Asymptotic Filtering and Entropy Rate of a Hidden Markov Process in the Rare Transitions Regime

Chandra Nair
Dept. of Elect. Engg.
Stanford University
Stanford CA 94305, USA
mchandra@stanford.edu

Erik Ordentlich
Information Theory Research Group
HP Laboratories
Palo Alto CA 94304, USA
erik.ordentlich@hp.com

Tsachy Weissman
Dept. of Elect. Engg.
Stanford University
Stanford CA 94305, USA
tsachy@stanford.edu

***Abstract*— Recent work by Ordentlich and Weissman put forth a new approach for bounding the entropy rate of a hidden Markov process via the construction of a related Markov process. We use this approach to study the behavior of the filtering error probability and the entropy rate of a hidden Markov process in the rare transitions regime. In this paper, we restrict our attention to the case of a two state Markov chain that is corrupted by a binary symmetric channel. Using this approach we recover the results on the optimal filtering error probability of Khasminskii and Zeitouni. In addition, this approach sheds light on the terms that appear in the expression for the optimal filtering error probability. We then use this approach to obtain tight estimates of the entropy rate of the process in the rare transitions regime. This leads to tight estimates on the capacity of the Gilbert-Elliot channel in the rare transitions regime.**

I. Introduction

Consider a stationary, finite-alphabet, Markov chain, $\{X_k\}$, and let $\{Z_k\}$ denote a noisy version when corrupted by a discrete memoryless channel. Let $\mathcal{K}$ denote the transition kernel of the Markov chain and $\mathcal{C}$ denote the channel transition matrix. The process $\{Z_k\}$ is known as a *hidden Markov process*, with the $\{X_k\}$ corresponding to the state process.

Hidden Markov processes occur naturally in the modeling of information sources [EM02]. They also arise as noise processes in additive noise channels, like the Gilbert-Elliot channel. It has been shown in [MBD89] that the characterization of the channel capacity for the Gilbert-Elliot channel boils down to finding the entropy rate of the noise.

Early work on the estimation of the underlying state (source) symbols from a hidden Markov process involved analysis of optimal filters [W65]. Later, sub-optimal filters were used to derive upper bounds [KL92], [KZ96] and information-theoretic arguments were used to obtain lower bounds [KZ96]. The lower and upper bounds in [KZ96] matched in the region of rare transitions and the optimal filtering error probability was obtained.

Our approach here is quite different; we use an alternative Markov process proposed in [OW04] to study the behavior of the optimal filter and use this to get tight estimates of the filtering error probability in the rare transitions regime. The analysis of the alternative Markov process also lays bare the terms that arise in the filtering error probability as obtained in [KZ96].

Work on the entropy rate of hidden Markov models used bounds [CT91], Monte Carlo simulations [HGG03], Lyapunov exponents [HGG03], [JSS04], Statistical Mechanics [ZKD04], and more [EBTBH04].

The analysis of the alternative Markov process proposed in [OW04], simultaneously provides us with the optimal filtering error probability as well as the entropy rate under the rare transitions regime. We perform the analysis for the simplest case of a symmetric 2-state Markov chain corrupted by a Binary Symmetric Channel (BSC). The extension to general finite alphabet Markov processes can be attempted along very similar lines, however the technical details of the arguments become more involved. The analysis of this toy model, however, sheds light on the behavior of the finite alphabet process.

The paper is organized as follows. In Section II we present the source and channel model and the alternative Markov process defined in [OW04]. Section III presents the main results of this paper and Section IV illustrates the analysis of the alternative Markov process that yields the claims. We conclude in Section V

II. The BSC-Corrupted Binary Markov Chain

Consider a source, X_k, that behaves according to a binary valued symmetric Markov chain with probability of transition π. Assume that the source symbols pass through a memoryless channel that flips the value of X_k with probability δ to give a corrupted sequence Z_k.

