International Journal of Applied Mathematics
Volume 2 No. 10 2000, 1165-1180

THE CAPACITY-COST FUNCTION OF A HÀRD-CONSTRAINED CHANNEL

Marcos Craizer[1], Artur O. Lopes[2]
[1]Departamento de Matemática - PUC - Rio
Rua Marquês de São Vicente, 225, Gávea, 22453-900
Rio de Janeiro, RJ, BRAZIL
e-mail: craizer@mat.puc-rio.br
[2]Instituto de Matemática - UFRGS, Av. Bento Gonçalves 9500
91540-000 Porto Alegre, RS, BRAZIL
e-mail: alopes@mat.ufrgs.br

Abstract: Let Σ denote the space of sequences in d symbols. A *hard-constrained channel* is determined by a subset $\Pi \subset \Sigma$ such that a message is transmitted without errors by the channel if it belongs to Π, and it is not transmitted if it does not belong to Π. The hard constrained channel is used to model the magnetic recording. Sometimes, it is interesting to add a cost for the usage of an undesirable sequence. The idea is to accept some of these sequences, but not too much of them. If a cost is assigned for each sequence, we have the *hard-constrained costly channel.*

The capacity-cost function $C(\rho)$, $\rho \in [\rho_{\min}, \rho_{\max}]$, of a channel represents the maximum code rate that we can achieve with average cost less than or equal to ρ. It is well-known that this function is continuous, strictly crescent and convex in this interval $[\rho_{\min}, \rho_{\max}]$. In this paper we show that, in the case of a hard-constrained costly channel, $C(\rho)$ is analytic in the interval $(\rho_{\min}, \rho_{\max}]$ and strictly convex in $[\rho_{\min}, \rho_{\max}]$. We prove also that $\lim_{\rho \to \rho_{\min}} \frac{dC}{d\rho}(\rho) = +\infty$ and $\frac{dC}{d\rho}(\rho_{\max}) = 0$. These are interesting theoretical results about $C(\rho)$.

The techniques we use are from the thermodynamic formalism. In our view, this is a main aspect of this paper. We show how to use deep results from thermodynamic formalism to obtain nice results in information theory. As a consequence of this approach, we can deal here with a general hard-constrained channels, not imposing the finite-state condition. Also, these techniques allow us to obtain estimates for the variations of the cost of the sequences generated by optimum codes.

AMS Subj. Classification: 94A40
Key Words: hard-constrained costly channel, capacity-cost function, thermodynamic formalism, information theory

Received: November 5, 1999

1. Introduction

Let Σ denote the space of sequences in d symbols,

$$\Sigma = \{x = (x_0, x_1, x_2, ...) | x_i \in \{1, ..., d\}\} .$$

A *channel* is a collection of rules that assigns to each *sent message* $x \in \Sigma$ the *received message* $y \in \Sigma$. The rules can be stochastic or deterministic. A *hard-constrained channel* has only deterministic rules: It transmits without errors only certain types of sequences. In other words, there exists a subset $\Pi \subset \Sigma$ such that if the sent message is in Π, the received message is exactly the same, and if the sent message is not in Π, there is no received message. If a cost is assigned for each sequence, we have the *hard-constrained costly channel*.

The hard constrained channel is used to model the magnetic recording. In this case, the constraints are due to physical limitation of the storage system. It can also be used in Shannon's telegraph channel, but in this case the constraints are dictated by the scheme of transmission used. Sometimes, it is interesting to add a cost for the usage of an undesirable sequence. The idea is to accept some of these sequences, but not too much of them. This is done by limiting the mean cost of the sequences. In [6], some examples of practical applications of these channels are presented.

A code is a scheme for the transmition of symbols through the channel, and the code rate is the average number of transmitted symbols per second. The capacity-cost function $C(\rho)$ of a channel is defined as the maximum code rate which is possible to achieve with average cost less than or equal to ρ. This function is only meaningful in a certain interval $[\rho_{\text{min}}, \rho_{\text{max}}]$ which are characterized by $\rho_{\text{min}} = \sup\{\rho | C(\rho) = 0\}$, and $\rho_{\text{max}} = \inf\{\rho | C(\rho) = C_{\text{max}}\}$. The function $C(\rho)$ has some well-known properties ([8]). It is continuous, strictly crescent and convex $\cap$ in this interval $[\rho_{\text{min}}, \rho_{\text{max}}]$. The hard-constrained costly channel has been studied by Khayrallah and Neuhoff in [6], where they show how to calculate its capacity-cost function, assuming that the channel is finite-state. They also show methods for constructing optimum codes.

