \sum_{j=1}^{r} 2^{-j} = 1 - 2^{-r} \sim 1 - (2\lambda - 1) = 2(1 - \lambda),$$

whence $1 - \nu(\mathfrak{A}_\lambda) = O(2\lambda - 1)$, and we are done. $\square$

This result characterises the decay of the channel capacity, R_λ, as λ exceeds $1/2$. If we write $|\log \lambda| = \log 2 - O(2\lambda - 1)$, it is easy to show that $\widetilde{R}_\lambda$ also decreases. Locally, at least, $\lambda = 1/2$ gives the maximum rate at which information can be transmitted by the channel if we assume that the input distribution is uniform.

Corollary 3.6.
$$0 < \frac{1}{\log 2} - \widetilde{R}_\lambda \asymp 2\lambda - 1 \text{ as } \lambda \to 1/2 + 0.$$

3.3.2. *Asymptotic behaviour as $\lambda \to 1 - 0$.* We turn now to the other limit—corresponding to arbitrarily large symbol input rates—and ask what is the asymptotic behaviour of R_λ and $\widetilde{R}_\lambda$ as $\lambda \to 1 - 0$? To begin with we can use the method above to obtain a simple—but revealing—estimate. Put

$$k = k(\lambda) = \max\ \left\{l \in \mathbb{N} : \lambda^l > 1/2\right\} = \left[-\frac{\log 2}{\log \lambda}\right],$$

$$N = N(\lambda) = \max\ \{l \in \mathbb{N} : \lambda + \lambda^l > 1\} = \left[\frac{\log(1 - \lambda)}{\log \lambda}\right].$$

We have $\nu(\mathfrak{A}_\lambda) = \sum_{C \in \mathcal{C}^{(k)}} \nu(\mathfrak{A}_\lambda \cap C)$; observe that $\mathfrak{A}_\lambda \cap C$ is nonempty for some $C \in \mathcal{C}^{(k)}$ if and only if C contains not more than one unity; this follows from the definition of k, as $\lambda^j + \lambda^{j'} > 1$ for $j, j' \le k$. Since there are only $k + 1$ cylinders containing not more than one unity,

$$\nu(\mathfrak{A}_\lambda) \le \frac{k + 1}{2^k} = O\left(\frac{1}{1 - \lambda} \cdot 2^{-\frac{\log 2}{1 - \lambda}}\right).$$

and hence, the mutual information tends to zero as the symbol input rate tends to infinity. Indeed, this simple bound shows that the mutual information vanishes quickly enough for the following assertion to be true:

Corollary 3.7. $\widetilde{R}_\lambda \to 0$ *as* $\lambda \to 1 - 0$.

This estimate can be easily tightened up to get a good picture of just how rapidly the mutual information vanishes. Let $\ell \ge 1$ be an arbitrary natural number. Using the same argument and the fact that $\lambda^{\ell k} > 2^{-\ell}$, we deduce that the total number

of unities in the cylinders from $\mathcal{C}^{(\ell k)}$ whose intersection with $\mathfrak{A}_\lambda$ is nonempty, cannot exceed 2^ℓ. Therefore,

$$\nu(\mathfrak{A}_\lambda) \leq 2^{-\ell k} \sum_{j=0}^{2^\ell} \binom{\ell k}{j} = O(2^{-\ell k} k^{2^\ell}).$$

Since ℓ is arbitrary, we conclude that

$$\nu(\mathfrak{A}_\lambda) = O\left(a^k\right)$$

for any $a \in (0, 1)$, whence follows

Proposition 3.8.

$$R_\lambda = O\left(a^{\frac{1}{1-\lambda}}\right), \quad \lambda \to 1 - 0,$$
$$\widetilde{R}_\lambda = O\left(a^{\frac{1}{1-\lambda}}\right), \quad \lambda \to 1 - 0,$$

for any $a \in (0, 1)$.

We have thus shown that both R_λ and $\widetilde{R}_\lambda$ decay faster than any exponential function of $\frac{1}{1-\lambda}$. The next step is to establish a more precise description of the behaviour of their logs.

The lower bound for the mutual information is straightforward. Since

$$\lambda^{N+1} + \lambda^{N+2} + \cdots = \frac{\lambda^{N+1}}{1-\lambda} < 1$$

(by the definition of N), the cylinder $[\varepsilon_1 = \mathbf{0}, \varepsilon_2 = \mathbf{0}, \ldots, \varepsilon_N = \mathbf{0}] \subset \mathfrak{A}_\lambda$, whence

(3.5) $$\nu(\mathfrak{A}_\lambda) \geq 2^{-N} \geq \text{const} \cdot (1-\lambda)^{\frac{\log 2}{1-\lambda}},$$

and a similar bound is valid for $\widetilde{R}_\lambda$. Our goal is to show that in the "logarithmic sense" the upper bound is the same.

Theorem 3.9. *We have*

$$\lim_{\lambda \to 1-0} \frac{(1-\lambda) \log R_\lambda}{\log(1-\lambda)} = \log 2.$$

Moreover,

(3.6) $$\left| \frac{(1-\lambda) \log R_\lambda}{\log(1-\lambda)} - \log 2 \right| = O\left(\frac{\log |\log(1-\lambda)|}{|\log(1-\lambda)|} \right), \quad \lambda \to 1 - 0.$$

The same is true for $\log \widetilde{R}_\lambda$.