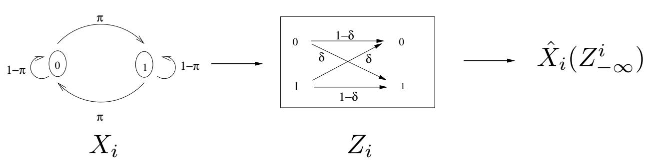

Fig. II.1. Source and Channel model

For this model the Markov transition kernel and the channel transition matrix are, respectively,

$$\mathcal{K} = \begin{pmatrix} 1-\pi & \pi \\ \pi & 1-\pi \end{pmatrix}, \quad \mathcal{C} = \begin{pmatrix} 1-\delta & \delta \\ \delta & 1-\delta \end{pmatrix}, \quad \text{(II.1)}$$

and we assume without loss of generality that $\delta \leq 1/2$.

Define the conditional probability of the source symbol conditioned on the entire set of output symbols by

$$\begin{aligned}\beta_k(1) &= \mathbb{P}(X_k = 1 | Z_{-\infty}^k) \\ \beta_k(0) &= \mathbb{P}(X_k = 0 | Z_{-\infty}^k).\end{aligned} \tag{II.2}$$

The log-likelihood ratio of the source given the present and past channel outputs is defined as

$$l_k = \ln \frac{\beta_k(1)}{\beta_k(0)} = \ln \frac{\beta_k(1)}{1 - \beta_k(1)}. \tag{II.3}$$

Let us consider the distribution of the log-likelihood ratio random variable conditioned on the event that $X_k = 1$. We now recall some relevant results that were established in [OW04].

Consider the following auto-regressively defined first order Markov process,

$$Y_k = r_k \ln \frac{1-\delta}{\delta} + s_k h(Y_{k-1}). \tag{II.4}$$

Here, $\{r_k\}$ and $\{s_k\}$ are independent i.i.d. sequences with

$$r_k = \begin{cases} -1 & \text{w.p. } \delta \\ +1 & \text{w.p. } 1-\delta \end{cases} \quad s_k = \begin{cases} -1 & \text{w.p. } \pi \\ +1 & \text{w.p. } 1-\pi \end{cases} \tag{II.5}$$

and the function $h(x)$ is given by

$$h(x) = \ln \frac{e^x(1-\pi) + \pi}{e^x \pi + 1 - \pi}. \tag{II.6}$$

It was shown in [OW04] that the unique stationary distribution of this 1st-order Markov process is given by $\mathbb{P}(l_k | X_k = 1)$. Let Y denote a random variable distributed according to the stationary distribution of the Markov process in (II.4).

Observe that the optimal filter estimates the source symbol to be 1 if the log-likelihood is positive and 0 if it is negative. Therefore, the probability of error for the optimal filtering estimator is given by

$$\mathcal{E}_{min} = \mathbb{P}(l_k < 0 | X_k = 1) = \mathbb{P}(Y < 0). \tag{II.7}$$

Consider the binary entropy function $h_b(x)$ (in nats) defined according to

$$h_b(x) = -x \ln x - (1-x)\ln(1-x). \tag{II.8}$$

Let $p*q = p(1-q)+q(1-p)$ denote the binary convolution. It was shown in [OW04] that the entropy rate of the hidden Markov process, $\{Z_k\}$, is given by

$$\bar{H}(Z) = \mathbb{E} h_b\left(\frac{e^Y}{1+e^Y} * \pi * \delta\right). \tag{II.9}$$

Thus it is clear from equations (II.7) and (II.9) that the filtering error probability and the entropy rate of the hidden Markov process are intimately connected to the stationary distribution of the *Alternative Markov Process* defined in equation (II.4).

