Noisy Constrained Capacity

Philippe Jacquet
INRIA
Rocquencourt
78153 Le Chesnay Cedex
France
Email: philippe.jacquet@inria.fr

Gadiel Seroussi
Hewlett-Packard Laboratories
Palo Alto, CA 94304
U.S.A.
Email: gseroussi@ieee.org

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.
Email: spa@cs.purdue.edu

***Abstract*— We study the classical problem of noisy constrained capacity in the case of the binary symmetric channel (BSC), namely, the capacity of a BSC whose input is a sequence from a constrained set. As stated in [4] "... while calculation of the noise-free capacity of constrained sequences is well known, the computation of the capacity of a constraint in the presence of noise ... has been an unsolved problem in the half-century since Shannon's landmark paper ...." We express the constrained capacity of a binary symmetric channel with (d,k)-constrained input as a limit of the top Lyapunov exponents of certain matrix random processes. We compute asymptotic approximations of the noisy constrained capacity for cases where the noise parameter ε is small. In particular, we show that when $k\leq 2d$, the error term with respect to the constraint capacity is $O(\varepsilon)$, whereas it is $O(\varepsilon\log\varepsilon)$ when $k>2d$. In both cases, we compute the coefficient of the error term. We also extend previous results on the entropy of a hidden Markov process to higher-order finite memory processes.**

I. Introduction

We consider a binary symmetric channel (BSC) with crossover probability ε, and a constrained set of inputs. More precisely, let $\mathcal{S}_n$ denote the set of binary sequences of length n satisfying a given (d,k)-RLL constraint [16], i.e., no sequence in $\mathcal{S}_n$ contains a run of zeros of length shorter than d or longer than k (we assume that the values d and k, $d\leq k$, are understood from the context). Furthermore, we denote $\mathcal{S}=\bigcup_{n>0}\mathcal{S}_n$. We assume that the input to the channel is a stationary process $X=\{X_k\}_{k\geq 1}$ supported on $\mathcal{S}$. We regard the BSC channel as emitting a Bernoulli noise sequence $E=\{E_k\}_{k\geq 1}$, independent of X, with $P(E_i=1)=\varepsilon$. The channel output is

$$Z_i = X_i \oplus E_i.$$

where $\oplus$ denotes addition modulo 2 (exclusive-or).

For ease of notation, we identify the BSC channel with its parameter ε. Let $C(\varepsilon)$ denote conventional BSC channel capacity (over unconstrained binary sequences), namely, $C(\varepsilon)=1-H(\varepsilon)$, where $H(\varepsilon)=-\varepsilon\log\varepsilon-(1-\varepsilon)\log(1-\varepsilon)$.[1] The *noisy constrained capacity* $C(\mathcal{S},\varepsilon)$ is defined [4] by

$$C(\mathcal{S},\varepsilon)=\sup_{X\in\mathcal{S}} I(X;Z)=\lim_{n\to\infty}\frac{1}{n}\sup_{X_1^n\in\mathcal{S}_n} I(X_1^n,Z_1^n)\,, \quad (1)$$

where the suprema are over all stationary processes supported on $\mathcal{S}$ and $\mathcal{S}_n$, respectively. The *noiseless capacity* of the constraint is $C(\mathcal{S})\triangleq C(\mathcal{S},0)$. This quantity has been extensively studied, and several interpretations and methods for its explicit derivation are known (see, e.g., [16] and extensive bibliography therein). As for $C(\mathcal{S},\varepsilon)$, the best results in the literature have been in the form of bounds and numerical simulations based on producing random (and, hopefully, typical) channel output sequences (see, e.g., [24], [21], [1] and references therein). These methods allow for fairly precise numerical approximations of the capacity for given constraints and channel parameters.

In order to find an expression for $C(\mathcal{S},\varepsilon)$ we first consider the corresponding mutual information,

$$I(X;Z)=H(Z)-H(Z|X). \quad (2)$$

Since $H(Z|X)=H(\varepsilon)$, the problem reduces to finding $H(Z)$, the entropy rate of the output process.

For any sequence $\{x_i\}_{i\geq 1}$, we denote by x_i^j, $j\geq i$, the subsequence $x_i,x_{i+1},\ldots,x_j$. It is well known (see, e.g., [16]) that we can regard the (d,k) constraint as the output of a kth-order finite memory (Markov) stationary process, uniquely defined by conditional probabilities $P(x_t|x_{t-k}^{t-1})$. For nontrivial constraints, some of these conditional probabilities must be set to zero in order to enforce the constraint (for example, the probability of a zero after seeing k consecutive zeros, or of a one after seeing less than d consecutive zeros). When the remaining free probabilities are assigned so that the entropy of the process is maximized, we say that the process is *maxentropic*, and we denote it by $P^{\max}$. The noiseless capacity $C(\mathcal{S})$ is equal to the entropy of $P^{\max}$ [16].

