

9505

CAPACITY OF NOISY,
DISCRETE MEMORYLESS CHANNELS
UNDER INPUT CONSTRAINTS

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS
ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

T. C.
Yükseköğretim Kurulu
Dokümantasyon Merkezi

By
Ezhan Karahan
June 1990

I certify that I have read this thesis and that, in my opinion, it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Associate Prof. Dr. Erdal Arıkan (Principal Advisor)

I certify that I have read this thesis and that, in my opinion, it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Associate Prof. Dr. Melek D. Yücel

I certify that I have read this thesis and that, in my opinion, it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.



Assistant Prof. Dr. Gürhan Şaplakoğlu

Approved for the Institute of Engineering and Sciences:



Prof. Dr. Mehmet Baray
Director of Institute of Engineering and Sciences



To Mom and Dad

ABSTRACT

In this thesis work, we examine the capacity of discrete memoryless channels under input constraints. We consider a certain class of input-restricted channels for which constrained sequences can be modeled as outputs of a finite-state machine(FSM). No efficient algorithm is known for computing the capacity of such a channel. For the noiseless case, i.e., when the channel input letter and the corresponding output letter are identical, it is shown that [1] the channel capacity is the logarithm of the largest eigenvalue of the adjacency matrix of the state-transition diagram of the FSM generating the allowed channel input sequences. Furthermore, the probability distribution on the input sequences achieving the channel capacity is first-order markovian.

Here, we discuss the noisy case. For a specific input-restricted channel, we show that unlike the noiseless case, the capacity is no longer achieved by a first-order distribution. We derive upper and lower bounds on the maximum rate achievable by a K -th order markovian distribution on the allowed input sequences. The computational results show that the second-order distribution does strictly better than the first-order distribution for this particular channel.

A sequence of upper bounds on the capacity of an input-restricted channel is also given. We show that this sequence converges to the channel capacity. The numerical results clarify that markovian distribution may achieve rates close to the capacity for the channel considered in this work.

Keywords: Channel capacity, Input-restricted channel, Finite-state machine, Markovian distribution.

ÖZET

Bu tez çalışmasında girdi kısıtlamaları altında ayırtık, hafızasız kanalların sığaları incelenmektedir. Burada tartışılan girdi kısıtlamalı kanallar için izin verilen girdi dizileri bir sonlu durumlu makinenin çıktıları olarak modellenenebilir. Böyle bir kanalın sığasını hesaplayabilmek için verimli bir algoritma bilinmemektedir. Gürültüsüz durumda (kanal girdi harfiyle karşılık gelen çıktı harfi aynı olduğu zaman), Shannon[1] kanal sığasının, kısıtlandırılmış kanal girdi dizilerini üreten sonlu durumlu makinenin bitişiklik matrisinin en büyük özdeğerinin logaritmasına eşit olduğunu göstermiştir. Ayrıca, kanal sığasına ulaşan girdi dizileri üzerindeki olasılık dağılımı birinci dereceden bir markov dağılımıdır.

Bu çalışmada biz gürültülü durumu tartışıyoruz. Belirli bir girdi kısıtlamalı kanal için, gürültüsüz durumdan farklı olarak, sığanın birinci dereceden bir markov dağılımı tarafından ulaşılamadığı gösterilmektedir. İzin verilen girdi dizileri üzerinde K 'nıncı dereceden bir markov dağılımının ulaşabileceği en yüksek hız üzerine alt ve üst sınırlar elde edilmektedir. Hesaplamalar sonucunda ikinci dereceden bir dağılımın birinci dereceden bir dağılıma göre daha yüksek hızlara ulaştığı görülmüştür.

Bu çalışmada ayrıca girdi kısıtlamalı kanalların sığaları üzerine bir üst sınırlar dizisi verilmektedir. Bu dizinin kanal sığasına yakınsadığı gösterilmektedir. Hesaplamalar sonucunda bu çalışmada kullanılan kanallar için markov dağılımlarının kanal sığasına oldukça yakın hızlara ulaştıkları gözlenmektedir.

Anahtar sözcükler: Kanal sığası, Girdi kısıtlamalı kanal, Sonlu durumlu makine, Markov dağılımı.

ACKNOWLEDGEMENT

I would like to express my gratitudes to Dr. Erdal Arıkan, my supervisor, for his invaluable guidance and suggestions throughout this work. I also thank to N. Cem Oğuz and M. Şenol Toygar for their helps in drawing the figures in the thesis. Special thanks are due to Oya Ekin for her moral support and help in typewriting the text.