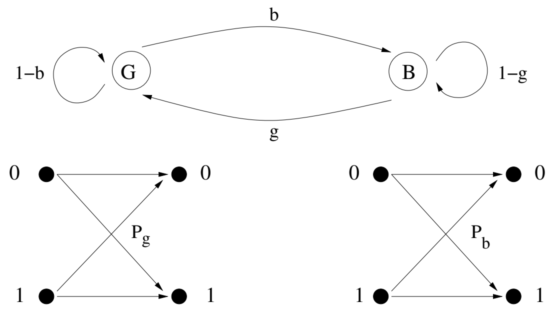

Fig. II.2. Gilbert-Elliot Channel model

1) Gilbert-Elliot Channel:: The Gilbert-Elliot channel model is described by the transition diagram on Figure II.2.

The channel exists in a good state or bad state, as determined by the 2-state Markov chain. If the channel is in the good state, the channel transition matrix, $\mathcal{C}$, behaves like a BSC with parameter P_g and if it is in the bad state, $\mathcal{C}$ behaves like a BSC with parameter P_b.

To make an equivalence between the capacity of the Gilbert-Elliot channel and the Source and Channel model in II.1, we make the following identification. X_k is said to be zero if the channel is in the good state and X_k is said to be one if the channel is in the bad state. The Markov transition kernel and the channel transition matrix for the equivalent source and channel model is given by

$$\mathcal{K} = \begin{pmatrix} 1-b & b \\ g & 1-g \end{pmatrix}, \quad \mathcal{C}_{GE} = \begin{pmatrix} 1-P_g & P_g \\ 1-P_b & P_b \end{pmatrix}. \tag{II.10}$$

Let the output of this equivalent source and channel model be $\tilde{Z}_k$. It was shown in [MBD89] that the capacity of the Gilbert-Elliot channel is given by

$$C_{GE} = 1 - \bar{H}(\tilde{Z}). \tag{II.11}$$

Now consider a Gilbert-Elliot Channel with parameters

$$\begin{aligned} g = b &= \pi, \\ P_g = 1 - P_b &= \delta. \end{aligned} \tag{II.12}$$

Observe that this reduces to the transition kernel and the channel matrix in (II.1).

In the next section we state the main results of this paper.

III. Main Results

We obtain the following main results of the paper in the rare transitions regime by analyzing the alternative Markov process.

Theorem 3.1: As $\pi \to 0$, the minimum probability of error is given by

$$\mathcal{E}_{min} = \frac{-\pi \ln \pi}{D(\delta \| 1-\delta)}(1 + o(1)). \tag{III.1}$$

Note, $D(\delta \| 1-\delta)$ represents the binary Kullback-Liebler distance given by

$$D(\delta \| 1-\delta) = \delta \ln \frac{\delta}{1-\delta} + (1-\delta) \ln \frac{1-\delta}{\delta}.$$

Theorem 3.2: In the asymptotic regime $\pi \to 0$, the entropy rate of the hidden Markov process, $\bar{H}(Z)$, is bounded by

$$h_b(\delta) + \frac{(1-2\delta)^2}{1-\delta}\pi \ln\frac{1}{\pi} \leq \bar{H}(Z) \leq h_b(\delta) + \pi \ln\frac{1}{\pi} \quad \text{(III.2)}$$

For the Gilbert-Elliot channel with parameters as defined in (II.12), we obtain the following result as an immediate consequence of Theorem 3.2 and equation (II.11).

Corollary 3.3: For $\pi \to 0$, we have

$$1 - h_b(\delta) - \pi\ln\frac{1}{\pi} \leq C_{GE} \leq 1 - h_b(\delta) - \frac{(1-2\delta)^2}{1-\delta}\pi\ln\frac{1}{\pi}.$$

The upper bound is quite straightforward and arises from the following observation

$$\bar{H}(Z) \leq h_b(\delta) + h_b(\pi)$$

In the next section, we establish Theorems 3.1 and 3.2 by analyzing the behavior of the alternative Markov process in the rare transitions regime, i.e. $\pi \to 0$.