If we restrict our attention to constrained processes X that are generated by Markov sources, the output process Z can be regarded as a *hidden Markov process* (HMP), and the problem of computing $I(X;Z)$ reduces to that of computing the entropy rate of this HMP.

The Shannon entropy (or, simply, *entropy*) of a HMP was studied as early as [2], where the analysis suggests the intrinsic complexity of the HMP entropy as a function of the process parameters. Blackwell [2] showed an expression of the entropy in terms of a measure Q, obtained by solving an integral equation dependent on the parameters of the process. The measure is hard to extract from the equation in any explicit way. Recently, we have seen a resurgence of interest in estimating HMP entropies [6], [7], [12], [17], [18], [25]. In particular, one

[1] We use natural logarithms throughout. Entropies are correspondingly measured in nats. The entropy of a random variable or process X will be denoted $\mathbf{H}(X)$, and the entropy rate by $H(X)$.

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 the Taylor expansion around $\varepsilon = 0$ is studied for a binary HMP of order one. In particular, the first derivative of the entropy rate at $\varepsilon = 0$ is expressed very compactly as a Kullback-Liebler divergence between two distributions on binary triplets, derived from the marginals of the input process X. It is also shown in [12] that the entropy rate of a HMP can be expressed in terms of the top Lyapunov exponent of a random process of $2{\times}2$ matrices (cf. also [9], where the capacity of certain channels with memory is also shown to be related to top Lyapunov exponents). Further improvements, and new methods for the asymptotic expansion approach were obtained in [17], [25], and [7]. In [18] the authors express the entropy rate for a binary HMP where one of the transition probabilities is equal to zero as an asymptotic expansion including a $O(\varepsilon \log \varepsilon)$ term. As we shall see in the sequel, this case is related to the $(1, \infty)$ (or the equivalent $(0, 1)$) RLL constraint. Analyticity of the entropy as a function of ε was studied in [6].

In Section II of this paper we extend the results of [12], [13] on HMP entropy to higher order Markov processes. We show that the entropy of a rth-order HMP can be expressed as the top Lyapunov exponent of a random process of matrices of dimensions $2^r \times 2^r$ (cf. Theorem 1), extending the result for $r = 1$ in [12], [13]. As an additional result of this work, of interest on its own, we derive the asymptotic expansion of the HMP entropy rate around $\varepsilon = 0$ for the case where all transition probabilities are positive (cf. Theorem 2). In particular, we derive an expression for the first derivative of the entropy rate as the Kullback-Liebler divergence between two distributions on $2r{+}1$-tuples, again generalizing the formula for $r{=}1$ [12].The results of Section II are applied, in Section III, to express the noisy constrained capacity as a limit of top Lyapunov exponents of certain matrix processes. These exponents, however, are notoriously difficult to compute [23]. Hence, as in the case of the entropy of HMPs, it is interesting to study asymptotic expansions of the noisy constrained capacity. In Section III-B, we study the asymptotics of the noisy constrained capacity, and we show that for (d, k) constraints with $k \le 2d$, we have $C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + K\,\varepsilon + O(\varepsilon^2 \log \varepsilon)$, where K is a well characterized constant. On the other hand, when $k > 2d$, we have $C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) + L\,\varepsilon \log \varepsilon + O(\varepsilon)$, where, again, L is a well-characterized constant. The latter case covers the $(0, 1)$ constraint (and also the equivalent $(1, \infty)$ constraint). Our formula for the constant L in this case is consistent with the one derived from the results of [18].

We remark that recently Han and Marcus [8] reached similar conclusions and obtained some generalizations.

II. Entropy of Higher Order HMPs

Let $X = \{X_i\}_{i \ge 1}$ be an rth-order stationary *finite memory (Markov) process* over a binary alphabet $\mathcal{A}{=}\{0, 1\}$. The process is defined by the set of conditional probabilities $P(X_t = 1 | X_{t-r}^{t-1} = a_1^r)$, $a_1^r \in A^r$. The process is equivalently interpreted as the Markov chain of its *states* $s_t = X_{t-r}^{t-1}$, $t > 0$ (we assume X_{-r+1}^0 is defined and distributed according to the stationary distribution of the process).[2] Clearly, a transition from a state $u{\in}A^r$ to a state $v{\in}A^r$ can have positive probability only if u and v satisfy $u_2^r{=}v_1^{r-1}$, in which case we say that (u, v) is an *overlapping* pair. The *noise process* $E = \{E_i\}_{i \ge 1}$ is Bernoulli (binary i.i.d.), independent of X, with $P(E_i{=}1) = \varepsilon$. Finally, the HMP is