Contents

1	INTRODUCTION	1
1.1	Examples of Input-Restricted Channels	2
1.2	Capacity of Input-Restricted DMC's	3
1.3	Computation of Channel Capacity	4
1.4	Capacity of Input-Restricted Channels:The Noiseless Case	5
1.5	Probability Assignments on Constrained Sequences and Summary of Results	5
2	CAPACITY OF NOISY, INPUT CONSTRAINED CHANNELS	8
2.1	Upper and Lower Bounds on the Maximum Rate Achievable by Markovian Distribution	8
2.2	Computational Results	11
2.3	Upper bounds on C	13
2.4	Rate of Convergence of C_N to the Capacity	15
3	CONCLUSIONS	20
A	Proof that the limit in (1.1) exists	21
B	List of Upper Bounds on the Capacity	23
	REFERENCES	26

List of Figures

1.1	State-transition diagram for an input constraint	1
1.2	State transition diagram for RLL(d,k) code	3
1.3	State transition diagram for charge constraint	3
1.4	Double-step transition diagram for the channel constraint in Figure 1.1 . . .	6
2.1	C_N vs. N for constrained BSC with $\epsilon=0.01$	17
2.2	C_N vs. N for constrained BSC with $\epsilon=0.05$	17
2.3	C_N vs. N for constrained BSC with $\epsilon=0.1$	18
2.4	C_N vs. N for constrained BEC with $\epsilon=0.1$	18
2.5	C_N vs. N for constrained BEC with $\epsilon=0.2$	19
2.6	C_N vs. N for constrained BEC with $\epsilon=0.3$	19

List of Tables

2.1	Computed bounds for BSC	13
2.2	Computed bounds for BEC	13
2.3	Upper bounds computed by Arimoto-Blahut algorithm(BSC)	14
2.4	Upper bounds computed by Arimoto-Blahut algorithm(BEC)	14
B.1	Upper bounds computed by Arimoto-Blahut algorithm(BSC)	23
B.2	Upper bounds computed by Arimoto-Blahut algorithm(BSC)	24
B.3	Upper bounds computed by Arimoto-Blahut algorithm(BSC)	24
B.4	Upper bounds computed by Arimoto-Blahut algorithm(BEC)	25

Chapter 1

INTRODUCTION

In this thesis work, we examine the capacity of discrete channels for which input sequences are restricted to satisfy certain channel input constraints. To introduce the problem, consider a binary symmetric channel (BSC) with the restriction that input sequences cannot contain consecutive 1's. For example, the sequence $(1,0,0,0,1,0,1,0,0,1,0)$ is an allowed input sequence, whereas $(1,0,1,1,0,0)$ is not. These constraints can be presented by the finite-state machine (FSM) in Fig.1.1. The binary numbers assigned to each state transition corresponds to the output of the FSM produced on that transition. It is readily seen that no sequence of transitions will generate a sequence containing consecutive 1's.

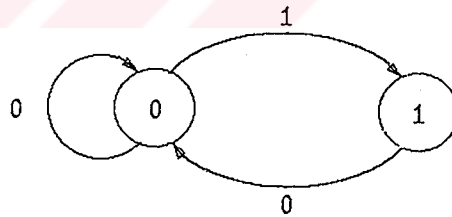


Figure 1.1: State-transition diagram for an input constraint

There is no known practical algorithm for computing the capacity of an input-restricted channel. Even for the simple channel with an input constraint represented by the FSM in Fig.1.1, the capacity is unknown. The difficulty is caused by the input constraints on the one hand and the noisy nature of the channel on the other. Although the channel is memoryless, each letter at the channel output depends on the corresponding input and on the past input and output letters. An efficient way for computing the capacity is presently unknown.

In this thesis work, we investigate maximum rates achievable by markovian channel

input distributions. We show that, unlike the noiseless case for input-restricted channels, improvements in the achievable rate are obtained by increasing the memory of markov distribution. We present numerical results for the channel in Fig.1.1 showing this improvement. We also derive upper bounds on the channel capacity to see how close one can get to capacity by using markovian channel input distributions.

In the rest of this introductory chapter, we define the above concepts more precisely, state the problem and give a summary of results. In section 1.1, we give examples of channel constraints that find various applications in practice. In section 1.2, we define the capacity of input-restricted, discrete, memoryless channels and give a brief review of channel coding theorem. Computation of the capacity of a discrete constrained channel is discussed in section 1.3. In section 1.4, we summarize the earlier work of Shannon [1] on the capacity of discrete, input-restricted, noiseless channels and on the input distribution achieving capacity. In the final section, we introduce markovian distributions on the allowed input sequences and the main contribution of this thesis is stated.

1.1 Examples of Input-Restricted Channels

In this section, we give two examples of input-restricted channels and explain the state diagram presentation of such channels. Both examples presented here belong to the class of input-restricted channels for which the allowed input sequences can be modeled as outputs of a FSM, with a set of states $S = \{S_1, S_2, \dots, S_J\}$ and at each state transition a letter from a discrete alphabet A is produced. Throughout this work, we consider input-restricted channels where constrained sequences are generated by FSM's.

Input-restricted channels play an important role in many digital transmission and recording systems, such as magnetic recording and optical transmission and recording. Recording codes based on runlength-limited sequences are widely used in optical and magnetic disk recording applications. In digital magnetic recording systems, encoding schemes (such as NRZI) encode a 1 by a change in the magnetization polarity and a 0 by no change in polarity. Systems using self-clocking for clock regeneration require frequent changes in polarity. Long sequences of like polarity may result in loss of synchronization. Therefore, the encoded sequences are restricted by design to have a maximum runlength $(k+1)$, i.e., the number of consecutive 0's between any two 1's does not exceed k . A second type of constraint is imposed by the intersymbol interference considerations. In order to reduce intersymbol interference, a constraint is placed on the minimum number of 0's between neighboring 1's which is denoted by d ; $(d+1)$ is called the minimum runlength of the code. The resulting code with parameters (d,k) is called a runlength limited (RLL) code. RLL(d,k) sequences may be thought as being generated by a finite-state machine with state transition diagram given in Fig. 1.2.