IV. Analysis of the alternative Markov chain

Consider the auto-regressively defined first-order Markov chain described by (II.4). In this section we will characterize the behavior of a typical sample path of this process. From this characterization we will use ergodicity to compute the various probabilities and expectations. For the convenience of the reader we reproduce (II.4) below and state some important observations,

$$Y_k = r_k \ln\frac{1-\delta}{\delta} + s_k h(Y_{k-1}), \quad \text{(IV.1)}$$

with r_k, s_k and $h(x)$ defined by (II.5) and (II.6), respectively. Let Y represent the stationary distribution of this Markov process. It was shown in [OW04] that the support of Y is given by $[-A, A]$ where

$$A = \ln\frac{(\alpha-1)(1-\pi) + \sqrt{4\alpha\pi^2 + (\alpha-1)^2(1-\pi)^2}}{2\pi} \quad \text{(IV.2)}$$

and $\alpha = \frac{1-\delta}{\delta}$. In the rare transitions regime $\pi \approx 0$, A becomes

$$\begin{aligned} A &= \ln\frac{(\alpha-1)(1-\pi)[2 + \frac{2\alpha\pi^2}{\alpha-1} + O(\pi^3)]}{2\pi} \\ &= \ln\frac{\alpha-1}{\pi}[1 + (\alpha-1)\pi + \frac{2\alpha\pi^2}{(\alpha-1)^2} + O(\pi^3)] \quad \text{(IV.3)} \\ &= \ln\frac{1}{\pi} + \ln(\alpha-1) + (\alpha-1)\pi + O(\pi^2), \end{aligned}$$

where $\alpha = \frac{1-\delta}{\delta} \geq 1$.

Remark 4.1: This can be readily seen by observing the dynamics in (IV.1) and observing that A should satisfy

$$A = h(A) + \ln\frac{1-\delta}{\delta}.$$

The next lemma states some properties of $h(x)$ that will be used for analyzing the typical sample path.

Lemma 4.2: The function $h(x) = \ln\frac{e^x(1-\pi)+\pi}{e^x\pi+1-\pi}$ satisfies:

(i) $h(x)$ and $x - h(x)$ are monotonically increasing functions
(ii) $h(x) > 0 (< 0)$ when $x > 0 (< 0)$ and $h(0) = 0$
(iii) $h(x) = -h(-x)$
(iv) $|h(x)| < |x|$ for all $x \in \mathbb{R}$
(v) If $|x| < \ln\frac{1}{\pi} - \sqrt{\ln\frac{1}{\pi}}$, then $|x - h(x)| < \frac{e^{-\sqrt{-\ln\pi}}}{1-\pi}$.

Proof: The proofs of items $(i) - (iv)$ are straightforward algebraic manipulations and is left to the reader. For part (v) of the Lemma, note the following observations: From parts $(i) - (iv)$, we know that $x - h(x)$ is an odd function and is monotonically increasing. Therefore it suffices to show that

$$x_0 - h(x_0) < \frac{e^{-\sqrt{-\ln\pi}}}{1-\pi},$$

where $x_0 = \ln\frac{1}{\pi} - \sqrt{\ln\frac{1}{\pi}}$.

Observe that $x - h(x) = \ln\frac{\pi e^x + 1 - \pi}{1-\pi+\pi e^{-x}}$ and $e^{x_0} = \frac{1}{\pi}e^{-\sqrt{-\ln\pi}}$. This implies

$$\begin{aligned} x_0 - h(x_0) &= \ln\frac{1-\pi+e^{-\sqrt{-\ln\pi}}}{1-\pi+\pi^2 e^{\sqrt{-\ln\pi}}} \\ &\leq \ln(1 + \frac{e^{-\sqrt{-\ln\pi}}}{1-\pi}) \quad \text{(IV.4)} \\ &< \frac{e^{-\sqrt{-\ln\pi}}}{1-\pi}. \end{aligned}$$