In the present paper we shall prove that the capacity-cost function $C(\rho)$ of a hard-constrained costly channel is *analytic* in the semi-open interval $(\rho_{\text{min}}, \rho_{\text{max}}]$ and *strictly convex* in the closed interval $[\rho_{\text{min}}, \rho_{\text{max}}]$. Moreover, we shall also show that $\frac{dC}{d\rho}(\rho_{\text{min}}) = +\infty$ and $\frac{dC}{d\rho}(\rho_{\text{max}}) = 0$. These are interesting theoretical properties of $C(\rho)$.

The techniques we use are from the thermodynamic formalism. In our view, this is a main aspect of this paper. We show how to use deep results from thermodynamic formalism to obtain nice results in information theory. As a consequence of this approach, we can deal here with a general hard-constrained channels, not imposing the finite-state condition and also consider very general cost functions.

These techniques allow us also to show two results concerning the cost of the sequences generated by optimum codes. One of them is that the average cost satisfies assymptotically a normal law. The variance $\sigma^2(\rho)$ of this normal distribution is given by

$$\sigma^2(\rho) = -\left(\frac{d^2C}{d\rho^2}(\rho)\right)^{-1} .$$

This parameter can help the choice of a good value of ρ for a given channel. If it is important that the average cost have a small variation, then we must choose an adequate value of ρ. We also give a formula for $\sigma^2(\rho)$ and calculate it in a practical situation.

The other result concerns large deviations of the cost of sequences generated by optimum codes. We show a formula that gives the assymptotic probability that the average cost of a codified sequence differs from its limit ρ. This formula is also related with the degree of convexity of the capacity-cost function.

A related problem that seems interesting is that of a costly hard-constrained channel where we fix not only the mean, but also the variance of the cost ([5]). This situation can occur when it is important that the average cost does not oscillate too much. The techniques used in this paper can also be applied to this case.

The organization of the paper is as follows: In Section 2, we define more precisely the capacity-cost function. In Sections 3 and 4 we prove the results about $C(\rho)$ mentioned above. In Section 5, the convergence to normal is studied and in Section 6 we show the relation between the capacity-cost function and large deviations of the average cost.

2. The Capacity-Cost Function

In this section, we shall define precisely the capacity-cost function of a hard-constrained channel and relate some of its properties. Let Σ denote the space of sequences in d symbols, and $S : \Sigma \to \Sigma$ the shift,

$$S(x_0, x_1, x_2, ...) = (x_1, x_2, ...) .$$

The hard-constrained channel is given by a closed subset Π of Σ invariant by S. We shall assume that Π is irredutible and aperiodic ([2],[6]). This is a natural hypothesis, since if it does not hold we can divide Π in several channels with these properties.

The cost is a Hölder-continuous function $k : \Pi \to R^+$. This means that there exist $M > 0$ and $0 < \gamma < 1$ such that if $x = (x_0, x_1, ..)$ and $y = (y_0, y_1, ..)$ are such that $x_i = y_i$, for $0 \le i \le N$, then $|k(x) - k(y)| \le M\gamma^N$. In particular, all functions which depend only on a finite number of coordinates are Hölder-continuous. The average cost k_n is given by

$$k_n(x) = \frac{1}{n}\sum_{j=0}^{n-1} k(S^j(x)) .$$

We say that k is homologous to a constant if, for each $x \in \Sigma$,

$$\lim_{n\to\infty} k_n(x) = K ,$$

where the limit K is independent of $x \in \Sigma$.

Given $\rho \geq 0$, the capacity $C(\rho)$ is defined by

$$C(\rho) = \max_{p\in\mathcal{M}(\Pi)} \left\{ H(p) | \int k dp \leq \rho \right\} ,$$

where $\mathcal{M}(\Pi)$ denotes the set of stationary probability measures with support in Π and $H(p)$ is the entropy of this probability p ([8]). In this paper, all of the logarithms considered are natural logarithms.

Let

$$\rho_{\min} = \sup\{\rho | C(\rho) = 0\}$$

and

$$\rho_{\max} = \min\{\rho | C(\rho) = C_{\max}\} ,$$

where $C_{\max} = \max_{p\in\mathcal{M}(\Pi)}\{H(p)\}$. Evidently $C(\rho) = 0$, if $\rho < \rho_{\min}$ and $C(\rho) = C_{\max}$, if $\rho \geq \rho_{\max}$. We observe also that if k is homologous to a constant, then $\rho_{\min} = \rho_{\max}$.

In [8], it is proved that the capacity-cost function $C(\rho)$ is continuous, strictly increasing and convex $\cap$ in the interval $[\rho_{\min}, \rho_{\max}]$.

