<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Asymptotics of entropy rate of hidden Markov chains at weak Black Holes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Han, G; Marcus, B</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2008</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/124816">http://hdl.handle.net/10722/124816</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; Proceedings of IEEE International Symposium on Information Theory. Copyright © IEEE.; ©2008 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.</td>
</tr>
</tbody>
</table>
Asymptotics of Entropy Rate of Hidden Markov Chains at Weak Black Holes

Guangyue Han∗, Brian Marcus†
∗ University of Hong Kong, Email: ghan@maths.hku.hk
† University of British Columbia, Email: marcus@math.ubc.ca

For a more detailed version of this paper, please refer to http://hkumath.hku.hk/ghan

Abstract—We generalize a result in [8] and derive an asymptotic formula for entropy rate of a hidden Markov chain around a “weak Black Hole”. We also discuss applications of the asymptotic formula to certain channels.

I. INTRODUCTION

Consider a finite-state stationary stochastic process \( Y = Y_{\infty} \). The entropy rate of \( Y \) is defined to be

\[
H(Y) = \lim_{n \to \infty} H(Y_n)/(n+1);
\]

here, \( H \) on finite length distributions is taken with the usual definition, with \( \log \) taken to mean the natural logarithm.

If \( Y = \{Y_{\infty}\} \) is a Markov chain with alphabet \( \{1,2,\cdots,B\} \) and transition probability matrix \( \Delta \), it is well known that \( H(Y) \) can be explicitly expressed with the stationary vector of \( Y \) and \( \Delta \). A function \( Z = \{Z_{\infty}\} \) of the Markov chain \( Y \) with the form \( Z = \Phi(Y) \) is called a hidden Markov chain; here \( \Phi \) is a finite valued function defined on \( \{1,2,\cdots,B\} \), taking values in \( A := \{1,2,\cdots,A\} \) (alternatively a hidden Markov chain is defined as a Markov chain observed in noise). For a hidden Markov chain \( Z \), \( H(Z) \) turns out (see Equation (1)) to be the integral of a certain function defined on a simplex with respect to a measure due to Blackwell [4]. However Blackwell’s measure is somewhat complicated and the integral formula appears to be difficult to evaluate in most cases.

Recently, the problem of computing the entropy rate of a hidden Markov chain has drawn much interest, and many approaches have been adopted to tackle this problem. For instance, Blackwell’s measure has been used to bound the entropy rate [14] and a variation on the Birch bound [3] was introduced in [5]. An efficient Monte Carlo method for computing the entropy rate of a hidden Markov chain was proposed independently by Arnold and Loeliger [1], Pfister et. al. [16], and Sharma and Singh [18]. The connection between the entropy rate of a hidden Markov chain and the top Lyapunov exponent of a random matrix product has been observed [10], [11], [12], [6]. In [7], it is shown that under mild positivity assumptions the entropy rate of a hidden Markov chain varies analytically as a function of the underlying Markov chain parameters.

Another recent approach is based on computing the coefficients of an asymptotic expansion of the entropy rate around certain values of the Markov and channel parameters. The first result along these lines was presented in [12], where for a binary symmetric channel with crossover probability \( \varepsilon \) (denoted by BSC(\( \varepsilon \))), the Taylor expansion of \( H(Z) \) around \( \varepsilon = 0 \) is studied for a binary hidden Markov chain of order one. In particular, the first derivative of \( H(Z) \) at \( \varepsilon = 0 \) is expressed very compactly as a Kullback-Liebler divergence between two distributions on binary triplets, derived from the marginal of the input process \( X \). Further improvements and new methods for the asymptotic expansion approach were obtained in [15], [19], [20] and [8]. In [15] the authors express the entropy rate for a binary hidden Markov chain where one of the transition probabilities is equal to zero as an asymptotic expansion including a \( O(\varepsilon \log \varepsilon) \) term. The asymptotic expansion is further generalized in [9], [13].

This paper is organized as follows. In Section II we give an asymptotic formula for the entropy rate of a hidden Markov chain around a “Weak Black Hole”. In Section III, we discuss applications of the formula to certain channels.