The last inequality follows from the fact that $\ln(1+x) < x$ when $x > 0$. ■

2) Outline of a typical sample path evolution: Consider a typical sample path of the Markov process in (IV.1). When π is small, s_k will almost always equal $+1$, with flips occurring roughly $\frac{1}{\pi}$ instances apart. During a long sequence when s_k is $+1$, equation (IV.1) becomes

$$Y_k = r_k \ln\frac{1-\delta}{\delta} + h(Y_{k-1}). \quad \text{(IV.5)}$$

We know that the support of Y lies in $[-A, A]$ and from (IV.3) that $A \approx -\ln\frac{1}{\pi}$. Whenever Y_k lies in $[-x_0, x_0]$, part (v) of Lemma 4.2 helps us conclude that we can approximate the evolution in equation (IV.5) by

$$Y_k = r_k \ln\frac{1-\delta}{\delta} + Y_{k-1}. \quad \text{(IV.6)}$$

This represents a random walk with a postive drift given by

$$\mathbb{E}(r_k)\ln\frac{1-\delta}{\delta} = D(\delta\|1-\delta).$$

Therefore, via usual martingale arguments, one can see that in $\frac{2x_0}{D(\delta\|1-\delta)}$ steps the walk reaches x_0 from $-x_0$. Because of a strong positive drift, the Markov process in (IV.5) tends to remain in the vicinity of x_0 and not drift downwards. Note that the time period of transition is $O(-\ln\pi)$ and is much smaller than the inter flip period of s_k, (whose expected value is $\frac{1}{\pi}$). Therefore at the occurence of the next flip $Y_k \approx -\ln\pi$ and $Y_{k+1} \approx \ln\pi$.

Again during this inter flip interval, Y_k performs the above mentioned random walk from around $\ln\pi$ to $-\ln\pi$ with a drift given by $D(\delta\|1-\delta)$. The number of steps required before this walk becomes positive is given by $\frac{-\ln\pi}{D(\delta\|1-\delta)}$. Since the flips of s_k occur at rate $\frac{1}{\pi}$, we obtain that the total fraction of time a typical sample path remains negative is given by

$$\frac{-\pi\ln\pi}{D(\delta\|1-\delta)}(1+o(1)).$$

By ergodicity, this implies that

$$\mathcal{E}_{min} = \mathbb{P}(Y<0) = \frac{-\pi\ln\pi}{D(\delta\|1-\delta)}(1+o(1)),$$

which is the statement of Theorem 3.1.

Remark 4.3: The above argument leaves out a lot of details that are required to complete the various claims in the explanation. Due to space constraints, as well as the fact that the details cloud the intuition, we omit them from this version of the paper.

To establish Theorem 3.2 we observe that in our previous analysis we showed that for a fraction of time $\approx 1-\frac{2\pi\ln\pi}{D(\delta\|1-\delta)}$ the sequence Y_k remains around $-\ln\pi$ and in the remaining fraction of time, it performs a random walk given by equation (IV.6) between $[\ln\pi, -\ln\pi]$.

This helps us break down the computation of $\mathbb{E}h_b(\frac{e^Y}{1+e^Y}*\pi*\delta)-h_b(\delta)$ into two parts. We know, using ergodicity, that for typical sample paths

$$\begin{aligned}
&\mathbb{E}h_b(\frac{e^Y}{1+e^Y}*\pi*\delta)-h_b(\delta)\\
&\quad=\lim_N\frac{1}{N}\sum_k h_b(\frac{e^{Y_k}}{1+e^{Y_k}}*\pi*\delta)-h_b(\delta)\\
&\quad=\lim_N\frac{1}{N}\sum_{k:Y_k\approx-\ln\pi}\left(h_b(\frac{e^{Y_k}}{1+e^{Y_k}}*\pi*\delta)-h_b(\delta)\right)\\
&\quad+\lim_N\frac{1}{N}\sum_{k:\ln\pi\overset{Y_k}{\rightsquigarrow}-\ln\pi}\left(h_b(\frac{e^{Y_k}}{1+e^{Y_k}}*\pi*\delta)-h_b(\delta)\right).
\end{aligned} \tag{IV.7}$$

We need to estimate the contributions of both the terms.