Proof. It suffices to prove the second statement for R_λ. We are going to follow the same line of ideas as in the previous items. Let us show first that

(3.7) $$\lambda^{N-k-1} + \lambda^{N-k} + \cdots + \lambda^N > 1.$$

Indeed, the inequality is equivalent to

$$\frac{\lambda^{N+1}}{1 - \lambda + \lambda^{N+1}} > \lambda^{k+2},$$

which in turn is equivalent to

$$k > N - 1 - \log_\lambda(1 - \lambda + \lambda^{N+1}).$$

By the definition of N, we have $\lambda^{N+1} < 1 - \lambda$, whence it would suffice to show that

$$k \geq N - 1 - \log_\lambda(2(1 - \lambda)).$$

Since $k + \log_\lambda 2 \in (-1, 0]$ and $N - \log_\lambda(1 - \lambda) \in (-1, 0]$, we are done.

We consider the cylinders of length N and, similarly to the above, utilise the fact that $\nu(\mathfrak{A}_\lambda) = \sum_{C \in \mathcal{C}^{(N)}} \nu(\mathfrak{A}_\lambda \cap C)$. By (3.7), if $\mathfrak{A}_\lambda \cap C \neq \emptyset$ for some $C \in \mathcal{C}^{(N)}$, then the number of unities in C cannot exceed $k + 2$. Therefore,

$$(3.8) \qquad \nu(\mathfrak{A}_\lambda) \leq 2^{-N} \sum_{j=0}^{k+2} \binom{N}{j}.$$

To simplify this estimate, we would like to show that for any $\lambda > 0.93$,

$$(3.9) \qquad \binom{N}{j} \leq 1/2 \binom{N}{j+1}, \quad j = 0, 1, \ldots, k+1.$$

Indeed, we have $\binom{N}{j} = \frac{j+1}{N-j}\binom{N}{j+1}$, and it suffices to show that $\frac{j+1}{N-j} < 1/2$ for $j \leq k+1$. In turn, this is equivalent to the inequality $N > 3k+5$. Now in view of the definitions of k and N, we have $N \geq \log_\lambda(1 - \lambda) - 1$ and $k < -\log_\lambda 2$. Now the last-mentioned inequality will follow from $\lambda^6 > 8(1 - \lambda)$, which actually holds for $\lambda > 0.93$.

Thus, from (3.8) and (3.9) it follows

$$(3.10) \qquad \nu(\mathfrak{A}_\lambda) \leq 2^{-N+1} \binom{N}{k+2}.$$

Using Stirling's Formula, we get

$$\log \binom{N}{k+2} \sim N \log N - N - (N - k - 2)\log(N - k - 2)$$
$$+ N - k - 2 - (k+2)\log(k+2) + k + 2,$$

whence

$$\log \binom{N}{k+2} \sim N \log \frac{N}{N-k} + k \log \frac{N}{k} \sim k \log \frac{N}{k}, \quad \lambda \to 1 - 0.$$

By (3.10) and (3.5), $-N \log 2 \leq \log \nu \mathfrak{A}_\lambda \leq (-N + 1)\log 2 + \log \binom{N}{k+2}$, whence

$$\left| \frac{\log R_\lambda}{N} + \log 2 \right| = O\left(\frac{k}{N} \log \frac{N}{k} \right).$$