II. ASYMPTOTIC FORMULA FOR ENTROPY RATE

Let \( W \) be the simplex, comprising the vectors

\[
\{w = (w_1, w_2, \cdots, w_B) \in \mathbb{R}^B : w_i \geq 0, \sum_i w_i = 1\},
\]

and for \( a \in A \), let \( W_a \) be all \( w \in W \) with \( w_i = 0 \) for \( \Phi(i) \neq a \). For \( a \in A \), let \( \Delta_a \) denote the \( B \times B \) matrix such that \( \Delta_a(i,j) = \Delta(i,j) \) for \( j \) with \( \Phi(j) = a \), and \( \Delta_a(i,j) = 0 \) otherwise. For every \( a \in A \), define the scalar-valued and vector-valued functions \( r_a \) and \( f_a \) on \( W \) by

\[
r_a(w) = w \Delta_a 1,
\]

and

\[
f_a(w) = w \Delta_a / r_a(w).
\]

Note that \( f_a \) defines the action of the matrix \( \Delta_a \) on the simplex \( W \).

If \( Y \) is irreducible, it turns out that

\[
H(Z) = - \int \sum_a r_a(w) \log r_a(w) dQ(w),
\]

where \( Q \) is Blackwell’s measure [4] on \( W \). This measure, which satisfies an integral equation dependent on the parameters of the process, is however very hard to extract from the equation in any explicit way.

Definition 2.1: (see [8]) Suppose that for every \( a \in A \), \( \Delta_a \) is a rank one matrix, and every column of \( \Delta_a \) is either strictly positive or all zeros. We call this the Black Hole case.
It was shown [8] that \( H(Z) \) is analytic around a Black Hole and the derivatives of \( H(Z) \) can be exactly computed around a Black Hole. In this sequel, we consider weakened assumptions and prove an asymptotic formula for entropy rate of a hidden Markov chain around a "weak Black Hole", generalizing the corresponding result in [8].

**Definition 2.2:** Suppose that for every \( a \in A \), \( \Delta_n \) is a rank one matrix. We call this the weak Black Hole case.

**Remark 2.3:** The weak Black Hole condition relaxes the Black Hole condition by eliminating the zero-positivity requirement. The differences and connections between these two conditions are illustrated briefly in simple examples (see section III-A).

In this paper we assume that \( \Delta \) is affinely parameterized by \( \varepsilon \) (\( \varepsilon \geq 0 \)), \( \Delta(\varepsilon) \) is strictly positive when \( \varepsilon > 0 \), and \( \varepsilon = 0 \) corresponds to the weak Black Hole case. Namely the stochastic matrix \( \Delta \) can be written as \( \Delta^0 + \varepsilon \Delta^1 \), where \( \Delta^0 \) and \( \Delta^1 \) are non-negative matrices independent of \( \varepsilon \), \( \Delta^0 \) corresponds to a weak Black Hole, and whenever some entry of \( \Delta^0 \) is zero, the corresponding entry of \( \Delta^1 \) is strictly positive. We use the standard notation: by \( \alpha = \Theta(\beta) \), we mean there exist positive constants \( C_1, C_2 \) such that \( C_1 \beta \leq |\alpha| \leq C_2 \beta \), while by \( \alpha = O(\beta) \), we mean there exists a positive constant \( C \) such that \( |\alpha| \leq C \beta \).

**Proposition 2.4:** For any fixed sequence \( z_{-1}^0 \in A^n \), \( p(z_{-1}^0) \) is the quotient of two polynomials of \( \varepsilon \). Moreover \( p(z_{-1}^0) \) is analytic around \( \varepsilon = 0 \).

**Proof:**

When \( \varepsilon > 0 \), \( \Delta(\varepsilon) \) is strictly positive. By Perron-Frobenius theory [17], \( \Delta(\varepsilon) \) has a unique positive stationary vector, say \( \pi(\varepsilon) \). Since

\[
\text{adj}(I - \Delta(\varepsilon))(I - \Delta(\varepsilon)) = \det(I - \Delta(\varepsilon))I = 0
\]