Example 1. Take Π to be the full shift Σ in 2 symbols and let $k : \Sigma \to R^+$ be given by $k(x_0, x_1, ...) = 0$ if $x_0 = 1$ and $k(x_0, x_1, ...) = 1$, if $x_0 = 2$. This channel is memoryless and so we can use i.i.d. source tests to find the capacity-cost function [8]. We have then

$$C(\rho) = \max_{p\in\mathcal{M}(\Pi)} \left\{ H(p) | \int k dp \leq \rho,\ p \text{ i.i.d.} \right\} .$$

Every i.i.d. source p is characterized by a number α in the interval $[0, 1]$ which represents $p\{x = (x_0, x_1, ...)|x_0 = 2\}$. It is clear then that $\int k dp = \alpha$ and $H(p)$ is given by $h(\alpha)$, where

$$h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha) .$$

Hence

$$C(\rho) = \max_{\alpha\in[0,1]} \{h(\alpha)|\ \alpha \leq \rho\} ,$$

and so we have $\rho_{\min} = 0$, $\rho_{\max} = \frac{1}{2}$ and for $\alpha \in [0, \frac{1}{2}]$, $C(\rho) = h(\rho)$ (Figure 3).

3. The Topological Pressure

We shall deal at the problem of maximizing the entropy by using a Lagrange multiplier μ. The corresponding problem is the to maximize the expression

$$H(p) + \mu \int_\Pi kdp ,$$

where p is any stationary probability with support in Π. It turns out that the function

$$P(\mu) = \sup_{p\in\mathcal{M}(\Pi)} \left\{H(p) + \mu \int_\Pi kdp\right\} .$$

is very well known in thermodynamic formalism and is called *topological pressure*. We shall use this fact here to derive our results.

The topological pressure has some striking properties that we list below [9]:

(1) P is an analytic function of μ.
(2) There exists a unique probability measure $p^* = p^*(\mu)$ in Π such that

$$P(\mu) = H(p^*) + \mu \int_\Pi kdp^* .$$

Moreover, the probability p^* is ergodic.

(3) The derivative of the topological pressure P is given by

$$\frac{dP}{d\mu}(\mu) = \int_\Pi kdp^* .$$

(4) The second derivative of the topological pressure P is given by

$$\frac{d^2P}{d^2\mu}(\mu) = \sigma^2(\mu) ,$$

where

$$\sigma^2(\mu) = \lim_{n\to\infty} \frac{1}{n} \int_\Pi (\sum_{j=0}^{n-1} k(S^j(x)) - n \int_\Pi kdp^*)^2 dp^* .$$

Moreover, if k is not homologous to a constant, then $\sigma^2(\mu) > 0$, for any $\mu \in R$.

From now on, we shall assume that $\rho_{\min} < \rho_{\max}$. In this case k is not homologous to a constant and hence P is a strictly convex $\cup$ function of μ.

Example 1. (Continuation) Returning to Example 1, we have that

$$P(\mu) = \sup_{\alpha\in[0,1]} \{h(\alpha) + \mu\alpha\} .$$

In order to find the maximum of the above expression we look for α satisfying

$$\frac{dh}{d\alpha}(\alpha) = -\mu .$$

Since

$$\frac{dh}{d\alpha}(\alpha) = \log(\frac{1-\alpha}{\alpha}) ,$$

we obtain

$$\alpha = \frac{1}{1+e^{-\mu}} .$$

This means that the probability p^* that maximizes the expression $H(p) + \mu\int_\Pi kdp$ is the Bernoulli proccess with $p^*\{(x_0, x_1, ..)|x_0 = 2\} = \alpha$, with α given by the above formula. Therefore

$$P(\mu) = \log(1 + e^{-\mu}) + \mu$$

(Figure 2). We obtain also expressions for

$$\frac{dP}{d\mu} = \frac{1}{1+e^{-\mu}}$$

and

$$\frac{d^2P}{d\mu^2} = \frac{e^{-\mu}}{(1+e^{-\mu})^2} .$$

In next lemma, we show that for $\mu \leq 0$, the probability measure that leads to the pressure also leads to the capacity.