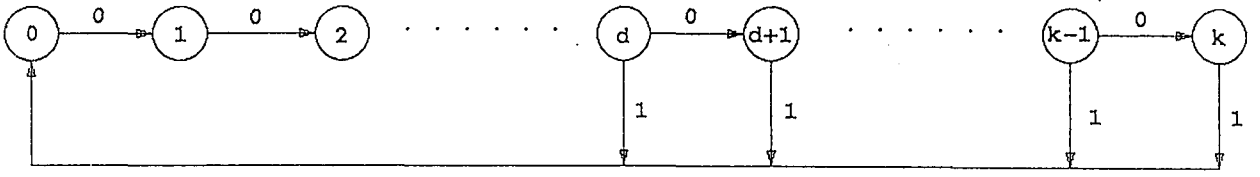


Figure 1.2: State transition diagram for RLL(d,k) code

The input-restricted channel in Fig. 1.1 corresponds to a runlength constraint $(1, \infty)$ on the input sequences.

Another kind of restriction on binary sequences is the charge constraint. Charge constrained sequences with maximum charge C are generated by an FSM with $(2C+1)$ states as shown in Fig. 1.3.

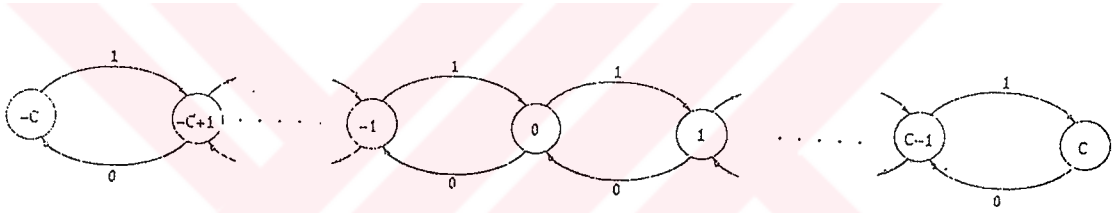


Figure 1.3: State transition diagram for charge constraint

1.2 Capacity of Input-Restricted DMC's

Consider a discrete memoryless channel (DMC), with input alphabet A , output alphabet B and channel transition probabilities $P(y|x)$, $y \in B$, $x \in A$. Let A_N be the set of channel input sequences $X^N = (x_1, \dots, x_N)$, $x_i \in A$, $i = 1, \dots, N$, of length N that satisfy the channel input constraints, starting from an arbitrary state (initial state not fixed).

The capacity C , of such a channel is defined by:

$$C \triangleq \lim_{N \rightarrow \infty} \sup_{P_N(X^N)} \frac{I(X^N; Y^N)}{N} \quad (1.1)$$

where $P_N(\cdot)$ is a probability distribution on A_N , $I(X^N; Y^N)$ is the average mutual information between channel input sequence X^N and channel output sequence Y^N :

$$I(X^N; Y^N) = \sum_{X^N, Y^N} P_N(Y^N|X^N) P_N(X^N) \log \frac{P_N(Y^N|X^N)}{\sum_{X^{N'}} P_N(Y^N|X^{N'}) P_N(X^{N'})}$$

and

$$P_N(Y^N|X^N) = \prod_{i=1}^N P(y_i|x_i)$$

since the channel is memoryless. The proof of the existence of the limit in (1.1) is given in Appendix A.

The channel capacity is a quantity of information theoretical significance [1]: At rates below it, we can communicate reliably through the channel by choosing proper encoding and decoding strategies (the term reliably means that, the probability of error averaged over the ensemble of codes can be made arbitrarily close to zero). Conversely, for rates exceeding the capacity, the probability of error for any encoder-decoder pair is bounded away from zero.

Determining channel capacity requires finding the probability distribution that maximizes the average mutual information. Knowledge of this probability distribution is also valuable in the construction of actual encoders and decoders.

1.3 Computation of Channel Capacity

The computation of channel capacity is a maximization of a nonlinear function over many variables with equality and inequality constraints. Fortunately, this problem is greatly simplified by the convexity of the average mutual information function, and there exists an algorithm for finding the capacity of a DMC (without input constraints) devised by Arimoto [4] and Blahut [5]. Starting from an arbitrary input distribution, the algorithm converges to the capacity.

When constraints on channel input sequences exist, making use of the Arimoto-Blahut algorithm we can compute

$$C_N = \frac{1}{N} \max_{P_N(X^N)} I(X^N; Y^N)$$

for any fixed $N \geq 1$, by considering the extended channel with input alphabet A_N and channel transition probabilities $P_N(Y^N|X^N)$.

However, we have no efficient way of computing the capacity $C = \lim_{N \rightarrow \infty} C_N$. Though it is possible in principle to approximate C as closely as desired by computing C_N for N sufficiently large, this quickly becomes impractical since the computation complexity grows exponentially in N . If we stop at a certain N , we have an upperbound on C , given by $C \leq C_N$ as stated in Appendix A.