(here \( \text{adj}(\cdot) \) denotes the adjugate operator on matrices), one can choose \( \pi(\varepsilon) \) to be any normalized row vector of \( \text{adj}(I - \Delta(\varepsilon)) \). So \( \pi(\varepsilon) \) can be written as

\[
\frac{(\pi_0(\varepsilon)), \ldots, \pi_B(\varepsilon))}{\pi_0(\varepsilon) + \pi_1(\varepsilon) + \cdots + \pi_B(\varepsilon)},
\]

where \( \pi_i(\varepsilon) \)'s are polynomials of \( \varepsilon \) and the first non-zero term of every \( \pi_i(\varepsilon) \) has a positive coefficient (we assume terms in the polynomials are increasingly ordered by the degree of \( \varepsilon \) unless otherwise specified). Let \( \text{mdeg}(\cdot) \) denote the degree of the first non-zero term of a polynomial, then we conclude that for each \( i \)

\[
\text{mdeg} (\pi_i(\varepsilon)) \geq \text{mdeg} (\pi_0(\varepsilon) + \cdots + \pi_B(\varepsilon)),
\]

and thus \( \pi(\varepsilon) \), which is uniquely defined on \( \varepsilon > 0 \), can be continuously extended to \( \varepsilon = 0 \) via setting \( \pi(0) = \lim_{\varepsilon \to 0} \pi(\varepsilon) \).

Now

\[
p(z_{-1}^0) = \pi(\varepsilon)\Delta_{-1}^0 \cdots \Delta_{-1}^1 = \frac{(\pi_0(\varepsilon)), \ldots, \pi_B(\varepsilon))\Delta_{-1}^0 \cdots \Delta_{-1}^1}{\pi_0(\varepsilon) + \pi_1(\varepsilon) + \cdots + \pi_B(\varepsilon)} = f(\varepsilon) g(\varepsilon),
\]

where \( \text{mdeg}(f(\varepsilon)) \geq \text{mdeg}(g(\varepsilon)) \). It then follows that \( p(z_{-1}^0) \) is analytic around \( \varepsilon = 0 \).

**Remark 2.5:** It immediately follows from Proposition 2.4 that there exists \( k \geq 0 \) such that \( p(z_{-1}^0) = \Theta(\varepsilon^k) \).

**Proposition 2.6:** For any fixed sequence \( z_{-1}^0 \in A^{n+1} \), \( p(z_0 z_{-1}^0) \) is the quotient of two polynomials of \( \varepsilon \). Moreover \( p(z_0 z_{-1}^0) \) is analytic around \( \varepsilon = 0 \), furthermore either \( p(z_0 z_{-1}^0) = \Theta(1) \) or \( p(z_0 z_{-1}^0) = \Theta(\varepsilon) \).

**Proof:**

Let \( x_{i-1,n} = x_{i-1,n}(z_0 z_{-1}^0) = p(y_i = \cdot | z_{i-1,n}^0) \), where \( \cdot \) denotes the possible states of Markov chain \( Y \). Then one checks that

\[
p(z_0 z_{-1}^0) = x_{-1,n} \Delta_{-1}^0 1
\]

and

\[
x_{i-1,n} = \frac{x_{i-1,n} \Delta_{-1}^0 1}{x_{i-1,n} \Delta_{-1}^0 1}, \quad -n \leq i \leq -1.
\]

Because \( \Delta \) is affinely parameterized by \( \varepsilon \) (\( \varepsilon \geq 0 \)) and \( \Delta(\varepsilon) \) is strictly positive when \( \varepsilon > 0 \), inductively we can prove (the proof is similar to the proof of Proposition 2.4) that for any \( i \), \( x_{i-1,n} \) can be written as follows:

\[
x_{i-1,n} = \left( f_1(\varepsilon), f_2(\varepsilon), \ldots, f_B(\varepsilon) \right)
\]

where \( f_i(\varepsilon) \)'s are certain polynomials of \( \varepsilon \) such that for each \( i \)

\[
\text{mdeg}(f_i(\varepsilon)) \geq \text{mdeg}(f_1(\varepsilon) + f_2(\varepsilon) + \cdots + f_B(\varepsilon)).
\]

The existence of the Taylor series expansion of \( x_{i-1,n} \) around \( \varepsilon = 0 \) (for any \( i \)) then follows. Together with (3), we conclude that \( p(z_0 z_{-1}^0) \) is analytic around \( \varepsilon = 0 \).

Now consider the Taylor series expansion of \( x_{-1,n} \) around \( \varepsilon = 0 \),

\[
x_{-1,n} = a_0(z_{-1}^0) + a_1(z_{-1}^0) \varepsilon + a_2(z_{-1}^0) \varepsilon^2 + \cdots
\]

Since for any \( \varepsilon, x_{i-1,n} \in W \), we conclude that \( a_0(z_{-1}^0) \geq 0 \), but \( a_0(z_{-1}^0) \neq 0 \).