Lemma 1. *Given $\mu \leq 0$, take $p^* = p^*(\mu)$ satisfying property (2) above and let $\rho = \int_\Pi kdp^*$. Then*

$$C(\rho) = H(p^*) .$$

Proof. If $p \in M(\Pi)$, $p \neq p^*$, then

$$H(p) + \mu\int_\Pi kdp < H(p^*) + \mu\int_\Pi kdp^* .$$

Since $\mu \leq 0$, if $\int_\Pi kdp \leq \rho$, then

$$H(p) < H(p^*) .$$

Therefore

$$H(p^*) = \sup_p\{H(p) | \int_\Pi kdp \leq \rho\} = C(\rho) .$$

Example 1. (Continuation) In Example 1, given $\mu \leq 0$, take $\alpha = \frac{1}{1+e^{-\mu}}$. Then $\rho = \alpha$ and by the above lemma,

$$C(\rho) = h(\rho) .$$

This is in accordance with our previous calculus of $C(\rho)$, since for $\mu \leq 0$, $\rho \in (0, \frac{1}{2}]$.

Remark 1. For each μ, we can obtain the value of $P(\mu)$ as the logarithm of the dominant eigenvalue $\lambda = \lambda(\mu)$ of a linear positive operator L in the space of continuous functions on Π . The operator L is called the Ruelle-Perron-Frobenius operator. The adjoint operator L^a defined on the space of measures on Π has also $\lambda(\mu)$ as its dominant eigenvalue. If we denote by ν the eigenvector of L^a associated to λ normalized to be a probability and by g the eigenvector of L associated to λ normalized by $\int gdv = 1$, then the probability p^* is equal to $g\nu$. For more details, see [9].

In case of a finite-state channel, the operator L reduces to a positive matrix B. And the probability p^* is determined by a Markov chain in the states (see [6]).

Example 2. Consider the hard $(1,3)$ constrained channel. It can be modeled as a 4 state channel, with transitions $1 \to 2$, $2 \to 3$, $3 \to 4$ and $2 \to 1$, $3 \to 1$, $4 \to 1$. To the first 3 transitions we associate the symbol 0 and to the other 3 we associate the symbol 1. In the soft $(1,3)$ channel, we also allow the transition $1 \to 1$ but with cost 1 (Figure 1). This channel was studied in [6] and the matrix B in this case is given by

$$B = \begin{bmatrix} e^\mu & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} .$$

In [6], the value of $\mu = -1.0885$ was chosen as an example. In this case the dominant eigenvalue is $\lambda = 1.5985$ and hence $P(1.0885) = \log(1.5985) =$

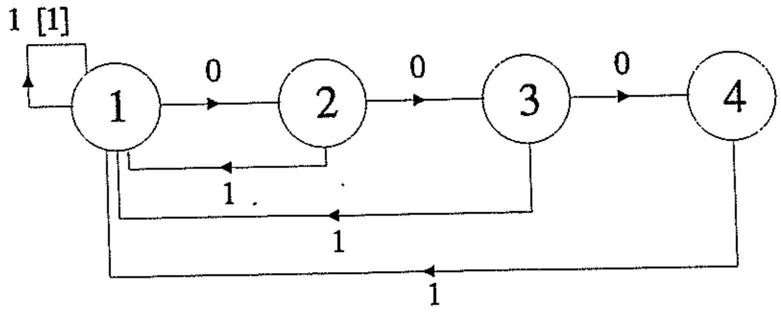

Figure 1: Graph of the (1,3) soft channel of Example 2.

0.4691. And the probability p^* is determined by a stationary Markov chain with the following distribution: The vector

$$Q = [0.4273, 0.3373, 0.1700, 0.0654]$$

gives the probabilities of the states $1, 2, 3$ and 4, respectively. And the vector

$$q = [0.2106, 0.7894, 0.4958, 0.5042, 0.6152, 0.3848, 1.0000]$$

gives the probabilities of the transitions $1 \to 1, 1 \to 2, 2 \to 1, 2 \to 3, 3 \to 1, 3 \to 4$ and $4 \to 1$, respectively. Therefore

$$\int_\Pi k dp^* = p^*(x_0 = 1, x_1 = 1) = Q_1 q_{11} = 0.0900 \, .$$

Hence we have $\rho = 0.0900$ and

$$C(0.0900) = 0.4691 + 0.0900 \times 1.0885 = 0.5671 \, .$$

4. The Legendre Transform

In this section, we show that the capacity-cost $C(\rho)$ is the Legendre transform of the topological pressure $P(\mu)$. Since we are assuming that $\rho_{\min}$ is strictly smaller than $\rho_{\max}$, we know that $P(\mu)$ is strictly convex $\cup$. Hence the limits

$$\rho_- = \lim_{\mu \to -\infty} \frac{dP}{d\mu}(\mu)$$

and

$$\rho_+ = \lim_{\mu \to +\infty} \frac{dP}{d\mu}(\mu)$$

are well defined. We denote by ρ_0 the derivative of P at $\mu = 0$.