To obtain (3.6), it suffices to use the definitions of N and k. $\qquad \square$

4. Channel Capacity

So far we have confined ourselves to considering the symmetric case $p = 1/2$. What happens when we allow p to vary? We noted in Section 2.2 that when $\mu_\lambda^{(p)}(E_A) = 0$ (i.e., when $\lambda \leq 1/2$), the vanishing of the conditional entropy, according to Proposition 2.1, implies that the mutual information is simply the entropy of the source. In this case we can say that the channel capacity—the maximum of the mutual information—is one bit per symbol and that the channel achieves this maximum when $p = 1/2$. If we consider the capacity measured in bits per unit time—found by scaling the channel capacity with $\tau = |\log \lambda|$—we find that in this region, this quantity increases with the rate at which symbols are input.
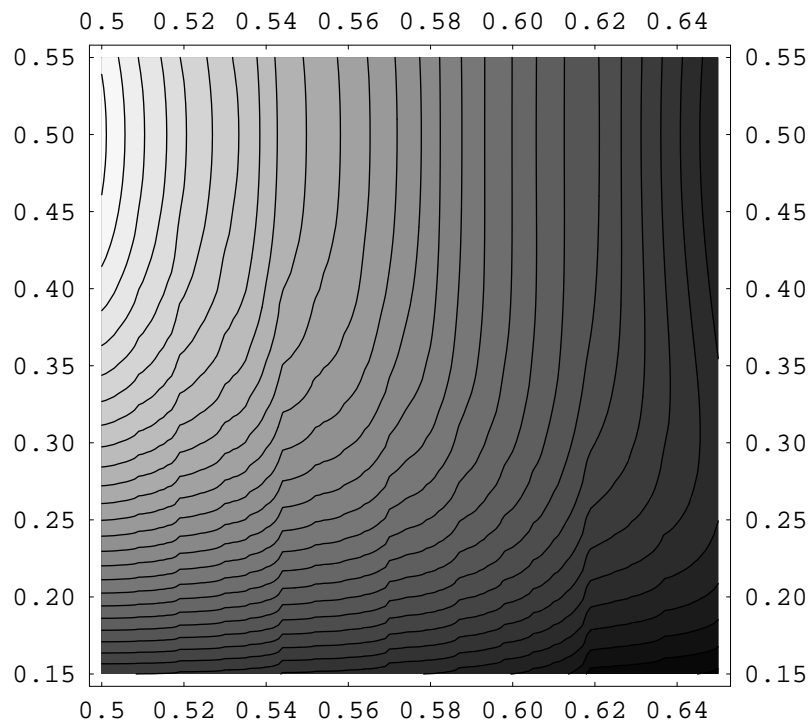
FIGURE 4. Plot of a numerical estimate the mutual information, $R^{(p)}_\lambda$, as a function of λ (abscissæ) and p (ordinates). The shading represents the value of the mutual information with dark representing small values and light representing large values. The range of the plot is from ~ 0.3 to ~ 0.7 bits per symbol.
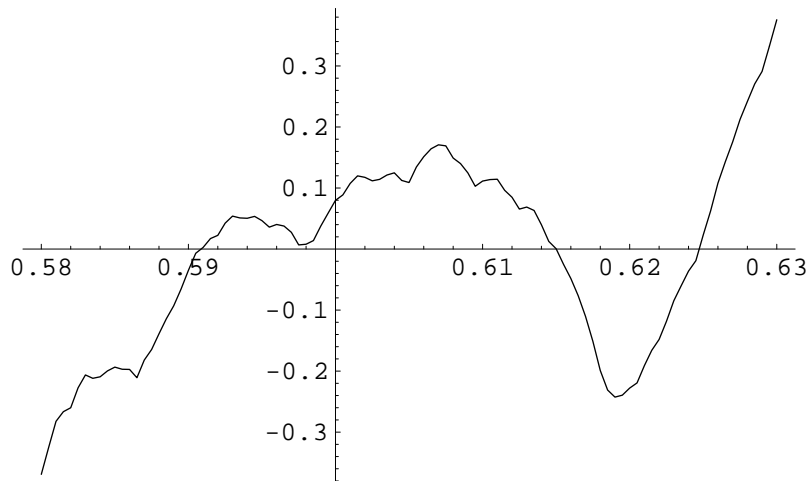

FIGURE 5. Plot of a numerical estimate of the second derivative of the mutual information with respect to p at $p = 1/2$ plotted as a function of λ. This suggests a complex transition between the mutual information having a local maximum and local minimum at $p = 1/2$.

When $\lambda > 1/2$, we have $\mu_\lambda^{(p)}(E_\mathsf{A}) \neq 0$, and the situation is not so clear cut and so far we have been content to study the mutual information in the case of $p = 1/2$. Wherever the maximum of the mutual information occurs at $p = 1/2$ we can carry over the results of Section 3 to make statements about the channel capacity of the simple model. Unfortunately, we are not able to obtain rigorous expressions for the range of λ over which the maximum of the mutual information occurs at $p = 1/2$.

Figure 4 shows a contour plot of a numerical estimate of the mutual information, $R_\lambda^{(p)}$, as a function of λ and p. To obtain the channel capacity from this figure, draw a vertical line at the value of λ which represents the channel and find the maximum value of the mutual information along this line. The figure suggests that for values of λ sufficiently close to 0.5 (the left hand side of the figure), the maximum in the mutual information occurs at $p = 1/2$. Also evident is the fact that at the right hand side of the figure the maximum of the mutual information has shifted to $p \approx 0.3$. As we increase λ going from left to right in the figure the transition seems to be quite complicated.

Figure 5 show a plot, as a function of λ, of numerical estimates of the second derivative of the mutual information evaluated at $p = 1/2$ (for the result on the smoothness of $p \mapsto R_\lambda^{(p)}$ see Proposition 4.3 below). This suggests that for sufficiently small λ ($\lambda \lesssim 0.59$), the mutual information does have a local maximum at $p = 1/2$. Figure 4 and direct plots of the mutual information against p in this region indicate that this maximum is actually the global maximum.

Conversely, Figure 5—in conformity with Figure 4—also suggests that for sufficiently large values of λ ($\lambda \gtrsim 0.625$), the mutual information actually is locally a minimum at $p = 1/2$. The following inequality supports this suggestion.