In the following we write for each \( i \)

\[
\Delta_{-1}^0 + \varepsilon \Delta_{-1}^1,
\]

here \( \Delta_{-1}^0 \) and \( \Delta_{-1}^1 \) are non-negative matrices independent of \( \varepsilon \). If \( a_0(z_{-1}^0) \Delta_{-1}^0 1 > 0 \), then

\[
p(z_0 z_{-1}^0) = x_{-1,n} \Delta_{-1}^0 1 = a_0(z_{-1}^0) \Delta_{-1}^0 1 + O(\varepsilon) = \Theta(1).
\]

Now consider the case when \( a_0(z_{-1}^0) \Delta_{-1}^0 1 = 0 \). Because \( \Delta_{-1}^0 + \varepsilon \Delta_{-1}^1 \) is strictly positive for \( \varepsilon > 0 \),

\[
a_0(z_{-1}^0) \Delta_{-1}^0 1 + \varepsilon \Delta_{-1}^1 1 > 0,
\]

which implies \( a_0(z_{-1}^0) \Delta_{-1}^0 1 > 0 \). On the other hand, for \( \varepsilon > 0 \)

\[
x_{-1,n} \Delta_{-1}^0 1 = a_0(z_{-1}^0) + a_1(z_{-1}^0) \varepsilon + \cdots \Delta_{-1}^1 1 > 0,
\]

which implies \( a_1(z_{-1}^0) \Delta_{-1}^0 1 > 0 \). So in this case,

\[
p(z_0 z_{-1}^0) = a_0(z_{-1}^0) \Delta_{-1}^0 1 + a_1(z_{-1}^0) \Delta_{-1}^1 1 \varepsilon + O(\varepsilon^2).
\]

It then follows that \( p(z_0 z_{-1}^0) = \Theta(\varepsilon) \), since the coefficient of \( \varepsilon \) has been proven to be strictly positive.
Lemma 2.7: Consider two formal series expansion $f(x), g(x) \in \mathbb{R}[x]$ such that $f(x) = \sum_{i=0}^{\infty} f_i x^i$ and $g(x) = \sum_{i=0}^{\infty} g_i x^i$, where $g_0 \neq 0$. Let $h(x) \in \mathbb{R}[x]$ be the quotient of $f(x)$ and $g(x)$ with $h(x) = \sum_{i=0}^{\infty} h_i x^i$. Then $h_i$ is a function only dependent on $f_0, \ldots, f_i$ and $g_0, \ldots, g_i$.

Proof:
Comparing the coefficients of all the terms in the following identity:
\[
\left( \sum_{i=0}^{\infty} h_i x^i \right) \left( \sum_{i=0}^{\infty} g_i x^i \right) = \sum_{i=0}^{\infty} f_i x^i,
\]
we obtain that for any $i$,
\[
h_0 g_i + h_1 g_{i-1} + \cdots + h_i g_0 = f_i.
\]
The lemma then follows from an induction (on $i$) argument.

Let $i$ be a fixed non-negative integer and let $s$ be a function defined on all hidden Markov strings $z_{-m}^i$ for $-m \leq i$. We say that $s$ stabilizes at a particular string $z_{-m}^i$ if for all $m \geq n$ and hidden Markov strings $z_{-m}^i$ such that $z_{-m}^i = z_{-m}^n$, we have
\[
s(z_{-m}^i) = s(z_{-m}^n).
\]

By Proposition 2.6, for any hidden Markov string $z_{-m}^0$, the Taylor series expansion of $p(z_0|z_{-m}^0)$ around $\varepsilon = 0$ exists. We use $b_j(z_{-m}^0)$ to represent the coefficient of $\varepsilon^j$ in the expansion, namely
\[
p(z_0|z_{-m}^0) = b_0(z_{-m}^0) + b_1(z_{-m}^0)\varepsilon + b_2(z_{-m}^0)\varepsilon^2 + \cdots.
\]

Lemma 2.8: Consider a hidden Markov chain $Z$ at a weak Black Hole corresponding to $\varepsilon = 0$. Suppose that for a fixed sequence $x_{-n}^0$, $p(z_{-m}^0) = \Theta(x^k)$ for some $k \geq 0$. Then the first $n-2k$ coefficients of $p(z_0|z_{-m}^0)$ are stabilized at $z_{-m}^0$, namely, for $j$ with $0 \leq j < n-2k-1$, the coefficient $b_j(z_{-m}^0)$ is stabilized at $z_{-m}^0$.