$$Z{=}\{Z_i\}_{i \ge 1}, \quad Z_i = X_i {\oplus} E_i, \ \ i \ge 1\,. \tag{3}$$

Let $\widetilde{Z}_i = (Z_i, Z_{i+1}, \ldots, Z_{i+r-1})$ and $\widetilde{E}_i = (E_i, \ldots, E_{i+r-1})$. Also, for $e{\in}\{0, 1\}$, let $\widetilde{E}_i^e = (e, E_2, \ldots, E_{i+r-1})$. We next compute[3] $P(\widetilde{Z}_1^n)$ (equivalently, $P(Z_1^{n+r-1})$). From the definitions of X and E, we have

$$\begin{aligned} P(\widetilde{Z}_1^n, \widetilde{E}_n) &= \sum_{e \in \mathcal{A}} P(\widetilde{Z}_1^n, \widetilde{E}_n, E_{n-1} = e) \qquad (4)\\ &= \sum_{e \in \mathcal{A}} P(\widetilde{Z}_1^{n-1}, Z_{n+r-1}, E_{n-1} = e, \widetilde{E}_n)\\ &= \sum_{e \in \mathcal{A}} P(Z_{n+r-1}, E_{n+r-1} | \widetilde{Z}_1^{n-1}, \widetilde{E}_{n-1}^e) P(\widetilde{Z}_1^{n-1}, \widetilde{E}_{n-1}^e)\\ &= \sum_{e \in \mathcal{A}} P(E_{n+r-1}) P_X(\widetilde{Z}_n {\oplus} \widetilde{E}_n | \widetilde{Z}_{n-1} {\oplus} \widetilde{E}_{n-1}^e) P(\widetilde{Z}_1^{n-1}, \widetilde{E}_{n-1}^e). \end{aligned}$$

Observe that in the last line the transition probabilities $P_X(\cdot|\cdot)$ are with respect to the original Markov chain.

We next derive, from (4) , an expression for $P(\widetilde{Z}_1^n)$ as a product of matrices. In what follows, vectors are of dimension 2^r, and matrices are of dimensions $2^r \times 2^r$. We denote *row* vectors by bold lowercase letters, matrices by bold uppercase letters, and we let $\mathbf{1} = [1, \ldots, 1]$; superscript t denotes transposition. Entries in vectors and matrices are indexed by vectors in A^r, according to some fixed order, so that $A^r = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{2^r}\}$. Let

$$\mathbf{p}_n = [P(\widetilde{Z}_1^n, \widetilde{E}_n{=}\mathbf{a}_1), P(\widetilde{Z}_1^n, \widetilde{E}_n{=}\mathbf{a}_2) \ldots P(\widetilde{Z}_1^n, \widetilde{E}_n{=}\mathbf{a}_{2^r})]$$

and let $\mathbf{M}(\widetilde{Z}_n | \widetilde{Z}_{n-1})$ be a $2^r \times 2^r$ matrix defined as follows: if $(\mathbf{e}_{n-1}, \mathbf{e}_n) \in \mathcal{A}^r \times \mathcal{A}^r$ is an overlapping pair, then

$$\mathbf{M}_{\mathbf{e}_{n-1}, \mathbf{e}_n}(\widetilde{Z}_n | \widetilde{Z}_{n-1}) = P_X(\widetilde{Z}_n {\oplus} \mathbf{e}_n | \widetilde{Z}_{n-1} {\oplus} \mathbf{e}_{n-1}) P(\widetilde{E}_n{=}\mathbf{e}_n). \tag{5}$$

All other entries are zero. Clearly, $\mathbf{M}(\widetilde{Z}_n | \widetilde{Z}_{n-1})$ is a random matrix, drawn from a set of 2^{r+1} possible realizations.

With these definitions, it follows from (4) that

$$\mathbf{p}_n = \mathbf{p}_{n-1} \mathbf{M}(\widetilde{Z}_n | \widetilde{Z}_{n-1}). \tag{6}$$

Since $P(\widetilde{Z}_1^n) = \mathbf{p}_n \mathbf{1}^t = \sum_{\mathbf{e} \in A^r} P(\widetilde{Z}_1^n, \widetilde{E}_n = \mathbf{e})$, after iterating (6), we obtain

$$P(\widetilde{Z}_1^n) = \mathbf{p}_1 \mathbf{M}(\widetilde{Z}_2 | \widetilde{Z}_1) \cdots \mathbf{M}(\widetilde{Z}_n | \widetilde{Z}_{n-1}) \mathbf{1}^t. \tag{7}$$

[2] We generally use the term "finite memory process" for the first interpretation, and "Markov chain" for the second.

[3] In general, the measures governing probability expressions will be clear from the context. In cases when confusion is possible, we will explicitly indicate the measure, e.g., P_X.