Denote by $\mathcal{L}(P)(\rho)$ the Legendre transform of $P(\mu)$, which is defined by

$$\mathcal{L}(P)(\rho) = \inf_{\mu}\{P(\mu) - \mu\rho\} ,$$

for $\rho \in (\rho_-, \rho_+)$. Since P is strictly convex, $\mathcal{L}(P)$ is well defined and there exists a unique $\mu^* \in R$ such that

$$\mathcal{L}(P)(\rho) = P(\mu^*) - \mu^*\rho .$$

The value μ^* is the solution of the equation

$$\frac{dP}{d\mu}(\mu^*) = \rho .$$

Lemma 2. *Take $\rho \in (\rho_-, \rho_0]$. Then $\mathcal{L}(P)(\rho) = C(\rho)$.*

Proof. We have

$$\mathcal{L}(P)(\rho) = P(\mu^*) - \mu^*\rho ,$$

with $\frac{dP}{d\mu}(\mu^*) = \rho$. By properties (2) and (3) in Section 3, there exists a unique p^* such that

$$P(\mu^*) = H(p^*) + \mu^* \int kdp^* ,$$

with $\frac{dP}{d\mu}(\mu^*) = \int kdp^*$. Hence $\int kdp^* = \rho$ and therefore

$$\mathcal{L}(P)(\rho) = H(p^*) .$$

Since $\rho \in (\rho_-, \rho_0]$, $\mu^* \leq 0$. Lemma 1 implies then that

$$\mathcal{L}(P)(\rho) = C(\rho) .$$

Lemma 3. *Given $\rho \in (\rho_-, \rho_0]$, take $\mu = \mu(\rho)$ such that $\frac{dP}{d\mu}(\mu) = \rho$. Then μ is an analytic function of ρ and the following formulas are valid:*

$$C(\rho) = P(\mu) - \rho\mu ,$$

$$\frac{dC}{d\rho}(\rho) = -\mu ,$$

and

$$\frac{d^2C}{d\rho^2}(\rho) = -\left(\frac{d^2P}{d\mu^2}(\mu)\right)^{-1} .$$

Proof. Given $\rho \in (\rho_{\min}, \rho_{\max}]$, let $\mu = \mu(\rho)$ be defined by the formula

$$\frac{dP}{d\mu}(\mu) = \rho\ .$$

By the implicit function theorem, $\mu(\rho)$ is analytic and

$$\frac{d\mu}{d\rho}(\rho) = \left(\frac{d^2P}{d\mu^2}(\mu)\right)^{-1}\ .$$

And since

$$C(\rho) = P(\mu(\rho)) - \rho\mu(\rho)\ ,$$

we have

$$\frac{dC}{d\rho}(\rho) = \frac{dP}{d\mu}(\mu)\frac{d\mu}{d\rho}(\rho) - \mu(\rho) - \rho\frac{d\mu}{d\rho}(\rho) = -\mu(\rho)\ .$$

Therefore

$$\frac{d^2C}{d\rho^2}(\rho) = -\frac{d\mu}{d\rho}(\rho) = -\left(\frac{d^2P}{d\mu^2}(\mu)\right)^{-1}\ .$$

Lemma 4. *We have that* $\rho_- = \rho_{\min}$ *and* $\rho_0 = \rho_{\max}$.

Proof. By Lemma 3, $C(\rho)$ is strictly convex $\cap$ in the interval $(\rho_-, \rho_0]$. Hence $C(\rho)$ is also strictly positive in this interval . This implies that $\rho_{\min} \leq \rho_-$. Also Lemma 3 implies that $\lim_{\rho\to\rho_-} \frac{dC}{d\rho}(\rho) = +\infty$. The convexity $\cap$ of $C(\rho)$ implies then that in fact $\rho_- = \rho_{\min}$.

At $\mu = 0$, $C(\rho_0) = P(0) = C_{\max}$ and therefore $\rho_0 \geq \rho_{\max}$. By Lemma 3, if $\rho < \rho_0$ then $\frac{dC}{d\rho}(\rho) = -\mu > 0$. This implies that in fact $\rho_0 = \rho_{\max}$.