Proposition 4.1. *For any fixed $p \in (0, 1/2)$ there exists $\lambda \in (g, 1)$ such that $R_\lambda^{(p)} > R_\lambda$.*

Proof. Let us recall that

$$R_\lambda^{(p)} = H_S^{(p)} - H^{(p)}(\text{in}|\text{out}) = H_S^{(p)} - \mu_\lambda^{(p)}(E_\mathsf{A}) H^{(p)}(\text{in}|E_\mathsf{A}).$$

Note that as $\lambda \to 1 - 0$, we have $\mu_\lambda^{(p)}(\mathbf{0}|E_\mathsf{A}) \to p$. This is because, according to Proposition 2.1 and Lemma 2.2,

$$
\begin{aligned}
\mu_\lambda^{(p)}(\mathbf{0}|E_\mathsf{A}) &= \frac{p - \mu_\lambda^{(p)}(E_\mathbf{0})}{\mu_\lambda^{(p)}(E_\mathsf{A})} \\
&= \frac{p - p\nu^{(p)}(\mathfrak{A}_\lambda)}{1 - p\nu^{(p)}(\mathfrak{A}_\lambda) - (1-p)\nu^{(1-p)}(\mathfrak{A}_\lambda)} \to p, \quad \lambda \to 1 - 0,
\end{aligned}
$$

since $\nu^{(q)}(\mathfrak{A}_\lambda) \to 0$ for any $q \in (0,1)$ (similarly to the symmetric case).

Therefore, $H^{(p)}(\text{in}|E_\mathsf{A}) \to H_S^{(p)}$ and

$$R_\lambda^{(p)} \sim H_S^{(p)}(\mu_\lambda^{(p)}(E_\mathbf{0}) + \mu_\lambda^{(p)}(E_\mathbf{1})).$$

Comparing this with the form of R_λ given in Proposition 2.3 we see that it suffices to show the following relation:

$$H_S^{(p)}(p\,\nu^{(p)}(\mathfrak{A}_\lambda) + (1-p)\,\nu^{(1-p)}(\mathfrak{A}_\lambda)) \gg \nu(\mathfrak{A}_\lambda)$$

as λ approaches 1. We will show that in fact

(4.1) $$\nu^{(1-p)}(\mathfrak{A}_\lambda) \gg \nu(\mathfrak{A}_\lambda).$$

We will use the notation of Section 3. Recall that it was shown that the cylinder $[\varepsilon_1 = \varepsilon_2 = \cdots = \varepsilon_N = 0] \subset \mathfrak{A}_\lambda$, whence

$$\nu^{(p)}(\mathfrak{A}_\lambda) \geq (1-p)^N.$$

Furthermore, we know that

$$\nu(\mathfrak{A}_\lambda) \leq 2^{-N+1} \binom{N}{k+2}$$

(see the relation (3.10)). It is left to show that $2^{-N}\binom{N}{k+2} = o\left((1-p)^N\right)$, which follows from

$$\log \binom{N}{k+2} = o(N),$$

which in turn is a consequence of the fact that $\log\binom{N}{k+2} \sim k \log \frac{N}{k}$. Since $N/k \to \infty$ as $\lambda \to 1-0$, this proves (4.1) and thus the claim. $\square$

4.1. **Other properties of $R_\lambda^{(p)}$.** In this subsection we are going to describe and give brief proofs of the properties of $R_\lambda^{(p)}$ that are analogous to the ones of R_λ given in Section 3.

Firstly, the function $\lambda \mapsto \nu^{(p)}(\mathfrak{A}_\lambda)$ is Hölder continuous for each $p \in (0,1)$ (the proof is exactly the same as in Section 3), whence so is the function $\lambda \mapsto R_\lambda^{(p)}$.

Secondly, the following claim about the asymptotic behaviour of $R_\lambda^{(p)}$ holds:

Proposition 4.2. *We have (assuming $p \leq 1/2$):*

(4.2) $$H_S^{(p)} - R_\lambda^{(p)} \asymp (2\lambda - 1)^{\log_2 p^{-1}}, \quad \lambda \to 1/2 + 0,$$

and

(4.3) $$\lim_{\lambda \to 1-0} \frac{(1-\lambda)\log R_\lambda^{(p)}}{\log(1-\lambda)} = -\log(1-p).$$

Proof. Since the cylinders involved in the proof of Proposition 3.5 and Theorem 3.9 are very simple, their product measure can be estimated for a general p in the same way as for $p = 1/2$. We thus omit the proof of (4.3) and give a sketch of the proof of (4.2).

Put

(4.4) $$P = \frac{p - p\nu^{(p)}(\mathfrak{A}_\lambda)}{1 - p\nu^{(p)}(\mathfrak{A}_\lambda) - (1-p)\nu^{(1-p)}(\mathfrak{A}_\lambda)}.$$

Then by Proposition 2.1 and Lemma 2.2,

(4.5) $$R_\lambda^{(p)} = H_S^{(p)} - \mu_\lambda^{(p)}(E_\mathsf{A}) \cdot H_S^{(P)},$$

and $H_S^{(P)} \sim H_S^{(p)}$. Relation (4.2) follows (by the same method as in Section 3) from the inequality

$$\frac{p(1-p)^{m-1}}{p(1-p)^{m-1} + (1-p)p^{m-1}} \leq P \leq \frac{p(1-p)^m}{p(1-p)^m + (1-p)p^m},$$

where m is given by (3.4). Indeed, we have $1 - P \asymp \alpha^m$, where $\alpha = \frac{p}{1-p}$, provided $p < 1/2$. Now a straightforward computation yields $H_S^{(p)} - R_\lambda^{(p)} \asymp p^m$, which is the same as (4.2). $\qquad \square$