The joint distribution $P(Z_1^n)$ of the HMP, presented in (7), has the form $\mathbf{p}_1\mathbf{A}_n\mathbf{1}^t$, where $\mathbf{A}_n$ is the product of the first $n-1$ random matrices of the process

$$\mathcal{M} = \mathbf{M}(\widetilde{Z}_2|\widetilde{Z}_1), \mathbf{M}(\widetilde{Z}_3|\widetilde{Z}_2), \ldots, \mathbf{M}(\widetilde{Z}_n|\widetilde{Z}_{n-1}), \ldots \quad (8)$$

Applying a subadditive ergodic theorem, and noting that $\mathbf{p}_1\mathbf{A}_n\mathbf{1}^t$ is a norm of $\mathbf{A}_n$, it is readily proved that $n^{-1}\mathbf{E}[\log P(Z_1^n)]$ must converge to a constant γ known as the *top Lyapunov exponent* of the random process $\mathcal{M}$ (cf. [5], [19], [23]). This leads to the following theorem.

Theorem 1: The entropy rate of the HMP Z of (3) satisfies

$$H(Z) = \lim_{n\to\infty} \mathbf{E}\left[-\frac{1}{n}\log P(Z_1^{n+r})\right]$$
$$= \lim_{n\to\infty}\frac{1}{n}\mathbf{E}\left[-\log\left(\mathbf{p}_1\mathbf{M}(\widetilde{Z}_2|\widetilde{Z}_1)\cdots\mathbf{M}(\widetilde{Z}_n|\widetilde{Z}_{n-1})\mathbf{1}^t\right)\right] = -\gamma,$$

where γ is the top Lyapunov exponent of the process $\mathcal{M}$ of (8).

Theorem 1 and its derivation generalize the results, for $r = 1$, of [12], [13]. It is known that computing top Lyapunov exponents is hard (maybe infeasible), as shown in [23]. Therefore, we shift our attention to asymptotic approximations.

We consider the entropy rate $H(Z)$ for the HMP Z as a function of ε for *small* ε. In order to derive expressions for the entropy rate, we resort to the following formal definition (which was also used in entropy computations in [11] and [12]):

$$R_n(s,\varepsilon) = \sum_{z_1^n\in\mathcal{A}^n} P_Z^s(z_1^n), \quad (9)$$

where s is a real (or complex) variable, and the summation is over all binary n-tuples. It is readily verified that

$$\mathbf{H}(Z_1^n) = \mathbf{E}\left[-\log P(Z_1^n)\right] = -\left.\frac{\partial}{\partial s}R_n(s,\varepsilon)\right|_{s=1}. \quad (10)$$

The entropy of the underlying Markov sequence is

$$\mathbf{H}(X_1^n) = -\left.\frac{\partial}{\partial s}R_n(s,0)\right|_{s=1}.$$

Furthermore, let $\mathbf{P} = [p_{\mathbf{e}_i,\mathbf{e}_j}]_{\mathbf{e}_i,\mathbf{e}_j\in\mathcal{A}^r}$ be the transition matrix of the underlying rth order Markov chain, and let $\boldsymbol{\pi} = [\pi_{\mathbf{e}}]_{\mathbf{e}\in\mathcal{A}^r}$ be the corresponding stationary distribution . Define also $\mathbf{P}(s) = [p^s_{\mathbf{e}_i,\mathbf{e}_j}]_{\mathbf{e}_i,\mathbf{e}_j\in\mathcal{A}^r}$ and $\boldsymbol{\pi}(s) = [\pi^s_{\mathbf{e}}]_{\mathbf{e}\in\mathcal{A}^r}$. Then

$$R_n(s,0) = \sum_{z^n} P_X^s(z_1^n) = \boldsymbol{\pi}(s)\mathbf{P}(s)^{n-1}\mathbf{1}^t. \quad (11)$$

Using a formal Taylor expansion near $\varepsilon = 0$, we write

$$R_n(s,\varepsilon) = R_n(s,0) + \varepsilon\left.\frac{\partial}{\partial\varepsilon}R_n(s,\varepsilon)\right|_{\varepsilon=0} + O(g(n)\varepsilon^2), \quad (12)$$

where $g(n)$ is the second derivative of $R_n(s,\varepsilon)$ with respect to ε, computed at some ε', provided these derivatives exist (the dependence on n stems from (9)).

Using analyticity at $\varepsilon = 0$ (cf. [6]), we find

$$\mathbf{H}(Z_1^n) = \mathbf{H}(X_1^n) - \varepsilon\left.\frac{\partial^2}{\partial s\partial\varepsilon}R_n(s,\varepsilon)\right|_{\substack{\varepsilon=0,\\ s=1}} + O(g(n)\varepsilon^2)$$
$$= \mathbf{H}(X_1^n) - \varepsilon\frac{\partial}{\partial s}\frac{\partial}{\partial\varepsilon}\left.\sum_{z_1^n} P_Z^s(z_1^n)\right|_{\substack{\varepsilon=0,\\ s=1}} + O(g(n)\varepsilon^2). \quad (13)$$

To compute the linear term in the Taylor expansion (13), we differentiate with respect to s, and evaluate at $s = 1$. Proceeding in analogy to the derivation in [12], [13], we obtain the following result.