Theorem 5. *The capacity-cost function $C(\rho)$ is analytic in the interval $(\rho_{\min}, \rho_{\max}]$ and strictly convex $\cap$ in the interval $[\rho_{\min}, \rho_{\max}]$. Moreover,*

$$\frac{dC}{d\rho}(\rho_{\max}) = 0$$

and

$$\lim_{\rho\searrow\rho_{\min}} \frac{dC}{d\rho}(\rho) = +\infty\ .$$

Proof. Since $C(\rho) = P(\mu(\rho)) - \rho\mu(\rho)$, with $\mu(\rho)$ analytic, $C(\rho)$ is also analytic. And the formula

$$\frac{d^2C}{d\rho^2}(\rho) = -\left(\frac{d^2P}{d\mu^2}(\mu)\right)^{-1}$$

implies that $\frac{d^2C}{d\rho^2}(\rho) < 0$. Therefore $C(\rho)$ is convex $\cap$ in the interval $(\rho_{\min}, \rho_{\max}]$. By the continuity of $C(\rho)$ in $[\rho_{\min}, \rho_{\max}]$, we conclude that in fact $C(\rho)$ is convex $\cap$ in the interval $[\rho_{\min}, \rho_{\max}]$. Finally, the equation $\frac{dC}{d\rho}(\rho) = -\mu(\rho)$ implies that $\frac{dC}{d\rho}(\rho_{\max}) = 0$ and $\lim_{\rho \searrow \rho_{\min}} \frac{dC}{d\rho}(\rho) = +\infty$.

Example 1. (Continuation) In Example 1 we have $\rho_{\min} = \rho_- = 0$ and $\rho_{\max} = \rho_0 = \frac{1}{2}$. Also, $\rho_+ = 1$ (Figure 3).

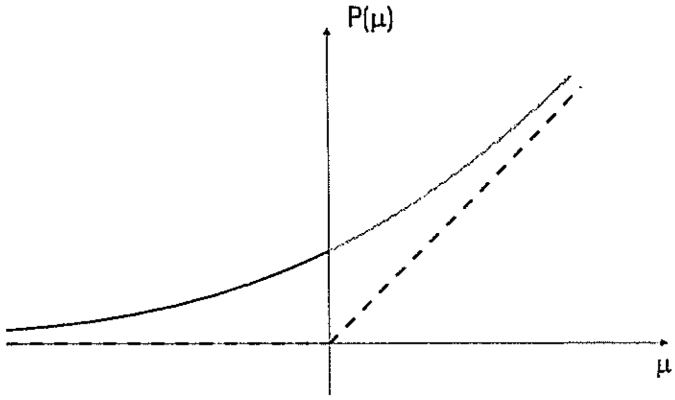

Figure 2: The topological pressure $P(\mu)$ of Example 1.

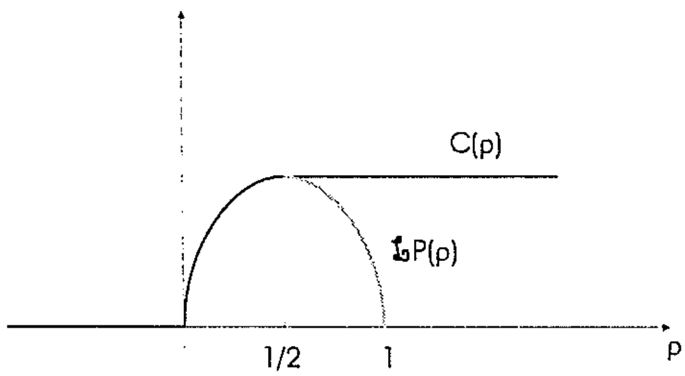

Figure 3: The function $C(\rho)$ and the Legendre transform of $P(\mu)$ of Example 1.

Example 2. (Continuation) By making $\mu = -\infty$, we obtain $\rho_{\min} = 0$ and $C(0) = \log(1.4656)$. And by making $\mu = 0$, we obtain $\rho_{\max} = 0.2938$ and $C_{\max} = \log(1.9276)$. The graph of $C(\rho)$ can be found in [6].

5. Central Limit Theorem

Given $\rho \in [\rho_{\min}, \rho_{\max}]$, there exist block codes with rates arbitrarily close to $C(\rho)$ and whose average cost is less or equal to ρ. In [4], it is shown how to construct these codes. If we assume that the source is generating symbols independently and with equal probability, then the codified sequences in Π will be distributed according to p^*. We observe that the distribution p^* in general has memory, that is, the stationary stochastic process $\{k(S^n x), p^*\}$ is not independent. In this section we show that the average cost of this process satisfies a normal law.