Figure 4 indicates that the function $\lambda \mapsto R_\lambda^{(p)}$ is decreasing for each p. Unfortunately, we have not succeeded in proving this rigorously. The behaviour of $R_\lambda^{(p)}$ as a function of p is given by

Proposition 4.3. *For any fixed $\lambda \geq 1/2$, the function $p \mapsto R_\lambda^{(p)}$ is real analytic for $p \in (0, 1)$.*

Proof. By (4.4), (4.5) and the fact that $\mu_\lambda^{(p)}(E_\mathsf{A}) = 1 - p\nu^{(p)}(\mathfrak{A}_\lambda) - (1-p)\nu^{(1-p)}(\mathfrak{A}_\lambda)$, it suffices to show that the function $p \mapsto \nu^{(p)}(\mathfrak{A}_\lambda)$ is analytic for all $p \in (0, 1)$. By definition of the product measure,

$$\nu^{(p)}([\varepsilon_1 = j_1, \ldots, \varepsilon_n = j_n]) = p^{\#\{k: j_k = 0\}}(1 - p)^{\#\{k: j_k = 1\}}.$$

Since we need to take into account only the cylinders of the form $[\varepsilon_1 = j_1, \ldots, \varepsilon_n = j_n] \subset \mathfrak{A}_\lambda$ with $j_n = 0$ such that $[\varepsilon_1 = j_1, \ldots, \varepsilon_n = 1] \not\subset \mathfrak{A}_\lambda$, we have

$$\nu^{(p)}(\mathfrak{A}_\lambda) = \sum_{n=1}^{\infty} \sum_{j=0}^{n} C_{n,j}(\lambda) p^j (1 - p)^{n-j},$$

where $C_{n,j}(\lambda)$ is the number of such cylinders with precisely j zeros. Thus, we have an infinite sum of polynomials, $M_n(p)$, say (we omit the index λ). It suffices to prove that one can rearrange this sum in such a way that it becomes $\sum_{n=0}^{\infty} a_n(\lambda) p^n$, which would imply the analyticity.

To do so, it is in turn sufficient to show that $\deg M_n(p)$ is nondecreasing in n, and that for any $n \geq 1$ there exists $n' \geq 1$ such that $\deg M_{n+n'}(p) > \deg M_n(p)$—then each a_n would be a finite sum of some $C_{k,j}(\lambda)$, which would make such a rearrangement possible.

In other words, we need to show that for any k there exists n such that $\Gamma_n(\lambda)$ (see (3.2)) does not contain words with less than k unities. This is obvious unless there exists an expansion of 1 (= an element of $\Gamma(\lambda)$) with the tail $\mathbf{1}^\infty$; this set is countable and thus does not affect the measure $\nu^{(p)}$ (which is continuous). $\qquad \square$

5. CONCLUDING REMARKS AND OPEN QUESTIONS

At a basic level, our calculations have shown that the model behaves in a physically reasonable manner. We have found that locally, at least, there is an optimum rate at which equiprobable symbols can be input to the channel and that an attempt to push data through the channel at rates significantly greater than this will result in a catastrophic loss of information which is more than sufficient to defeat the purpose (Theorem 3.9 and Corollary 3.7). To an extent it appears that the information loss at high data rates can be offset by use of an asymmetric code in which one symbol is more likely than the other. Thus we find—even for such a very simple channel model—an unusual symmetry-breaking effect. It can be argued that this is a result of the link which exists, *via* the stationary measure, between the input distribution and the conditional measures which relate input to output.
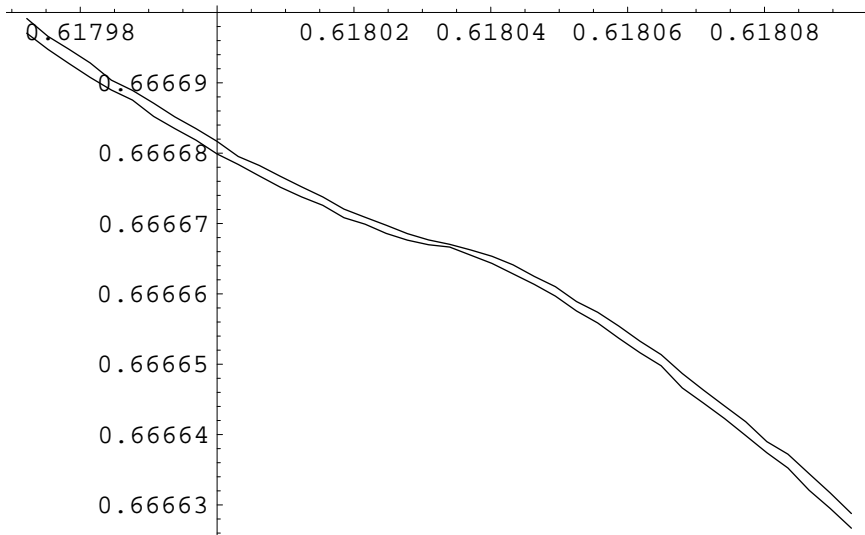
FIGURE 6. As with Figure 1, a plot of a numerical estimate of R_λ as a function of λ showing more detail in the region of $\lambda = g$. The two curves show upper and lower bounds obtained by approximating $\mathfrak{A}_\lambda$ using cylinders.