Let $\tilde{Y}_0\approx-\ln\pi$ denote the initial state of the Markov process in the second phase (i.e. the first time the random walk crosses x_0.) Since the jumps are in fixed amounts of $\ln\frac{1-\delta}{\delta}$, and the fact that x_0 is approximately the upper fixed point of the actual walk defined by (IV.5) helps us approximate the walk in this region by a birth-death Markov chain.

The states of this Markov chain are defined by

$$S_k = \tilde{Y}_0 - k\ln\frac{1-\delta}{\delta}, \quad k\geq 0$$

and the birth-death process is linked to the actual random walk as follows: Whenever $r_i=-1$ the Markov chain jumps from current state S_k to state S_{k+1} and when $r_i=1$ it jumps down from current state S_k to state S_{k-1}. If at current time the Markov chain is at state S_0 and $r_i=-1$ then however the chain continues to remain at state S_0.

We wish to study the birth-death process for a time equal to the inter-flip duration of the coin with bias π, the variable s_k, in (II.4). The birth death process described above starts from S_0 at time 0 and evolves as described previously.

The expected hitting time for a state j of a birth-death process conditioned on the event that the process starts at origin is given by [PT96]

$$\mathbb{E}_0 T_j = \frac{1}{\alpha^{j-1}}\frac{1-\alpha^j}{\delta(1-\alpha)^2} - \frac{j\alpha}{\delta(1-\alpha)}$$

Here, $\alpha=\frac{1-\delta}{\delta}$. Thus the hitting time is proportional to $(\frac{1-\delta}{\delta})^j$. Since the interflip duration is governed by a coin of bias π, the Markov chain will not hit any states S_k for $\frac{k}{\ln\frac{1}{\pi}} > \left(\ln\frac{1-\delta}{\delta}\right)^{-1}$ as $\pi\to 0$.

Consider the states S_k for $\frac{k}{\ln\frac{1}{\pi}} < \left(\ln\frac{1-\delta}{\delta}\right)^{-1}$ as $\pi\to 0$. Observe that the expected hitting time is of a smaller order than the interflip duration. Hence the fraction of time (during an interflip interval) that the birth death chain occupies such a state would have converged to its stationary probability measure.

Using these observations, we can lower bound the first term, H_1, by restriction the summation to the appropriate states of the birth-death approximation and substituting fraction of times in states by their stationary probabilities.

$$\begin{aligned}
H_1 &\geq \frac{1}{N}\sum_{k:Y_k\approx-\ln\pi}\left(\frac{e^{-|\hat{Y}_k|}}{1+e^{-|\hat{Y}_k|}}\right)D(\delta\|1-\delta))(1+o(1))\\
&= D(\delta\|1-\delta))\sum_{i=0}^{\ln\frac{1}{\pi}\left(\ln\frac{1-\delta}{\delta}\right)^{-1}}\pi\left(\frac{1-\delta}{\delta}\right)^i\\
&\qquad\frac{1-2\delta}{1-\delta}\left(\frac{\delta}{1-\delta}\right)^i(1+o(1))\\
&= \frac{(1-2\delta)^2}{1-\delta}\pi\ln\frac{1}{\pi}(1+o(1))
\end{aligned} \tag{IV.8}$$

The second term, H_2, comes from the random walk between $[\ln\pi,-\ln\pi]$, initiated by s_k taking the value -1. We now show that the contribution from this term is negligible compared to the first term.

Remark 4.4: We proceed to bound the second term in the following fashion. We express the summation $\sum_{k:\ln\pi\overset{Y_k}{\rightsquigarrow}-\ln\pi}\left(h_b(\frac{e^{Y_k}}{1+e^{Y_k}}*\pi*\delta)-h_b(\delta)\right)$ as the expected contribution to the sum coming from one particular flip of s_k multiplied by the total number of flips in time N. The contribution coming from one particular flip of s_k is termed $\tilde{H}_2$ and since the rate of flips is π we get that $H_2\approx\pi\mathbb{E}(\tilde{H}_2)$. Note: Though per flip of s_k, the term $\tilde{H}_2$ is a random variable, since we average over a large number of flips we can replace the average of these terms by its expected value.