The cost k have a mean ρ and the assymptotic variation is given by ([9])

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \int_\Pi (\sum_{j=0}^{n-1} k(S^j(x)) - n\rho)^2 dp^* .$$

This formula can be simplified to ([9])

$$\begin{aligned}\sigma^2 &= \int_\Pi (k(x) - \rho)^2 dp^* \\ &\quad + 2\sum_{i=1}^{\infty} \int_\Pi (k(x) - \rho)(k(S^i(x)) - \rho) dp^* .\end{aligned}$$

This variance corresponds exactly to the second derivative of $P(\mu)$ at $\mu^* = -\frac{dC}{d\rho}(\rho)$. In case $\rho_{\min} < \rho_{\max}$, then it is strictly positive. We remember also that, by Lemma 4,

$$\sigma^2 = \sigma^2(\rho) = -\left(\frac{d^2C}{d\rho^2}(\rho)\right)^{-1} .$$

This fact shows us that the variance of the cost at ρ determines the degree of convexity of $C(\rho)$ at this point.

In [9], it is proved that

$$\frac{1}{\sqrt{n}} \left[\sum_{j=0}^{n-1} k(S^j(x)) - n\rho\right]$$

converges in distribution to the normal $\mathcal{N}(0,\sigma)$ with mean 0 and variance σ^2.

Example 1. (Continuation) In Example 1, it is clear that, for any $i \geq 1$, $k(S^i(x))$ is independent of $k(x)$. Therefore,

$$\int_\Pi (k(x) - \rho)(k(S^i(x)) - \rho) dp^* = 0$$

and

$$\int_{\Pi} (k(x)-\rho)^2 dp^* = \alpha(1-\rho)^2 + (1-\alpha)(0-\rho)^2 .$$

Since $\alpha = \rho^*$, we have

$$\sigma^2 = \rho(1-\rho) .$$

Example 2. (Continuation) We return to the $(1,3)$ soft channel to calculate the value of $\sigma^2 = \sigma^2(\rho)$ for $\rho = 0.0900$. In a Markov chain, given the symbol in coordinate 2, the ocurrence of symbols in coordinates 0 and 1 are independent of the ocurrence of symbols in coordinates 3, 4, ... Therefore

$$\int_{\Pi} (k(x)-\rho)(k(S^i x)-\rho) dp^* = 0 ,$$

for any $i \geq 3$. So we have to calculate this correlation only for $i = 0, 1$ and 2.

Observe first that

$$p^*((x_0, x_1) = (1,1)) = Q_1 q_{11} = 0.0900$$

and hence

$$\begin{aligned} \int_{\Pi} (k(x)-\rho)^2 dp^* &= \\ Q_1 q_{11}(1-\rho)^2 + (1-Q_1 q_{11})(0-\rho)^2 &= 0.0819 . \end{aligned}$$

Also

$$\begin{aligned} p^*((x_0,x_1) &= (1,1),(x_1,x_2) = (1,1)) \\ &= Q_1 q_{11} q_{11} = 0.0190 , \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &= (1,1),(x_1,x_2) \neq (1,1)) \\ &= Q_1 q_{11} q_{12} = 0.0710 , \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &\neq (1,1),(x_1,x_2) = (1,1)) \\ &= \sum_{i=2}^{4} Q_i q_{i1} q_{11} = 0.0710 , \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &\neq (1,1),(x_1,x_2) \neq (1,1)) \\ &= 0.9390 , \end{aligned}$$

and therefore

$$\int_{\Pi} (k(x)-\rho)(k(Sx)-\rho) dp^* = 0.0117 .$$

Finally

$$\begin{aligned} p^*((x_0,x_1) &= (1,1),(x_2,x_3)=(1,1)) \\ &= Q_1q_{11}q_{11}q_{11} = 0.0040\,, \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &= (1,1),(x_2,x_3)\neq(1,1)) \\ &= Q_1q_{11}(1-q_{11}q_{11}) = 0.0860\,, \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &\neq (1,1),(x_2,x_3)=(1,1)) \\ &= \sum_{i=2}^{4} Q_iq_{i1}q_{11}(1+q_{11}) = 0.0860\,, \end{aligned}$$

$$\begin{aligned} p^*((x_0,x_1) &\neq (1,1),(x_2,x_3)\neq(1,1)) \\ &= 0.8240\,, \end{aligned}$$

and hence

$$\int_\Pi (k(x)-\rho)(k(S^2x)-\rho)dp^* = -0.0041\,.$$

Summing these terms we obtain

$$\sigma^2 = 0.0895\,.$$

6. Large Deviations

As in last section, assume that the source is generating symbols independently and with equal probability. If we consider an optimum code with average cost ρ, the codified sequences in Π will be distributed according to p^*. Since p^* is ergodic, we know that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}k(T^j(x)) = \int_\Pi kdp^* = \rho\,.$$

In this section, we shall estimate the probability of deviations of $\frac{1}{n}\sum_{j=0}^{n-1}k(T^j(x)$ from its limit ρ. We call this deviation a large deviation ([2],[3],[4],[7]). Let

$$p_n^*(A) = p^*(\frac{1}{n}\sum_{j=0}^{n-1}k(T^j(x))\in A)\,.$$

It is clear that

$$\lim_{n\to\infty} p_n^*(A) = 1 \text{ if } \rho \in A$$

and

$$\lim_{n\to\infty} p_n^*(A) = 0 \text{ if } \rho \notin A \, .$$

In this section we shall study the velocity of convergence of $p_n^*(A)$ to 0, in the case where $\rho \notin A$.