Proof: Recall that $x_{i,-m} = x_{i,-m}(z_{i,-m}^0) = p(y_i = \cdot|z_{i,-m}^0)$, where $\cdot$ denotes the possible states of Markov chain $Y$. Consider the Taylor series expansion of $x_{i,-m}$ around $\varepsilon = 0$,
\[
x_{i,-m} = a_0(z_{i,-m}^0) + a_1(z_{i,-m}^0)\varepsilon + a_2(z_{i,-m}^0)\varepsilon^2 + \cdots.
\]
Assume that $p(z_{i,-m}^0) = \Theta(x^k_0)$ ($0 \leq k \leq k$), we shall show that $x_{i,-m}$ has the first $n+i+1-2k$ coefficients stabilized at $z_{-m}^0$, namely for $j$ with $0 \leq j < n+i+1-2k$, $a_j(z_{i,-m}^0)$ is stabilized at $z_{-m}^0$.

We proceed by induction on $i$ (from $-n$ to $-1$). The case when $i = -n$ is straightforward. Now we consider $i > -n$ and we suppose that $a_j(z_{i,-m}^0)$ ($0 \leq j < n+i-2k$) is stabilized at $z_{-m}^0$. Recall that for each $i$
\[
\Delta_i = \Delta_i^0 + \varepsilon \Delta_i^1,
\]
where $\Delta_i^0$ and $\Delta_i^1$ are non-negative matrices independent of $\varepsilon$. Note that with this notation, we have
\[
x_{i+1,-m} = x_{i,-m} \frac{\Delta_{i+1}}{x_{i,-m} \Delta_{i+1}}
\]
\[
= (a_0(z_{i,-m}^0) + a_1(z_{i,-m}^0)\varepsilon + \cdots) (\Delta_{i+1}^0 + \varepsilon \Delta_{i+1}^1) = (a_0(z_{i,-m}^0) + a_1(z_{i,-m}^0)\varepsilon + \cdots) (\Delta_{i+1}^0 + \varepsilon \Delta_{i+1}^1) + O(\varepsilon).
\]

It follows from Proposition 2.6 that either $k_{i+1} = k_i + 1$ or $k_{i+1} = k_i$. If $k_{i+1} = k_i + 1$, we have for $m \geq n$
\[
a_0(z_{i,-m}^0) \Delta_{i+1}^0 = a_0(z_{i,-m}^0) \Delta_{i+1}^0 = 0.
\]

Applying Lemma 2.7 to expression (8), we conclude that for all $j$, $a_j(z_{i,-m}^0)\varepsilon^j$ depends only on
\[
a_0(z_{i,-m}^0), a_1(z_{i,-m}^0), \ldots, a_{j+1}(z_{i,-m}^0), \Delta_{i+1}^0, \Delta_{i+1}^1.
\]

Thus for $j$ with $0 \leq j < n+i-2k$, $a_j(z_{i,-m}^0)$ is stabilized at $z_{-m}^0$. This implies that the first $n+i-2k$ coefficients of $x_{i+1,-m}$ are stabilized at $z_{-m}^0$, namely for $j$ with $0 \leq j < n+i-2k$, the coefficient $a_j(z_{i,-m}^0)$ is stabilized at $z_{-m}^0$.

If $k_{i+1} = k_i$, we have
\[
a_0(z_{i,-m}^0) \Delta_{i+1}^0 = \Delta_{i+1}^0 \neq 0.
\]

Again by Lemma 2.7 applied to expression (8), for any $j$, $a_j(z_{i,-m}^0)\varepsilon^j$ depends only on
\[
a_0(z_{i,-m}^0), a_1(z_{i,-m}^0), \ldots, a_{j+1}(z_{i,-m}^0), \Delta_{i+1}^0, \Delta_{i+1}^1.
\]

Thus, for any $j$ with $0 \leq j < n+i-2k$, $a_j(z_{i,-m}^0)$ is stabilized at $z_{-m}^0$. Now, let $j = n+i-1+2k_i+1$, then we have
\[
a_j(z_{i+1}^0) = \left( a_0(z_{i,-m}^0) \Delta_{i+1}^0 \right)^2 + \text{other terms},
\]
here
\[
* = a_{n+i+2k_i}(z_{i,-m}^0) \Delta_{i+1}^0 a_0(z_{i,-m}^0) \Delta_{i+1}^0 1
\]
\[-a_0(z_{i,-m}^0) \Delta_{i+1}^0 a_{n+i+2k_i}(z_{i,-m}^0) \Delta_{i+1}^0 1.
\]