Returning to bounding the second part, consider the random walk with $\tilde{Y}_0=y\approx\ln\pi$, and

$$\tilde{Y}_k = r_k\ln\frac{1-\delta}{\delta}+\tilde{Y}_{k-1}.$$

Further let $T = \inf_k \tilde{Y}_k > x_0$. Define the random variable $\tilde{H}_2^y$ as

$$\tilde{H}_2^y = \sum_{k=0}^{T} h_b(\frac{e^{\tilde{Y}_k}}{1+e^{\tilde{Y}_k}} * \pi * \delta) - h_b(\delta). \quad \text{(IV.9)}$$

We can bound the expected value of the expression in equation (IV.9) as follows. First observe that the contribution from $\tilde{Y}_k$ and $-\tilde{Y}_k$ is the same, i.e.

$$h_b(\frac{e^{\tilde{Y}_k}}{1+e^{\tilde{Y}_k}} * \pi * \delta) = h_b(\frac{e^{-\tilde{Y}_k}}{1+e^{-\tilde{Y}_k}} * \pi * \delta). \quad \text{(IV.10)}$$

Consider a new random walk $\hat{Y}_k$ with $\hat{Y}_0 = 0$ and

$$\hat{Y}_k = r_k \ln \frac{1-\delta}{\delta} + \hat{Y}_{k-1}.$$

We claim that

$$\begin{aligned} \mathbb{E}(\tilde{H}_2^y) &\leq 2\mathbb{E}(\sum_{k=0}^{\infty} h_b(\frac{e^{\hat{Y}_k}}{1+e^{\hat{Y}_k}} * \pi * \delta) - h_b(\delta)) \\ &= 2\mathbb{E}(\sum_{k=0}^{\infty} h_b(\frac{e^{-|\hat{Y}_k|}}{1+e^{-|\hat{Y}_k|}} * \pi * \delta) - h_b(\delta)). \end{aligned} \quad \text{(IV.11)}$$

The factor 2 takes care of the contribution of the walk Y_k from y to 0, as it can be considered in reverse time as a walk starting from 0 and an identical negative drift. Now observe that

$$\begin{aligned} &2\mathbb{E}(\sum_{k=0}^{\infty} h_b(\frac{e^{-|\hat{Y}_k|}}{1+e^{-|\hat{Y}_k|}} * \pi * \delta) - h_b(\delta)) \\ &\leq 2\mathbb{E}(\sum_{k=0}^{\infty} h_b(\frac{e^{-|\hat{Y}_k|}}{1+e^{-|\hat{Y}_k|}} * \delta) - h_b(\delta)) \\ &\overset{(a)}{\leq} \mathbb{E}(2\sum_{k=0}^{\infty} (\frac{e^{-|\hat{Y}_k|}}{1+e^{-|\hat{Y}_k|}} * \delta - \delta) D(\delta \| 1-\delta)) \\ &\leq \mathbb{E}(2\sum_{k=0}^{\infty} e^{-|\hat{Y}_k|}(1-2\delta) D(\delta \| 1-\delta)), \end{aligned} \quad \text{(IV.12)}$$

where (a) follows from the concavity of $h_b(x)$.

We make the following claim:

Lemma 4.5:

$$\sum_{k=0}^{\infty} \mathbb{E}(e^{-|\hat{Y}_k|}) \leq \frac{1}{1-2\sqrt{\delta(1-\delta)}}. \quad \text{(IV.13)}$$

Proof: The proof is omitted due to space constraints.