As an aside, it is worth emphasising the crucial role of the stationary measure. For example, the Lebesgue measure of $E_{\mathbf{0}} \cup E_{\mathbf{1}}$, is $2(1 - \lambda)$ and, although this behaves reasonably in the sense that it decreases from unity to zero as λ is varied from $1/2$ to unity, the use of it in an approximation of the information rate would lead to the erroneous conclusion that $\widetilde{R}$ is a strictly increasing function of the symbol input rate.

There are some mathematical questions for the future that arise from this work and which invite interesting speculations about the physics of digital data transmission.

1. Is it possible to improve the Hölder exponent in (3.3)? For example, let $\lambda = g$. Then taking any subsequence (g_n) such that $|g - g_n| \asymp g^n$, one may show that $|R_{g_n} - R_g| \asymp n 2^{-n}$. The idea of the proof is in the explicit description of $\mathfrak{A}_g$: from the relations $g^2 + g^3 + g^4 + \cdots = g + g^2 = g + g^3 + g^4 = \cdots = 1$ it follows that

$$(5.1) \qquad \mathfrak{A}_g = \bigcup_{j=0}^{\infty} [(10)^j 0] \cup \text{ a countable set,}$$

where $[i_1 \ldots i_s]$ denotes the cylinder $[\varepsilon_1 = i_1, \ldots, \varepsilon_s = i_s]$. Then taking the subsequences g_n and g'_n which are the roots of the equations $x + x^2 + x^n = 1$ and $x^2 + x^3 + \cdots + x^n = 1$ respectively, one can easily estimate the measure of $\mathfrak{A}_{g_n} \setminus \mathfrak{A}_g$ and $\mathfrak{A}_g \setminus \mathfrak{A}_{g'_n}$. We leave the details to the reader.

Thus, in this case,

$$|R_{g'} - R_g| = O(|g' - g|^\alpha), \quad \alpha = \frac{\log 2}{\log 1/\lambda} \approx 1.44,$$

whence $\frac{dR_\lambda}{d\lambda}(g) = 0$ (see Figure 6 for numerical results which suggest this claim). Note that since $\gamma_n(g) = n + 1$ (see [8]), $\gamma(\lambda) = 1$. This means that the actual Hölder exponent of R_λ at $\lambda = g$ is in reality a factor of two larger than the estimate given in (3.3).

Is this true for all λ? If it is, this would have some interesting implications; recall that in [8] it is shown that there exists an uncountable set of $\lambda \in (1/2, g)$, for which $\gamma_n(\lambda) \equiv 1$ (i.e., 1 has a unique expansion in base λ). Hence if one can actually drop 2 in the denominator of α in (3.3), then there will be $2^{\aleph_0}$ points of inflection for R_λ!

On the other hand, it would be interesting to find λ, at which R_λ is not Lipschitz.

2. Is it true that the function $\lambda \mapsto \widetilde{R}_\lambda$ has an infinite number of intervals of monotonicity, as Figure 3 suggests?

3. Figure 3 also suggests that there exists λ_* such that $\widetilde{R}_\lambda$ strictly decreases on $[\lambda_*, 1)$. What is the minimal possible value of such λ_*?

4. Is it possible to "improve" the behaviour of R_λ by considering a nonlinear model? More precisely, we recall that by the famous Erdős theorem, for any λ which is the reciprocal of a *Pisot number* (an algebraic integer greater than 1 whose Galois conjugates are all less than 1 in modulus), μ_λ is singular [7]. It is conceivable that the intervals of monotonicity for the rate capacity $\widetilde{R}_\lambda$ are somehow related to these specific values (see Figure 3).

APPENDIX: A FEW WORDS ABOUT NUMERICS

The numerical calculations reported in this paper were largely obtained by evaluating truncations of the formula

$$(5.2) \qquad R_\lambda = 2\lambda - \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{\sin(2\pi k\lambda)}{k}\, \phi_\lambda(k),$$

where $\phi_\lambda(k)$ is the Fourier transform of μ_λ and is given by the infinite product

$$(5.3) \qquad \phi_\lambda(k) = \operatorname{Re} \prod_{j=0}^{\infty} 1/2 \left(1 + e^{2\pi i k \lambda^j (1-\lambda)}\right).$$

Formula (5.2) is obtained by integrating the Fourier series for the characteristic function of $E_{\mathsf{A}} = [1-\lambda, \lambda]$ with respect to μ_λ. Equation (5.3) is a well-known formula for the moment generating function of the Bernoulli convolution (on $[0,1]$).

Convergence of finite truncations of these two formulae is a delicate issue. This approach was unsophisticated. We studied the effect of various truncations by comparing the different numerical results obtained. We also checked the numerical accuracy against known values of R_λ (for example, $R_\lambda = 2/3$ when $\lambda = g$—see (5.1)). In practise we used 5000 terms in the summation and 200 terms in the product.