1.4 Capacity of Input-Restricted Channels: The Noiseless Case

Suppose that we transmit the constrained sequences through a noiseless channel, i.e., the channel input letter and the corresponding output letter are identical. In this case, the probability distribution on the constrained channel input sequences achieving the capacity is first-order markovian, as shown by Shannon [1, Theorem 8]. That is, no improvement in maximum rate is possible by considering higher order distributions. Also, a formula is given in [1] for computing the capacity in the noiseless case:

$$C = \log W$$

where W is the largest eigenvalue of the adjacency matrix of the state transition diagram of the FSM [1, Theorem 1]. In fact, Shannon's result covers the case where letters may have different time durations.

In this work, we are concerned with the noisy, input-restricted DMC's and we ask whether the capacity is still achieved by a first-order markovian distribution on the constrained input sequences. Before discussing this problem, we will consider markovian distributions on the constrained input sequences that are generated by a FSM.

1.5 Probability Assignments on Constrained Sequences and Summary of Results

In section 1.1, examples of constraints where allowed channel input sequences can be represented as outputs of a FSM were presented. Suppose we have a FSM with a set of states $S = \{S_1, S_2, \dots, S_J\}$ and at each time instant, a state transition occurs and a letter from a discrete alphabet A is produced.

The probability distribution on the constrained sequences is first-order markovian if

$$Pr\{x_{t+1}|s_t, s_{t-1}, \dots\} = Pr\{x_{t+1}|s_t\} \quad (1.2)$$

all $s_t \in S$, $x_{t+1} \in A$, $t \in Z$ (Z is the set of integers, where s_t is the state of the machine at time t and x_{t+1} is the output produced at time $(t+1)$).

If probabilities $\{p_{ij}\}$ are assigned to the transition from state S_i to state S_j , with $p_{ij} \geq 0$, $\sum_{j=1}^J p_{ij} = 1$ all $i = 1, 2, \dots, J$, then the probability of the letter emitted at any time instant depends only on the previous state and the distribution on the allowed sequences will be first-order markovian.

We can also consider a K -th order markovian distribution on the channel input sequences so that each produced letter depends statistically only on the K previous states.

$$Pr\{x_{t+1}|s_t, s_{t-1}, \dots\} = Pr\{x_{t+1}|s_t, s_{t-1}, \dots, s_{t-K+1}\}$$

One can view the K -th order Markov distribution for a FSM as a one-step Markov distribution for an extended FSM, which also models the same input constraints. For a given FSM, consider the possible state sequences of length K and construct a new FSM with each state corresponding to to such a sequence. Let $S' = \{S'_1, S'_2, \dots, S'_L\}$ represent the state space of the new FSM, where $S'_i = (S_{i_1}, S_{i_2}, \dots, S_{i_K})$, $1 \leq i_k \leq J$. Assign an edge from $S'_i = (S_{i_1}, S_{i_2}, \dots, S_{i_K})$ to $S'_j = (S_{j_1}, S_{j_2}, \dots, S_{j_K})$ if $S_{j_l} = S_{i_{l+1}}$, $l = 1, \dots, (K - 1)$ and the letter emitted at that transition is the letter produced at the transition from S_{i_K} to S_{j_K} in the original FSM.

If a first-order markovian distribution is defined on the sequences generated by the new FSM, then by (1.2),

$$\begin{aligned} Pr\{x_{t+1}|s'_t, s'_{t-1}, \dots\} &= Pr\{x_{t+1}|s'_t\} \\ &= Pr\{x_{t+1}|s_{t_1}, \dots, s_{t_K}\} \end{aligned}$$

where $s'_t = (s_{t_1}, \dots, s_{t_K}) \in S'$ and $s_{t_i} \in S$, $i = 1, \dots, K$. The sets of sequences generated by both machines are identical. Hence, we obtain a K -th order markovian distribution on the sequences generated by the original FSM.

As an illustration, a two-step transition diagram for the channel constraint given in Figure 1.1 is shown in Figure 1.4. The set of states $S = \{S_0 = (00), S_1 = (01), S_2 = (10)\}$

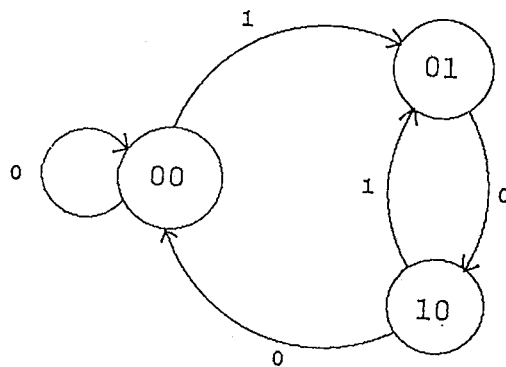


Figure 1.4: Double-step transition diagram for the channel constraint in Figure 1.1

includes all possible state sequences of length two that can be generated by the FSM in Figure 1.1. Letters produced at each state transition are assigned as labels to the edges. If we assign non-negative probabilities $\{p_{ij}\}$ on each directed branch with $p_{00} + p_{01} =$

1, $p_{12} = 1$, $p_{21} + p_{20} = 1$, then the distribution on the sequences generated by the FSM in Fig. 1.4 will be first-order markovian which corresponds to a second-order distribution on the sequences generated by the FSM in Fig. 1.1.