Using equations (IV.11), (IV.12), and (4.5), we obtain that

$$\mathbb{E}(\tilde{H}_2^y) \leq 2(1-2\delta) D(\delta \| 1-\delta) \frac{1}{1-2\sqrt{\delta(1-\delta)}}.$$

Since this contribution occurs at rate π corresponding to a flip in s_k, we obtain that

$$H_2 \leq \pi 2(1-2\delta) D(\delta \| 1-\delta) \frac{1}{1-2\sqrt{\delta(1-\delta)}} (1+o(1)). \quad \text{(IV.14)}$$

Hence the second term is negligible compared to the first one and combining (IV.8) and (IV.14) we obtain as $\pi \to 0$,

$$\frac{(1-2\delta)^2}{1-\delta} \pi \ln \frac{1}{\pi} \leq \bar{H}(Z) - h_b(\delta) \leq \pi \ln \frac{1}{\pi} \quad \text{(IV.15)}$$

which is the statement of Theorem 3.2.

Remark 4.6: As in the proof of Theorem 3.1, we omit the details of the justification for the various approximations. The main idea behind this justification is, however, contained in part (v) of Lemma 4.2, where we see that the approximation of $h(x)$ by x is negligible in the region of interest.

V. Conclusion

This paper presents an alternative approach for determining the optimal filtering error rate in the rare transitions regime. This approach involves the analysis of an alternative Markov process. The method also yields bounds on the entropy rate of the hidden Markov process in the rare transitions regime as well. The resulting bounds translate directly into a statement regarding the capacity of the Gilbert-Elliot channel. Though the analysis performed here has been on a simple model: a binary chain with symmetric transition probabilities, it is quite easy to carry over the intuition obtained to finite state Markov chains as well. The expressions for general finite state Markov chains can be understood via a similar analysis, and the filtering error rate can be obtained, again, essentially as the time a random walk with a drift takes to cross a certain threshhold.

VI. Acknowledgements

The authors wish to thank Or Zuk for pointing out an error in a preliminary version of the paper.

References

[CT91] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[EBTBH04] S. Egner, V. B. Balakirsky, L. M. G. .M. Tolhuizen, S. P. M. J. Baggen, and H. D. L. Hollmann, "On the Entropy Rate of a Hidden Markov Model," *Int. Symp. Inf. Th.,* p.12, Chicago, IL, June-July 2004.

[EM02] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inform. Theory*, vol. IT-48, no. 6,pp. 1518–1569, June 2002

[HGG03] T. Holliday, P. Glynn and A. Goldsmith, "Capacity of Finite State Markov Channels with General Inputs," *Int. Symp. Inf. Th.,* p. 289, Yokohama, Japan, June-July 2003.

[JSS04] P. Jacquet, G. Seroussi, and W. Spankowski. "On the Entropy of a Hidden Markov Process," *Int. Symp. Inf. Th.,* p.10, Chicago, IL, June-July 2004.

[KL92] R. Z. Khasminskii and B. V. Lazareva, "On some filtration procedure for jump Markov process observed in white Gaussian noise," *Ann. Statistics*, vol. 20, pp. 2153–2160, 1992.

[KZ96] R. Z. Khasminskii and O. Zeitouni, "Asymptotic filtering for finite state Markov chains," *Stoch. Proc. Appl.*, vol. 63, pp. 1–10, 1996.

[MBD89] M. Mushkin and I. Bar-David, "Capacity and Coding for the Gilbert-Elliot Channel," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1277–1290, November 1989

[OW04] E. Ordentlich and T. Weissman, "New bounds on the Entropy Rate of Hidden Markov Processes," *San Antonio Information Theory workshop*, October 2004.

[PT96] J. L. Palacios and P. Tetali, "A note on expected hitting times for birth and death chains,: *Statistics and Probability Letters*, vol. 30, pp. 119-125, 1996.

[W65] W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," *SIAM J. Control*, vol. 2, pp. 347–368, 1965.

[ZKD04] O. Zuk, I. Kanter, and E. Domany, "Asymptotics of the Entropy Rate for a Hidden Markov Process," Submitted to SSPIT, 2004.