Theorem 2: If the conditional symbol probabilities in the finite memory (Markov) process X satisfy $P(a_{r+1}|a_1^r) > 0$ for all $a_1^{r+1}\in\mathcal{A}^{r+1}$, then the entropy rate of Z for small ε is

$$H(Z) = \lim_{n\to\infty}\frac{1}{n}H_n(Z^n) = H(X) + f_1(P)\varepsilon + O(\varepsilon^2), \quad (14)$$

where, denoting by $\bar{z}_i$ the Boolean complement of z_i, and $\check{z}^{2r+1} = z_1 \ldots z_r\bar{z}_{r+1}z_{r+2}\ldots z_{2r+1}$, we have

$$f_1(P) = \sum_{z_1^{2r+1}} P_X(z_1^{2r+1})\log\frac{P_X(z_1^{2r+1})}{P_X(\check{z}_1^{2r+1})}$$
$$= \mathbb{D}\left(P_X(z_1^{2r+1})||P_X(\check{z}_1^{2r+1})\right). \quad (15)$$

Here, $\mathbb{D}(\cdot||\cdot)$ is the Kullback-Liebler divergence, applied here to distributions on $\mathcal{A}^{2r+1}$ derived from the marginals of X.□

The proof of Theorem 2, omitted here due to space limitations, generalizes and follows along the lines of [12], [13], where it was derived for the case $r = 1$.

Remark. A question arises about the asymptotic expansion of the entropy $H(Z)$ when some of the conditional probabilities are zero. Clearly, when some transition probabilities are zero, then certain sequences x_1^n are not reachable by the Markov process, which provides the link to constrained sequences. For example, consider a Markov chain with the following transition probabilities

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix} \quad (16)$$

where $0 \le p \le 1$. This process generates sequences satisfying the $(1,\infty)$ constraint (or, under a different interpretation of rows and columns, the equivalent $(0,1)$ constraint). The output sequence Z, however, will generally not satisfy the constraint. The probability of the constraint-violating sequences at the output of the channel is polynomial in ε, which will generally contribute a term $O(\varepsilon\log\varepsilon)$ to the entropy rate $H(Z)$ when ε is small. This was already observed for the transition matrix $\mathbf{P}$ of (16) in [18], where it is shown that

$$H(Z) = H(P) - \frac{p(2-p)}{1+p}\varepsilon\log\varepsilon + O(\varepsilon) \quad (17)$$

as $\varepsilon\to 0$. Recently, Han and Marcus [8] showed that in general

$$H(Z) = H(P) - f_0(P)\varepsilon\log\varepsilon + O(\varepsilon)$$

when at least one of the transition probabilities in the Markov chain is zero. If all transition probabilities are positive, then $f_0(P) = 0$ and the coefficient $f_1(P)$ at ε is then computed as in Theorem 2.

III. Capacity of the Noisy Constrained System

We now apply the results on HMPs to the problem of noisy constrained capacity.

A. Capacity in terms of Lyapunov exponents

Recall that $I(X;Z) = H(Z) - H(\varepsilon)$ and, by Theorem 1, when X is a Markov process, we have $H(Z) = \mu(P)$ where $\mu(P)$ is the top Lyapunov exponent of the process $\{\mathbf{M}(\widetilde{Z}_i|\widetilde{Z}_{i-1})\}_{i>0}$. The process optimizing the mutual information can be approached by a sequence of Markov representations $P^{(r)}$ of the constraint, of increasing order [3]. We conclude the following.

Theorem 3: The noisy constrained capacity $C(\mathcal{S},\varepsilon)$ for a (d,k) constraint through a BSC channel of parameter ε is given by

$$C(\mathcal{S},\varepsilon) = \lim_{r\to\infty} \sup_{P^{(r)}} \mu(P^{(r)}) - H(\varepsilon) \tag{18}$$

where $P^{(r)}$ denotes the probability law of an rth-order Markov process generating the (d,k) constraint $\mathcal{S}$.

In the next subsection, we turn our attention to asymptotic expansions of $C(\mathcal{S},\varepsilon)$ near $\varepsilon = 0$.

B. Asymptotic behavior

A nontrivial constraint will necessarily have some zero-valued conditional probabilities. Therefore, the associated HMP will not be covered by Theorem 2, and generally, as discussed in the remark following Theorem 2, we expect to have

$$H(Z) = H(P) - f_0(P)\varepsilon\log\varepsilon + f_1(P)\varepsilon + o(\varepsilon) \tag{19}$$

for some $f_0(P)$ and $f_1(P)$ where P is the underlying Markov process [8]. Notice that expanding around $\varepsilon = 0$ corresponds to taking the maxentropic process $P \triangleq P^{\max}$ in (19) . Therefore, after subtracting $H(\varepsilon)$, recalling that $H(\varepsilon) = -\varepsilon\log\varepsilon + \varepsilon - O(\varepsilon^2)$ for small ε, and arguing as in [8], the expansion for $C(\mathcal{S},\varepsilon)$ becomes