Since $\Delta_{i+1}^0$ is a rank one matrix, we find $* = 0$ and “other terms” are functions of
\[
a_0(z_{i,-m}^0), a_1(z_{i,-m}^0), \ldots, a_{n+i+2k_i}(z_{i,-m}^0), \Delta_{i+1}^0, \Delta_{i+1}^1.
\]

So we conclude $a_j(z_{i+1}^0)$ is also stabilized at $z_{-m}^0$, and thus $x_{i+1,-m}$ has the first $n+i+1+2k_i = n+i+1+1-2k_i$ coefficients stabilized at $z_{-m}^0$.

The lemma then immediately follows from (3) and the proven fact that $x_{i,-m}$ has the first $n-2k_i = n-2k$ coefficients stabilized at $z_{-m}^0$.
where \( j \)’s and \( g \)’s are functions only dependent on \( \Delta \), the transition probability matrix of the underlying Markov chain \( Y \).

**Proof:**

First we fix a large \( n \) and consider the Birch upper bound on \( H(Z) \)

\[
H_n(Z) = H(Z_0|Z_n^{-1}) = -\sum p(z_0|z_n^{-1}) \log p(z_0|z_n^{-1}).
\]

Note that for \( j \geq k + 1 \),

\[
\left| \sum p(z_0|z_n^{-1}) \log p(z_0|z_n^{-1}) \right| = O(\varepsilon^{k+1}).
\]

So, in the following we only consider the sequences \( z_n \) with \( p(z_n^{-1}) = \Theta(\varepsilon^j), j \leq k \). For such sequences, by Lemma 2.8, as long as \( n \) is large enough, \( p(z_0|z_n^{-1}) \) will have sufficiently many initial coefficients stabilized at \( z_n^{-1} \). Moreover, for any sequence \( z_n \) with \( n \geq 3k + 1 \), \( p(z_n^{-1}) = \Theta(\varepsilon^j), p(z_n^{-1}) = \Theta(1) \), we have

\[
|\log p(z_0|z_n^{-1}) - \log p(z_n^{-1})| = O(\varepsilon^{k+1});
\]

for any sequence \( z_n \) with \( n \geq 3k + 2 \), \( p(z_n^{-1}) = \Theta(\varepsilon^j) \), \( p(z_0|z_n^{-1}) = \Theta(\varepsilon^j) \), we have

\[
|\log p(z_0|z_n^{-1}) - \log p(z_n^{-1}|z_0) - \log p(z_0) - \log p(z_n^{-1})| = O(\varepsilon^{k+1}).
\]

Now fix \( n_0 \geq 2k + 1 \). For any \( n \geq n_0 \), using (10) and (11), we can show

\[
H_n(Z) = -\sum p(z_0) \log p(z_0|z_n^{-1}) + O(\varepsilon^{k+1})
\]

\[
= H(Z) + \sum \sum p(z_0) f_j \varepsilon^j + O(\varepsilon^{k+1}),
\]

where \( f_j \) and \( g \) are functions dependent only on \( Y \).

Using Remark 2.9, one can apply similar analysis to the Birch lower bound

\[
\bar{H}_n(Z) = H(Z_0|Z_n^{-1}Y_n^{-1}).
\]

For the same \( n_0 \), one can show that \( \bar{H}_n(Z) \) takes the same form as \( H_n(Z) \) in (12) with exactly the same coefficients of \( \varepsilon^j \) for \( n \geq 2k + 1 \) when \( n \) is sufficiently large. We thus prove the theorem.

**Remark 2.11:** Note that as a Black Hole, \( g_j = 0 \) for all \( j \)’s. So Formula (9) is consistent with the Taylor series expansion of \( H(Z) \) around a Black Hole.

**III. APPLICATIONS TO CERTAIN CHANNELS**

**A. Binary Markov Chains Corrupted by BSC(\( \varepsilon \))**

Consider a binary symmetric channel with crossover probability \( \varepsilon \). At time \( n \) the channel can be characterized by the following equation

\[
Z_n = X_n ⊕ E_n,
\]

where \( X_n \) denotes the input process, \( ⊕ \) denotes binary addition, \( E_n \) denotes the i.i.d. binary noise with \( p(0) = 1 - \varepsilon \) and \( p(1) = \varepsilon \), and \( Z_n \) denotes the corrupted output.