In addition to the Fourier series method, we used a numerical technique which is based on the analytical tools used in Section 3 to study $\mathfrak{A}_\lambda$. The idea is use an algorithm to generate systematically cylinder sets which are subsets of $\mathfrak{A}_\lambda$. Each iteration of the computation produces a set of strings of given length which are "suspects"; that is, cylinders which are not subsets of $\mathfrak{A}_\lambda$ but have a nonempty intersection with it (see (3.2)). Given a suspect of length n, say $(\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n)$, we test it by computing the following quantities

$$s_< = \sum_{k=1}^{n} \varepsilon_k \lambda^k$$

and

$$s_> = s_< + \frac{\lambda^{n+1}}{1-\lambda}$$

which we compare with unity. This comparison will yield one of three outcomes:

(1) if $s_> \leq 1$, then $[\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n] \subseteq \mathfrak{A}_\lambda$;
(2) if $s_< \leq 1 < s_>$, then $[\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n] \not\subseteq \mathfrak{A}_\lambda$ and $[\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n] \cap \mathfrak{A}_\lambda \neq \emptyset$;
(3) if $1 < s_<$, then $[\varepsilon_1 \varepsilon_2 \ldots \varepsilon_n] \cap \mathfrak{A}_\lambda = \emptyset$.

In the case of the first outcome the string we have tested corresponds to a cylinder which is a subset of $\mathfrak{A}_\lambda$. In this case we save it and remove it from the list of suspects. If the test gives the second outcome, the string does not correspond to a subset of $\mathfrak{A}_\lambda$ but it is a suspect and so should be retained in the set of suspects. If the string gives the third outcome it is neither a subset nor a suspect and so we remove it from the list of suspects. Each string that remains in the list of suspects is used to generate two new strings by appending a **0** and a **1**. This new set of strings of length $n + 1$ is then tested in the same way.

This recursive process is initiated by testing the two strings: (**0**) and (**1**). The process is stopped when all cylinders up to some string length n_{stop} have been considered. We then calculate a lower bound for $\nu^{(p)}(\mathfrak{A}_\lambda)$ using the cylinders we have found to be subsets of $\mathfrak{A}_\lambda$ and the product measure. An upper bound can be found be calculating the measure of the remaining suspects and adding it to the lower bound.

As far as the images shown in the paper are concerned these two methods are indistinguishable. In principle, the cylinder based method seems better because it provides its own error bounds. In practise however, as the value of λ tends towards unity the number of suspects begins to grow rather quickly with n_{stop} and this limits the utility of the method in this parameter region. On the other hand when $\lambda = g$ the number of suspects grows linearly with n_{stop} and the method provides highly accurate results. Figure 7 shows estimates of $\widetilde{R}_\lambda$ made using this approach. This figure should be compared with Figure 3 which was produced using the Fourier technique. The comparison suggests that even the small detail of Figure 3 is significant. The curves plotted in Figure 6 were also computed using this technique.

References

[1] M. Barnsley, *Fractals Everywhere*, 1988, Academic Press Inc., San Diego.

[2] D. S. Broomhead, J. P. Huke and M. R. Muldoon, *Fractals, Linear Channels and Delay Methods*, Proc. 4th Institute of Mathematics Conference, 'Mathematics of Signal Processing' ed. J.G. McWhirter & I.K. Proudler, Clarendon Press, Oxford, 1998.

[3] D. S. Broomhead, J. P. Huke and M. R. Muldoon, *Digital Channels*, Proceedings of IEEE 2000 Adaptive Systems for Signal Processing, Communications and Control Symposium, 123–128, 2000.

[4] D. S. Broomhead, J. P. Huke, M. R. Muldoon and A. G. Brown, *Nonlinear Thoughts about Linear Signal Processing*, Proc. 5th Institute of Mathematics Conference, 'Mathematics of Signal Processing' ed. J.G. McWhirter & I.K. Proudler, Clarendon Press, Oxford, 2001.

[5] D. S. Broomhead, J. P. Huke, M. R. Muldoon, J. Stark *Iterated Function System Models of Digital Channels* Proc Roy Soc A **460** (2004), appeared electronically 13/7/04.

[6] P. Diaconis and D. Freedman, *Iterated random functions*, SIAM Review, **41** (1999), 45–76.

[7] P. Erdős, *On a family of symmetric Bernoulli convolutions*, Amer. J. Math. **61** (1939), 974–975.

[8] P. Erdős, M. Horvath and I. Joó, *On the uniqueness of the expansions* $1 = \sum q^{-n_i}$, Acta Math. Hungar. **58** (1991), 333–342.

[9] P. Erdős and I. Joó, *On the number of expansions* $1 = \sum q^{-n_i}$, Ann. Univ. Sci. Budapest Eötvös Sect. Math. **35** (1992), 129–132.