Note that the set of first-order markovian probability distributions on the constrained sequences is a subset of the set of second order distributions.

Our main contribution in this context is an answer to the question mentioned in section 1.4 stating that the input distribution maximizing the average mutual information is not first-order markovian. In particular, our lower and upper bounds on the maximum rate achievable by markovian input distribution show that using a second-order markovian distribution on the allowed input sequences for the input-restricted BSC for which constrained sequences are generated by the FSM in Fig. 1.1, higher rates can be achieved as compared with the first-order distribution. Hence, assuming a markovian distribution on the constrained input sequences of a noisy channel, we may achieve higher rates by considering higher order markovian distributions. Using the observation made in section 1.5, we may believe that second-order markovian distribution also does not achieve capacity, since it can be viewed as a first-order distribution for another FSM.

In the next chapter, our main results are presented and discussed. In chapter 3, conclusions and suggestions for future research are stated.

Chapter 2

CAPACITY OF NOISY, INPUT-CONSTRAINED CHANNELS

In section 2.1, upper and lower bounds on the maximum mutual information achieved by markovian input distribution are derived. The results obtained by using these bounds on the capacity of binary symmetric channel (BSC) and binary erasure channel (BEC) with the input constraint given in Fig.1.1 are tabulated and discussed in section 2.2. In section 2.3, the upper bounds C_N computed by using Arimoto-Blahut algorithm for the same channels are given both tabularly and graphically. In the final section, a new problem dealing with the convergence in the capacity is discussed.

2.1 Upper and Lower Bounds on the Maximum Rate Achievable by Markovian Distribution

By (1.1), the capacity of an input-restricted, discrete channel is given by

$$C \triangleq \lim_{N \rightarrow \infty} \sup_{P_N(X^N)} \frac{I(X^N; Y^N)}{N}$$

where for each N , the supremum is taken over all distributions $\{P_N(X^N)\}$ on the constrained channel input sequences of length N . Since the set of such distributions on the constraint set is compact and the average mutual information function is continuous in $P_N(X^N)$, the supremum in (1.1) can be replaced by maximum [6, p.49].

In this work, we ask whether a first order markovian distribution achieves C . The results clarify that it does not. We prove this by showing that for the channel in Fig.1.1, a second order markovian distribution does strictly better than a first order markovian distribution. We define

$$R_K \triangleq \lim_{N \rightarrow \infty} \max_{P_N(X^N)} \frac{I(X^N; Y^N)}{N} \quad (2.1)$$

where the maximum is taken over all K -th order markovian distributions on the constrained channel input sequences that are generated by a FSM. In particular, we will show that $R_1 < R_2$ for the channel of Fig.1.1.

To show this, we derive upper and lower bounds on R_K . Assume that the channel input distribution is stationary (when constrained input sequences are generated by a FSM with an irreducible and aperiodic set of states and the probability distribution on the sequences is markovian, then starting infinitely far in the past, the distribution becomes stationary since ergodicity implies the existence of limiting state probability distribution). When the channel input distribution is stationary,

$$\begin{aligned} P(y_n = \beta_n, \dots, y_{n+k} = \beta_{n+k}) &= \sum_{\alpha_n, \dots, \alpha_{n+k}} P(x_n = \alpha_n, \dots, x_{n+k} = \alpha_{n+k}) \prod_{i=n}^{n+k} P(y_i = \beta_i | x_i = \alpha_i) \\ &= \sum_{\alpha_n, \dots, \alpha_{n+k}} P(x_0 = \alpha_n, \dots, x_k = \alpha_{n+k}) \prod_{i=0}^k P(y_i = \beta_{n+i} | x_i = \alpha_{n+i}) \\ &= P(y_0 = \beta_n, \dots, y_k = \beta_{n+k}) \end{aligned}$$

for all integers n, k and channel output sequences $(\beta_n, \dots, \beta_{n+k})$ of length $(k+1)$, where the second equality is obtained by using the stationarity of the input distribution and the discrete memoryless channel assumption. Therefore, the channel output distribution is also stationary.

We can write

$$I(X^N; Y^N) = H(Y^N) - H(Y^N | X^N)$$

Define $H_\infty(Y)$ as $\lim_{N \rightarrow \infty} \frac{H(Y^N)}{N}$ and by Theorem 3.5.1 of [2]

$$H_\infty(Y) = \lim_{N \rightarrow \infty} H(Y_N | Y_{N-1}, \dots, Y_1) \quad (2.2)$$

On the other hand,

$$\frac{H(Y^N | X^N)}{N} = \frac{1}{N} E \left\{ \log \frac{1}{P_N(Y^N | X^N)} \right\} = \frac{1}{N} \sum_{n=1}^N H(Y_n | X_n) \quad (2.3)$$

where the second equality follows since the channel is memoryless.

$$H(Y_n | X_n) = \sum_{\alpha \in A, \beta \in B} P(y_n = \beta | x_n = \alpha) P(x_n = \alpha) \log \frac{1}{P(y_n = \beta | x_n = \alpha)}$$