$$C(\mathcal{S},\varepsilon) = C(\mathcal{S}) - (1-f_0(P^{\max}))\varepsilon\log\varepsilon + (f_1(P^{\max})-1)\varepsilon + o(\varepsilon) \tag{20}$$

where $C(\mathcal{S})$ is the capacity of noiseless RLL system. Various methods exist to derive $C(\mathcal{S})$ [16]. In particular, one can write [14] $C(\mathcal{S}) = -\log\rho_0$, where ρ_0 is the smallest real root of

$$\sum_{\ell=d}^{k} \rho_0^{\ell+1} = 1. \tag{21}$$

We will show that for some RLL constraints, we have $f_0(P) = 1$ in (20), and the noisy constrained capacity is of the form $C(\mathcal{S},\varepsilon) = C(\mathcal{S}) + O(\varepsilon)$ (cf. Theorem 4 below). The first two terms of the expansion (20) were independently derived in [8], using a different methodology.

Next, we compute the leading terms of the expansion (20), leading to Theorems 4 and 5 below. Summing over the number of errors introduced by the channel, we write

$$\begin{aligned} P_Z(Z_1^n) &= P_X(X_1^n)(1-\varepsilon)^n \\ &+ \varepsilon(1-\varepsilon)^{n-1}\sum_{i=1}^{n} P_X(X_1^n\oplus e_i) + O(\varepsilon^2) \end{aligned} \tag{22}$$

where $e_j = (0,\ldots,0,1,0,\ldots,0) \in \mathcal{A}^n$ with a 1 at position j. Let $B_n \subseteq \mathcal{A}^n$ denote the set of sequence z_1^n at Hamming distance one from $\mathcal{S}_n$, and $C_n = \mathcal{A}^n \setminus (\mathcal{S}_n \cup B_n)$. Notice that sequences in C_n are at distance at least two from $\mathcal{S}_n$, and contribute to the $O(\varepsilon^2)$ term in. From (22), recalling the definition (9), we have

$$\begin{aligned} R_n(s,\varepsilon) &= \\ &\sum_{z_1^n\in\mathcal{S}_n} (1-\varepsilon)^{ns}\left(P_X(z_1^n)^s + \overbrace{\frac{\varepsilon}{1-\varepsilon}\sum_{i=1}^{n} P_X(z_1^n\oplus e_i)}^{(*)} \right)^s \\ &+ \sum_{z_1^n\in B_n\setminus\mathcal{S}_n} \varepsilon^s(1-\varepsilon)^{(n-1)s}\left(\sum_{i=1}^{n} P_X(z_1^n\oplus e_i)\right)^s + O(\varepsilon^2)\,. \end{aligned} \tag{23}$$

We now restrict our attention to the case $k \le 2d$. In this case, a one-bit flip on a (d,k) sequence is guaranteed to violate the constraint, and thus we have $\mathcal{S}_n \cap B_n = \phi$. Let $N_i(z_1^n)$ denote the number of (d,k) sequences at Hamming distance one from $z_1^n \oplus e_i$. Then, the term marked $(*)$ in (23) vanishes, and after some manipulations we obtain

$$R_n(s,\varepsilon) = R_n(s,0)(1-\varepsilon)^{ns} + \varepsilon^s(1-\varepsilon)^{(n-1)s}Q_n(s) + O(\varepsilon^{2s}), \tag{24}$$

where

$$Q_n(s) = \sum_{z_1^n\in\mathcal{S}_n}\sum_{i=1}^{n}\frac{1}{N_i(X_1^n)}\left(\sum_{j=1}^{n} P(X_1^n\oplus e_i\oplus e_j)\right)^s \tag{25}$$

Observing that $N_i(z_1^n) = N_j(z_1^n\oplus e_i\oplus e_j)$ whenever $(z_1^n, z_1^n\oplus e_i\oplus e_j)\in S_n^2$, it follows from (25) that $Q_n(1)=n$. We compute the entropy $H(Z_1^n)$ by taking the derivative of $R_n(s,\varepsilon)$ at $s=1$, which yields, after setting $Q_n(1)=n$ and some further algebraic manipulations,

$$\mathbf{H}(Z_1^n) = \mathbf{H}(X_1^n) - n\varepsilon\log\varepsilon - (Q_n'(1) + \mathbf{H}(X_1^n))\varepsilon + O(\varepsilon^2\log\varepsilon) \tag{26}$$

where $Q_n'(1)$ is the derivative of $Q_n(s)$ at $s=1$. Thus, in the case $k \le 2d$, we have $f_0(P) = 1$, and the term $O(\varepsilon\log\varepsilon)$ in (20) cancels out in this case.

Further computations are required to compute $Q_n'(1)$ and obtain the coefficient of ε in (26) . The complete derivation is presented in the full paper. Here, we provide the necessary definitions, and state the result.