[10] P. Erdős, I. Joó and V. Komornik, *Characterization of the unique expansions* $1 = \sum_{i=1}^{\infty} q^{-n_i}$ *and related problems*, Bull. Soc. Math. Fr. **118** (1990), 377–390.
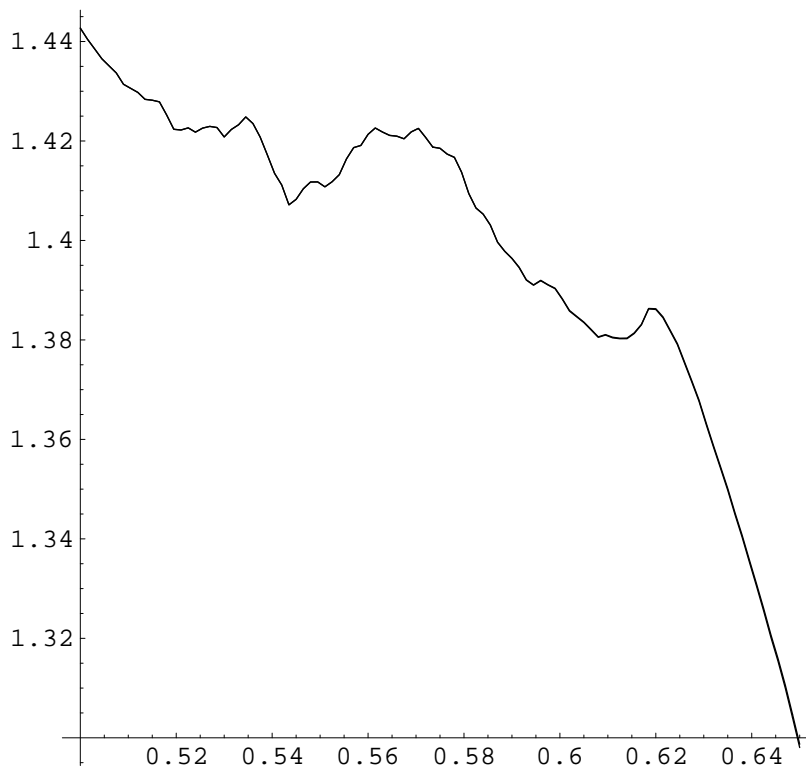
FIGURE 7. A plot of a numerical estimate of $\widetilde{R}_\lambda$ as a function of λ using the cylinder method described in the Appendix. The value of n_{stop} used for this calculation was 20. Although this plot seems to show a single curve there are in fact two curves giving the upper and lower bound of $\widetilde{R}_\lambda$ found in the computation. In this figure the error bounds are comparable with the width of the lines. In places it is just possible to see the upper and lower bounds separately.

[11] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley and Sons Inc., New York, 1990.

[12] P. Glendinning and N. Sidorov, *Unique representations of real numbers in non-integer bases*, Math. Res. Lett. **8** (2001), 535–543.

[13] S. Haykin, *Communication Systems*, John Wiley and Sons Inc., New York, 1994.

[14] J. E. Hutchinson, *Fractals and self-similarity*, Indiana Math. J. **30** (1981), 713–747.

[15] B. Jessen and A. Wintner, *Distribution functions and the Riemann zeta function*, Trans. Amer. Math. Soc. **38** (1938), 48–88.

[16] V. Komornik and P. Loreti, *Unique developments in non-integer bases*, Amer. Math. Monthly **105** (1998), 636–639.

[17] V. Komornik and P. Loreti, *Subexpansions, superexpansions and uniqueness properties in non-integer bases*, preprint.

[18] R. J. McEliece *The Theory of Information and Coding*, Encyclopedia of Mathematics and its Applications, **86**, Cambridge Univ. Press, 2002.

[19] M. J. Nicol, N. Sidorov and D. S. Broomhead *On the fine structure of stationary measures in systems which contract-on-average*, J. Theoret. Prob. **15** (2002), 715–730.

[20] W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hung. **11** (1960), 401–416.

[21] Y. Peres, W. Schlag and B. Solomyak, *Sixty years of Bernoulli convolutions*, Fractal geometry and stochastics, II (Greifswald/Koserow, 1998), Progr. Probab. **46** (2000), 39–65, Birkhauser, Basel.

[22] R. Salem, *A remarkable class of algebraic integers: proof of a conjecture of Vijayaraghavan*, Duke Math. J. **11** (1944), 103–108.

[23] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Tech. J., **27** (1948), 379–656.

[24] N. Sidorov, *Almost every number has a continuum of β-expansions*, Amer. Math. Monthly **110** (2003), 838–842.

[25] N. Sidorov, *Arithmetic Dynamics*, London Math. Soc. Lecture Note Ser. **310** (2003), 145–189.

[26] B. Solomyak, *On the random series $\sum \pm \lambda^i$ (an Erdös problem)*, Ann. of Math. (2) **142** (1995), 611–625.

[27] B. Solomyak, *Notes on Bernoulli convolutions*, preprint (2002), to appear in Proc. Symp. in Pure Math., AMS.

DEPARTMENT OF MATHEMATICS, UMIST, P.O. BOX 88, MANCHESTER M60 1QD, UNITED KINGDOM. E-MAIL: D.S.BROOMHEAD@UMIST.AC.UK

DEPARTMENT OF MATHEMATICS, UMIST, P.O. BOX 88, MANCHESTER M60 1QD, UNITED KINGDOM. E-MAIL: NIKITA.A.SIDOROV@UMIST.AC.UK