Now, suppose \( X = \{X_n\} \) is a first order Markov chain with the transition probability matrix

\[
\Pi = \begin{bmatrix}
\pi_{00} & \pi_{01} \\
\pi_{10} & \pi_{11}
\end{bmatrix}.
\]

Then \( Y = \{Y_n\} = \{\{X_n, E_n\}\} \) is jointly Markov with transition probability matrix:

\[
\Delta = \begin{bmatrix}
\pi_{00}(1 - \varepsilon) & \pi_{01}(1 - \varepsilon) & \pi_{01}\varepsilon \\
\pi_{10}(1 - \varepsilon) & \pi_{10}(1 - \varepsilon) & \pi_{11}\varepsilon \\
\pi_{10}(1 - \varepsilon) & \pi_{10}(1 - \varepsilon) & \pi_{11}(1 - \varepsilon)
\end{bmatrix},
\]

and \( Z = \Phi(Y) \) is a hidden Markov chain with \( \Phi(0, 0) = \Phi(1, 0) = \Phi(0, 1) = \Phi(1, 1) = 0 \). When \( \varepsilon = 0 \), one checks that both \( \Delta_0 \) and \( \Delta_1 \) have rank one. If \( \pi_{ij} \)’s are all positive, then we have a Black Hole case, for which one can derive the Taylor series expansion of \( H(Z) \) around \( \varepsilon = 0 \) [19], [8]; if some \( \pi_{ij} \)’s are zeros, then this is a weak Black Hole case, for which Theorem 2.10 can be applied and an asymptotic formula for \( H(Z) \) can be derived.

For instance, consider a first order Markov chain \( X \) with the following transition probability matrix

\[
\begin{bmatrix}
1 - p & p \\
1 & 0
\end{bmatrix},
\]

where \( 0 < p < 1 \). It has been shown [15] that

\[
H(Z) = H(X) - \frac{p(2 - p)}{1 + p} \log \varepsilon + O(\varepsilon),
\]

as \( \varepsilon \rightarrow 0 \). This result has been further generalized [9], [13] to the following formula:

\[
H(Z) = H(X) + f(X) \varepsilon \log(1/\varepsilon) + g(X) \varepsilon + O(\varepsilon^2 \log \varepsilon),
\]

where \( X \) is the input Markov chain of any order, \( Z \) is the output process obtained by passing \( X \) through a BSC(\( \varepsilon \)), and \( f(X) \) and \( g(X) \) are explicitly computed.

Theorem 2.10 claims that higher order asymptotic terms together with their coefficients can be derived as well. Now suppose the input is an \( m \)-th order Markov chain \( X \) defined by the transition probabilities \( P(X_i = a_0|X_{i-1} = a_{-1}) \), \( a_{-m} = X^m, \) \( X^m \)

\[
= \{(0, 1), \delta(0, \varepsilon), \delta(1, \varepsilon)\},
\]

\[
X_i = X_{-m+1}^{m+1}, \hat{E}_i = E_{-m+1}^{m+1}, \hat{Z}_i = Z_{-m+1}^{m+1},
\]

where \( \hat{X}_i \) is a first order Markov chain with state space \( X^m \), whose transition probability from state \( \hat{x} \in X^m \) to \( \hat{y} \in \delta(\hat{y}, \hat{e}) \) is \( \hat{\Delta}(\hat{x}, \hat{y}, \hat{e}) \) is

\[
\hat{\Delta}(\hat{x}, \hat{y}, \hat{e}) = P_{\hat{y} \hat{e}} \prod_{i=1}^m \delta(\hat{e}_i, \hat{e}),
\]

where \( \hat{\Delta} \) denotes the transition probability matrix of \( \hat{X}, \hat{E} \) and

\[
\delta(0, \varepsilon) = 1 - \varepsilon, \quad \delta(1, \varepsilon) = \varepsilon.
\]
It is easy to check that when $\varepsilon = 0$, $\hat{\Delta}$ is a rank one matrix for every $\hat{z} \in \mathcal{X}^m$. In particular, when $P(X_t = a_0 | X_{t-m} = a_{-m}^{-1}) > 0$ for $a_{-m}^{-1} \in \mathcal{X}^{m+1}$, we have a Black Hole. In this case, the derivatives of the entropy rate $H(\hat{Z})$ at $\varepsilon = 0$ can be exactly computed. Then together with

$$H(Z) = H(\hat{Z})/m,$$

we conclude that the derivatives of the entropy rate $H(Z)$ at $\varepsilon = 0$ can be exactly computed as well. When $P(X_t = a_0 | X_{t-m} = a_{-m}^{-1}) = 0$ for some $a_{-m}^{-1} \in \mathcal{X}^{m+1}$, we have a weak Black Hole. In this case, Theorem 2.10 can be applied and an asymptotical formula for $H(Z)$ around $\varepsilon = 0$ can be obtained.