Since the input distribution is stationary, for any n

$$P(x_n = \alpha) = P(x = \alpha) \quad (2.4)$$

and since the channel is memoryless

$$P(y_n = \beta | x_n = \alpha) = P(y = \beta | x = \alpha) \quad (2.5)$$

Then, by using (2.4) and (2.5), we obtain

$$H(Y^N | X^N) = N H(Y | X) \quad (2.6)$$

where $H(Y | X)$ is given by

$$H(Y | X) = \sum_{\alpha, \beta} P(y = \beta | x = \alpha) P(x = \alpha) \log \frac{1}{P(y = \beta | x = \alpha)} \quad (2.7)$$

By (2.2) and (2.6), (2.1) becomes

$$\begin{aligned} R_K &= \lim_{N \rightarrow \infty} \max_{P_N(X^N)} [H(Y_N | Y_{N-1}, \dots, Y_1) - H(Y | X)] \\ &= \lim_{N \rightarrow \infty} \max_{P_N(X^N)} [H(Y_0 | Y_{-1}, \dots, Y_{-N}) - H(Y_0 | X_0)] \end{aligned} \quad (2.8)$$

where the second equality follows from the stationarity of the channel output.

Proposition 1 *The maximum rate achievable by a K -th order markovian distribution on the input sequences of an input-restricted channel satisfies*

$$(i) \ R_K \leq \max_{P_N(X^N)} [H(Y_0 | Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0 | X_0)]$$

$$(ii) \ R_K \geq \max_{P_{L+K}(X^{L+K})} [H(Y_0 | Y_{-1}, \dots, Y_{-(L-1)}, X_{-L}, \dots, X_{-(L+K-1)}) - H(Y_0 | X_0)]$$

where the maximums are taken over all K -th order markovian probability distributions on the allowed input sequences.

Proof:

(i)

$$\begin{aligned} \max_{P_N(X^N)} [H(Y_0 | Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0 | X_0)] \\ \leq \max_{P_M(X^M)} [H(Y_0 | Y_{-1}, \dots, Y_{-(M-1)}) - H(Y_0 | X_0)] \end{aligned} \quad (2.9)$$

for integers N, M ; $M < N$, where the first inequality follows from the fact that conditioning can not increase entropy. Hence, from (2.8) and (2.9) we can write an upper bound for capacity

$$R_K \leq \max_{P_N(X^N)} [H(Y_0 | Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0 | X_0)] \quad (2.10)$$

for any positive integer N , where the maximum is taken over all markovian probabilities $\{P_N(X^N)\}$ on the sequences $\{X^N\}$ satisfying the input constraints.

(ii) Suppose the probability distribution on the constrained channel input sequences is K -th order markovian. Since conditioning can not increase entropy,

$$\begin{aligned} \lim_{N \rightarrow \infty} H(Y_0|Y_{-1}, Y_{-2}, \dots, Y_{-(N-1)}) &\geq \lim_{N \rightarrow \infty} H(Y_0|X_{-L}, \dots, X_{-(L+K-1)} Y_{-1}, \dots, Y_{-(N-1)}) \\ &= H(Y_0|Y_{-1}, \dots, Y_{-(L-1)} X_{-L}, \dots, X_{-(L+K-1)}) \end{aligned} \quad (2.11)$$

for any positive integer L , where the equality follows from the fact that the channel is memoryless and input distribution is K -th order markovian.

$$\begin{aligned} \max_{P_{L+K}(X^{L+K})} [H(Y_0|Y_{-1}, Y_{-2}, \dots, Y_{-(L-1)} X_{-L}, \dots, X_{-(L+K-1)}) - H(Y_0|X_0)] \\ \leq \lim_{N \rightarrow \infty} \max_{P_N(X^N)} [H(Y_0|Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0|X_0)] \\ = R_K \end{aligned} \quad (2.12)$$

the inequality follows from (2.11), where the maximums are taken over all K -th order markovian input probability distributions $\square$.

Thus, combining the two bounds (2.10) and (2.12), for any two integers N and L greater than 1, we obtain

$$\begin{aligned} \max_{P_{L+K}(X^{L+K})} [H(Y_0|Y_{-1}, \dots, Y_{-(L-1)} X_{-L}, \dots, X_{-(L+K-1)}) - H(Y_0|X_0)] &\leq R_K \\ &\leq \max_{P_N(X^N)} [H(Y_0|Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0|X_0)] \end{aligned} \quad (2.13)$$

These bounds are computable for finite N and L . As K increases, i.e., the order of the markovian distribution increases, the lower bound will become closer to channel capacity. In general, as K , N , L approach infinity the two inequalities become equalities. However, even for moderately large parameters K , N and L , these computations may become impractical because of the extremely long run-times.

2.2 Computational Results

In this section, we present the computational results obtained by using (2.13) for the input constraint given in Figure 1.1 on BSC and BEC. The bounds in (2.13) are computed for two different input distributions: first-order and second-order markovian. The parameters are taken as $N=7$ and $L=3$. In Tables 2.1 and 2.2, R_1^u and R_1^l denote the upper and lower bounds respectively on R_1 , whereas R_2^u and R_2^l denote the same bounds for R_2 . All bounds are in the unit of nats per channel uses. ϵ denotes the cross-over probability of BSC and the erasure probability for the BEC.