Except for border effects that do not affect the asymptotics, we can represent the (d,k) sequence X_1^n as a sequence over the *extended* alphabet (of *phrases*)

$$\mathcal{B} = \{\, 0^d1, 0^{d+1}1, \ldots, 0^k1 \,\}.$$

Unconstrained sequences over $\mathcal{B}$ correspond to (d,k) sequences, and, conversely (again neglecting border effects), every (d,k) sequence can be written as a sequence over $\mathcal{B}$. Let $P_\beta^{\max}$ denote the measure induced on $\mathcal{B}$ by the maxentropic distribution $P^{\max}$, and define

$$p_\ell = P_\beta^{\max}(0^\ell 1), \quad d \le \ell \le k\,. \tag{27}$$

Note that $p_\ell = \rho_0^{\ell+1}$, with ρ_0 as in (21). The expected length of a super-symbol in $\mathcal{B}$ is

$$\lambda = \sum_{\ell=d}^{k} (\ell+1) p_\ell \,. \tag{28}$$

For integers ℓ_1, ℓ_2, $d \le \ell_1, \ell_2 \le k$, let $\mathcal{I}_{\ell_1,\ell_2}$ denote the interval

$$\mathcal{I}_{\ell_1,\ell_2} = \{\ell\colon -\min_+\{\ell_1-d, k-\ell_2-1\} \le \ell \le \min_+\{\ell_2-d, k-\ell_1-1\}\}\,,$$

where $\min_+\{a,b\} = \max\{\min\{a,b\},0\}$. Also, let $\mathcal{I}^*_{\ell_1,\ell_2} = \mathcal{I}_{\ell_1,\ell_2} \setminus \{0\}$. Now, define

$$\begin{aligned}
\tau(s) &= \sum_{\ell_1,\ell_2} (\ell_1 - d + \ell_2 - d + 1 - |I_{\ell_1,\ell_2}|) p_{\ell_1}^s p_{\ell_2}^s \\
&\quad + \sum_{\ell_1,\ell_2} \sum_{\theta \in I^*_{\ell_1,\ell_2}} \frac{1}{2} \left(p_{\ell_1} p_{\ell_2} + p_{\ell_1+\theta} p_{\ell_2-\theta}\right)^s \\
&\quad + \sum_{\ell_1,\ell_2} \frac{1}{|I_{\ell_1,\ell_2}|} \left(\sum_{\theta \in I_{\ell_1,\ell_2}} p_{\ell_1+\theta} p_{\ell_2-\theta} \right)^s,
\end{aligned}$$

and

$$\alpha(s) = \sum_{\ell=d}^{k} (2d - \ell)\, p_\ell^s.$$

The following theorem summarizes our findings for the case $k \le 2d$.

Theorem 4: Consider the constrained system $\mathcal{S}$ with $k \le 2d$. Then,

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) - (2 - \frac{\tau'(1) + \alpha'(1)}{\lambda})\varepsilon + O(\varepsilon^2 \log \varepsilon)\,.$$

Here, the derivatives of α and τ are with respect to s, evaluated at s=1. □

In the complementary case $k > 2d$, the term marked $(*)$ in (23) does not vanish, and thus the $O(\varepsilon \log \varepsilon)$ term in (20) is generally nonzero. For this case, using techniques similar to the ones leading to Theorem 4, we obtain the following result.

Theorem 5: Consider the constrained system $\mathcal{S}$ with $k \ge 2d$, let p_ℓ be as defined in (27) , define

$$\gamma = \sum_{\ell > 2d} (\ell - 2d) p_\ell \,, \qquad \delta = \sum_{\ell_1 + \ell_2 + 1 \le k} p_{\ell_1} p_{\ell_2} \,,$$

and λ as in (28) . Then,

$$C(\mathcal{S}, \varepsilon) = C(\mathcal{S}) - (1 - f_0(P^{\max}))\, \varepsilon \log \varepsilon^{-1} + O(\varepsilon)\,, \tag{29}$$

where $f_0(P^{max}) = 1 - \dfrac{\gamma + \delta}{\lambda}$. □

Example. We consider the $(1,\infty)$ constraint with transition matrix $\mathbf{P}$ as in (16) . Computing the quantities called for in Theorem 5 for $d = 1$ and $k = \infty$, we obtain $p_\ell = (1-p)^{\ell-1} p$, $\lambda = \frac{1+p}{p}$, $\gamma = \frac{(1-p)^2}{p}$, and $\delta = 1$. Thus,

$$f_0(P) = 1 - \frac{\gamma + \delta}{\lambda} = \frac{p(p-2)}{p-1}\,,$$

consistent with the calculation of the same quantity in [18]. The noisy constrained capacity is obtained when $P = P^{\max}$, i.e., $p = 1/\varphi^2$, where $\varphi = (1+\sqrt{5})/2$, the golden ratio. Then, $f_0(P^{\max}) = 1/\sqrt{5}$, and the coefficient of $\varepsilon \log(1/\varepsilon)$ in (29) is $(1/\sqrt{5}) - 1 \approx -0.553\,.$ □

ACKNOWLEDGMENT

The authors thank B. Marcus for very helpful discussions during the summer of 2006 when the first draft of this research was shaping up.