**B. Binary Markov Chains Corrupted by BEC(\delta)**

Consider a binary erasure channel with fixed erasure rate $\delta$ (denoted by BEC(\delta)). We say the channel is in state 1 if the input digit is erased after passing through the channel ($\varepsilon$ will be used to denote an erasure), otherwise we say the channel is in state 0. Let $C$ denote the channel state, and let $X$ denote the first order input Markov chain with transition probability matrix $\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}$, and let $Z$ denote the output process. Then $Y = (X, C)$ is jointly Markov with

$$\Delta = \begin{bmatrix} \pi_{00}(1 - \delta) & \pi_{00}\delta & \pi_{01}(1 - \delta) & \pi_{01}\delta \\ \pi_{10}(1 - \delta) & \pi_{10}\delta & \pi_{11}(1 - \delta) & \pi_{11}\delta \\ \pi_{00}(1 - \delta) & \pi_{00}\delta & \pi_{01}(1 - \delta) & \pi_{01}\delta \\ \pi_{10}(1 - \delta) & \pi_{10}\delta & \pi_{11}(1 - \delta) & \pi_{11}\delta \end{bmatrix},$$

and $Z = \Phi(Y)$ is hidden Markov with $\Phi(0, 1) = \Phi(1, 1) = e$, $\Phi(0, 0) = 0$ and $\Phi(1, 0) = 1$.

Now one checks that if $\Pi$ is a rank one matrix, $\Delta_0, \Delta_1, \Delta_\varepsilon$ will all be rank one matrices, then we have a Black Hole case. In particular, if we use the following parameterization:

$$\pi_{00} = k/(k+1) + \varepsilon, \pi_{10} = k/(k+1) - \varepsilon, \pi_{01} = 1/(k+1) - \varepsilon, \pi_{11} = 1/(k+1) + \varepsilon,$$

where $k \geq 0$, then we have a Black Hole case corresponding to $\varepsilon = 0$.

**C. Binary Markov Chains Corrupted by Gilbert-Elliott Channel**

Consider a binary Gilbert-Elliott channel. The channel is said to be in state 0 if it behaves like a BSC($\varepsilon_0$) and state 1 if it behaves like a BSC($\varepsilon_1$). The channel state varies as a Markov chain with transition probability matrix:

$$C = \begin{bmatrix} c_{00} & c_{01} \\ c_{10} & c_{11} \end{bmatrix}.$$

Let $X$ denote the first order input Markov chain with transition probability matrix $\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}$, and let $Z$ denote the output process. Then $Y = (X, C, E)$ is jointly Markov with a $8 \times 8$ transition probability matrix $\Delta$ and $Z = \Phi(X, C, E)$ is hidden Markov with

$$\Phi(0, 0, 0) = \Phi(0, 1, 0) = \Phi(1, 0, 1) = \Phi(1, 1, 1) = 0,$$

$\Phi(0, 0, 1) = \Phi(0, 1, 1) = \Phi(1, 1, 0) = \Phi(1, 1, 0) = 1$.

If $\varepsilon_0 = \varepsilon_1 = 0$ and $C$ is a rank one matrix, both $\Delta_0$ and $\Delta_1$ will be rank one matrices and we have a weak Black Hole. In particular, if we use the following parameterization: $\varepsilon_0 = k_1\varepsilon, \varepsilon_1$ and $c_{00} = k_2/(k_2 + 1) + \varepsilon, c_{10} = k_2/(k_2 + 1) - \varepsilon$, $c_{01} = 1/(k_2 + 1) - \varepsilon, c_{11} = 1/(k_2 + 1) + \varepsilon$, where $k_1, k_2 \geq 0$, then we have a Black Hole case corresponding to $\varepsilon = 0$.

**REFERENCES**