We define

$$\delta_{min} = \frac{R_2^l - R_1^u}{R_1^u} 100$$

This corresponds to a lower bound on the incremental improvement in percentage gained by using second order markovian distribution on the input sequences of a constrained channel. Tables 2.1 and 2.2 also list the δ_{min} values.

The results show that for both channels, when the channel is noisy, the input distribution achieving the capacity is not first-order markovian, since $R_1 \leq R_1^u < R_2^l \leq R_2 \leq C$. By considering higher order distributions on the allowed channel input sequences, it seems that further improvements in the maximum rate can be obtained.

We know that for the noiseless channel, first-order markovian distribution achieves capacity. When the channel introduces a low level of noise, i.e., ϵ in BSC and BEC is close to zero, δ_{min} is quite small. As the noise level increases, we can gain more in rate by increasing the order. At the extreme point, $\epsilon=0.5$ in BSC, we know that the capacity is zero and any probability distribution achieves this rate. Hence, at those values of ϵ near 0.5 we obtain small gains by considering second-order markovian distribution on the channel input sequences. For BSC, higher improvements are obtained for intermediate values of ϵ between 0 and 0.5. Similar comments can be made for constrained BEC as ϵ goes to 1.

A situation arising in BSC for $\epsilon = 0.4, 0.45, 0.49$ is that the upper bound on the capacity for the first-order markovian input distribution exceeds the lower bound for the second-order distribution. The same situation occurs also for constrained BEC with $\epsilon=0.95$. This situation is related to the tightness of the bounds. For higher noise levels, knowledge about a particular input letter gives more information about a later channel output letter; thus the entropy of the output letter decreases considerably when conditioned on a previous input letter. The lower bound on the maximum achievable rate by markovian input distribution of constrained BSC is not tight for values of ϵ near 0.5 (same situation is valid for BEC with $\epsilon=1$).

This observation may explain the above situation. We can still expect that second-order markovian distribution improves the rate by examining the upper bounds on these values of ϵ . By increasing the parameter L , it is possible to obtain tighter lower bounds and in this case we can have a more informative answer to this problem.

Another comparison can be made between results of BSC and BEC when ϵ is between 0 and 0.3. The output of a BEC resembles the input more as compared with BSC having a cross-over probability equal to the erasure probability of BEC, since observing a non-erasure letter at the output, we are sure about the corresponding input letter. Therefore, the perturbation on the input sequences of a BEC can be thought to be less, compared with BSC having the same cross-over probability. Consequently, optimal one-step distribution achieves rates closer to the capacity for BEC. As ϵ goes to 0.5, the capacity of constrained BSC approaches zero and improvements become smaller, whereas for constrained BEC

there are improvements for ϵ near 0.5 (capacity of BEC goes to zero as ϵ approaches one).

ϵ	R_1^l	R_1^u	R_2^l	R_2^u	δ_{min}
0.01	0.447914	0.447930	0.447952	0.447968	0.005
0.05	0.359617	0.359805	0.360047	0.360215	0.067
0.1	0.277278	0.277748	0.278378	0.278749	0.227
0.2	0.154919	0.156008	0.157079	0.157756	0.687
0.3	0.069924	0.071511	0.072128	0.072840	0.863
0.4	0.017769	0.018854	0.018701	0.019149	-
0.45	0.004463	0.004871	0.004729	0.004895	-
0.49	0.000179	0.000199	0.000189	0.000199	-

Table 2.1: Computed bounds for BSC

ϵ	R_1^l	R_1^u	R_2^l	R_2^u	δ_{min}
0.01	0.477374	0.477374	0.477375	0.477375	$2.10 \cdot 10^{-4}$
0.05	0.461891	0.461892	0.461915	0.461915	$4.98 \cdot 10^{-3}$
0.1	0.442237	0.442238	0.442329	0.442331	0.021
0.2	0.401858	0.401871	0.402206	0.402217	0.083
0.3	0.359922	0.359973	0.360650	0.360668	0.188
0.4	0.316246	0.316380	0.317426	0.317521	0.331
0.5	0.270600	0.270889	0.272244	0.272435	0.500
0.6	0.222699	0.223244	0.224758	0.225081	0.678
0.7	0.172185	0.173100	0.174532	0.175007	0.827
0.8	0.118609	0.119957	0.120990	0.121566	0.861
0.9	0.061426	0.062978	0.063280	0.063925	0.480
0.95	0.031286	0.032542	0.032457	0.032698	-

Table 2.2: Computed bounds for BEC

2.3 Upper bounds on C

The upper bounds $\{C_N\}$ on the capacity of BSC and BEC with the input constraint given in Fig. 1.1 are tabulated in Tables 2.3 and 2.4. In Tables 2.3 and 2.4, for each channel the upper capacity C_N is provided for the largest value of N that could be computed. The complete list of upper bounds C_N is given in Appendix B. In Figures (2.1)-(2.6), C_N vs. N are plotted for various channels considered in this work for giving a visual inspection to the reader.

These values are computed by using Arimoto-Blahut algorithm with an error term 10^{-8} . The error term is defined as the difference between results of two consecutive iterations of the algorithm. The number of iterations performed in order to achieve the desired error term are noted.