The work of W. Szpankowski was supported in part by the NSF Grants CCF-0513636 and DMS-0503742, NIH Grant R01 GM068959-01, and AFOSR Grant 073071.

REFERENCES

[1] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, W. Zeng, "Simulation-Based Computation of Information Rates for Channels With Memory," *IEEE Trans. Information Theory*, **52**, 3498–3508, 2006.

[2] D. Blackwell, "The entropy of functions of finite-state Markov chains," in *Trans. First Prague Conf. Information Theory, Statistical Decision Functions, Random Processes*, (Prague, Czechoslovakia), pp. 13–20, Pub. House Chechoslovak Acad. Sci., 1957.

[3] J. Chen and P. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *Proc. ISIT 2004*, p. 346, Chicago, 2004.

[4] J. Fan, T. L. Poo, and B. Marcus, "Constraint Gain," *IEEE Trans. Information Theory*, **50**, 1989–1999, 2001.

[5] H. Furstenberg and H. Kesten, "Products of random matrices," *Ann. Math. Statist*, pp. 457–469, 1960.

[6] G. Han and B. Marcus, "Analyticity of Entropy Rate of Hidden Markov Chains," *IEEE Trans. Information Theory*, **52**, 5251–5266, 2006.

[7] G. Han and B. Marcus, "Analyticity of Entropy Rate in Families of Hidden Markov Chains (II)," *Proc. ISIT 2006*, 103–107, Seattle, 2006.

[8] G. Han and B. Marcus, "Capacity of noisy constrained channels, *Proc. ISIT 2007*, Nice, 2007.

[9] T. Holliday, A. Goldsmith, and P. Glynn, "Capacity of Finite State Channels Based on Lyapunov Exponents of Random Matrices," *IEEE Trans. Information Theory*, **52**, 3509- 3532, 2006.

[10] K. A. Schouhammer Immink, *Codes for Mass Data Storage Systems*, Shannon Foundation Publishers, Eindhoven, 2004.

[11] P. Jacquet and W. Szpankowski, "Entropy computations via analytic depoissonization," *IEEE Trans. Inform. Theory*, **45**, 1072–1081, 1999.

[12] P. Jacquet, G. Seroussi, and W. Szpankowski, "On the Entropy of a Hidden Markov Process" (extended abstract), *Data Compression Conference*, 362–371, Snowbird, 2004.

[13] P. Jacquet, G. Seroussi, and W. Szpankowski, "On the Entropy of a Hidden Markov Process" (full version), to appear in *Theoretical Computer Science*, 2007.

[14] P. Jacquet and W. Szpankowski, "On (d,k) Sequences Not Containing a Given Word," *Proc. ISIT 2006*, Seattle, 2006.

[15] S. Karlin and H. Taylor, *A First Course in Stochastic Processes*. New York: Academic Press, 1975.

[16] B. Marcus, R. Roth and P. Siegel, "Constrained Systems and Coding for Recording Channels," Chap. 20 in *Handbook of Coding Theory* (eds. V. S. Pless and W. C. Huffman), Elsevier Science, 1998.

[17] E. Ordentlich and T. Weissman, "New Bounds on th Entropy Rate of Hidden Markov Process," *Information Theory Workshop*, 117–122, San Antonio, 2004.

[18] E. Ordentlich and T. Weissman, "On the optimality of symbol by symbol filtering and denoising," *IEEE Trans. Information Theory*, **52**, 19-40, 2006.

[19] V. Oseledec, "A multiplicative ergodic theorem," *Trudy Moskov. Mat. Obshch.*, 1968.

[20] E. Seneta, *Non-Negative Matrices*. New York: John Wiley & Sons, 1973.

[21] S. Shamai (Shitz) and Y Kofman, "On the capacity of binary and Gaussian channels with run-length limited inputs," *IEEE Trans. Commun.*, **38**, 584–594, 1990.

[22] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*. New York: John Wiley & Sons, Inc., 2001.

[23] J. Tsitsiklis and V. Blondel, "The Lyapunov exponent and joint spectral radius of pairs of matrices are hard - when not impossible – to compute and to approximate," *Mathematics of Control, Signals, and Systems*, **10**, 31–40, 1997.

[24] E. Zehavi and J. K. Wolf, "On runlength codes," *IEEE Trans. Information Theory*, **34**, 45–54, 1988.

[25] O. Zuk, I. Kanter and E. Domany, "Asymptotics of the Entropy Rate for a Hidden Markov Process," *J. Stat. Phys.*, **121**, 343–360, 2005.