Define

$$I(\rho, \delta) = \lim_{\varepsilon\to 0}\lim_{n\to\infty} -\frac{1}{n}\log p_n^*((\delta-\varepsilon, \delta+\varepsilon)) \, .$$

In order to simplify the notation, we shall drop the symbol ρ and denote $I(\rho,\delta)$ simply by $I(\delta)$. The following facts are true [7]:

(1) For any interval $A \subset [\rho_{\min}, \rho_{\max}]$,

$$\lim_{n\to\infty} -\frac{1}{n}\log p_n^*(A) = \inf_{\delta\in A}(I(\delta)) \, .$$

(2) We can calculate $I(\delta)$ by the formula

$$I(\delta) = -\mathcal{L}\{P(\mu+\mu^*) - P(\mu^*)\} \, ,$$

where $\mu^* = -\frac{dC}{d\rho}(\rho)$ is fixed and the Legendre transform is calculated with respect to the variable μ.

Lemma 6. *For any $\delta \in [\rho_{\min}, \rho_{\max}]$,*

$$I(\delta) = C(\rho) + (\delta-\rho)\frac{dC}{d\rho}(\rho) - C(\delta) \, .$$

Proof. We know by (2) above that

$$I(\delta) = -\inf_{\mu}\{P(\mu+\mu^*) - P(\mu^*) - \mu\delta\} \, .$$

So

$$I(\delta) = -\inf_{\mu}\{P(\mu+\mu^*) - (\mu+\mu^*)\delta\} + P(\mu^*) - \mu^*\delta \, .$$

The first parcel in the second member corresponds to the Legendre tranform of P, which is equal to the capacity. Hence

$$I(\delta) = -C(\delta) + \{P(\mu^*) - \mu^*\rho\} - (\delta-\rho)\mu^* \, .$$

Now the second parcel corresponds to $C(\rho)$. Hence

$$I(\delta) = -C(\delta) + C(\rho) - (\delta-\rho)\mu^* \, .$$

And since $\mu^* = -\frac{dC}{d\rho}(\rho)$, we conclude that

$$I(\delta) = C(\rho) + (\delta - \rho)\frac{dC}{d\rho}(\rho) - C(\delta) .$$

We observe from the last lemma that the deviate function $I(\delta)$ is obtained from the capacity-cost function in a very simple way. Observe also that $I(\rho) = \frac{dI}{d\rho}(\rho) = 0$ and $\frac{d^2I}{d\rho^2}(\rho) = -\frac{d^2C}{d\rho^2}(\rho)$. Therefore the deviation function near ρ is directly related to the degree of convexity of $C(\rho)$. If $C(\rho)$ is more convex $\cap$ at ρ, the deviation function becomes greater, which implies less deviations. If it is less convex $\cap$, the deviate function becomes smaller, which implies more deviations. This is in accordance with Section 5.

References

[1] R.L. Adler, D. Coppersmith, M. Hassner, Algorithms for sliding block codes, *IEEE Transactions on Information Theory*, **29**, No.1 (1983).

[2] P.H. Algoet, B.H. Marcus, Large deviation theorems for empirical types of Markov chains constrained to thin sets, *IEEE Transactions on Information Theory*, **38**, No. 4 (1992).

[3] P. Billingsley, *Probability and Measure*, John Wiley (1979).

[4] K. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer-Verlag (1985).

[5] J. Justesen, T. Hoholdt, Maxentropic Markov chains, *IEEE Transactions on Information Theory*, **30**, No. 4 (1984).

[6] A.S. Khayrallah, D.L. Neuhoff, Coding for channels with cost constraints, *IEEE Transactions on Information Theory*, **42**, No.3 (1996).

[7] A. Lopes, Entropy, Pressure and large deviations, *Cellular Autamata, Dynamical Systems and Neural Networks*, (E. Goles and S. Martinez, Eds.), Kluwer (1994), 79-146.

[8] R.J. McEliece, *The Theory of Information and Coding*, Addison-Wesley (1977).

[9] W. Parry, M. Pollicott, *Zeta Functions and the Periodic Structure of Hyperbolic Dynamics*, Asterisque 187-188 (1990).