The bounds may give some idea about how close markovian distributions on the channel input sequences get to capacity. Obtaining closer bounds by computing C_N for larger values of N was not possible due to extremely long run-times and huge memory requirements.

By examining results, one can conclude that assuming markovian distribution on the channel input sequences it is possible to achieve rates near capacity for the channels considered in this work. However, for a more complicated channel the situation may be quite different so that it is possible to achieve considerably much higher rates by using nonmarkovian distributions as compared with the rates achievable by markovian distributions.

ϵ	C_N	N	#of iter.
0.01	0.458085	14	6
0.05	0.367535	14	13
0.1	0.284136	13	55
0.2	0.160613	11	1009
0.3	0.073944	9	3182
0.4	0.019271	9	2891
0.45	0.004921	8	35010
0.49	0.000199	7	207862

Table 2.3: Upper bounds computed by Arimoto-Blahut algorithm(BSC)

ϵ	C_N	N	#of iter.
0.1	0.456009	10	8
0.2	0.413863	10	13
0.3	0.370379	10	26

Table 2.4: Upper bounds computed by Arimoto-Blahut algorithm(BEC)

2.4 Rate of Convergence of C_N to the Capacity

In this section, we will discuss the rate of the convergence C_N to the capacity as N goes to infinity, where

$$C_N = \max_{P_N(X^N)} \frac{I(X^N; Y^N)}{N}$$

Using (A.5), we can write

$$NC_N \leq (N-1)C_{N-1} + C_1$$

Rearranging the terms, we obtain

$$C_{N-1} - C_N \geq \frac{C_{N-1} - C_1}{N}$$

Therefore, the difference between consecutive terms of the sequence $\{C_N\}$ goes to zero at fastest reciprocally in N as N goes to infinity. Since we use Arimoto-Blahut algorithm for computing C_N , which has a computation complexity that is exponential in N , the rate of convergence of $\{C_N\}$ is quite small. In Appendix B, we give the complete list of $\{C_N\}$ computed for the channels considered in this work.

A computationally more efficient way of getting closer to C may be to assume a markovian distribution on the allowed channel input sequences and to find a practical method for computing R_K . By increasing the memory of the input distribution, we can get better estimates of C . The problem is the rate of convergence to R_K as one considers (2.1). We think that the convergence in (2.8) to R_K is achieved exponentially, i.e., the difference between two consecutive terms goes exponentially to zero as N tends to infinity.

As shown in section 2.1, the maximum mutual information of a DMC under input constraints achieved by K -th order stationary markovian distribution can be written in the form (2.8):

$$R_K = \lim_{N \rightarrow \infty} \max_{P_N(X^N)} [H(Y_0|Y_{-1}, Y_{-2}, \dots, Y_{-N}) - H(Y|X)]$$

We investigate the difference Δ_N , between two consecutive terms as we approach the limit in (2.8).

$$\Delta_N = H(Y_0|Y_{-1}, \dots, Y_{-(N-1)}) - H(Y_0|Y_{-1}, \dots, Y_{-N}) \quad (2.14)$$

Since conditioning can not increase entropy, $\Delta_N \geq 0$. Furthermore, convergence in the capacity assures that $\lim_{N \rightarrow \infty} \Delta_N = 0$. We can write (2.14) by using (2.2.26) of [2] in the form

$$\Delta_N = I(Y_0; Y_{-N} | Y_{-1}, \dots, Y_{-(N-1)}) \quad (2.15)$$

We can upperbound Δ_N as

$$\Delta_N \leq I(X_0; X_{-N} | Y_{-1}, \dots, Y_{-(N-1)})$$

If the rate of convergence in (2.8) is exponential, then Δ_N can be upperbounded with a function which goes to zero exponentially in N as $N \rightarrow \infty$. When the channel output sequences are K -th order markovian, $\Delta_N = 0$ for $N \geq K + 1$. However, for an arbitrary distribution on Y^N , Δ_N may be positive for all N . We have proved that when the channel input distribution is ergodic and markovian $I(Y_0; Y_{-N})$ goes to zero exponentially in N as N tends to infinity. There is no general equality or inequality condition between $I(Y_0; Y_{-N} | Y_{-1}, \dots, Y_{-(N-1)})$ and $I(Y_0; Y_{-N})$. However, in our case the sequence Y^N is not completely arbitrary. The sequences correspond to the output sequences of a DMC when the probability distribution on channel input sequences is markovian. For distributions satisfying the condition that $I(Y_0; Y_{-N})$ is larger than $I(Y_0; Y_{-N} | Y_{-1}, \dots, Y_{-(N-1)})$, the desired result is shown.

If this statement is true, the upper bound given in (2.13) will be exponentially tight and hence, we can get a good estimate of maximum achievable rate with a markovian distribution on the constrained input sequences. And we can use (2.13) for estimating R_K , since the convergence will be exponential, we can get close to R_K even for small values of N . Even the optimal distribution is not markovian in the noisy case, we can still find a good approximation to the capacity by assuming higher order markovian distributions on the channel input sequences.

The proof or disproof of this exponential convergence remains as an open problem.

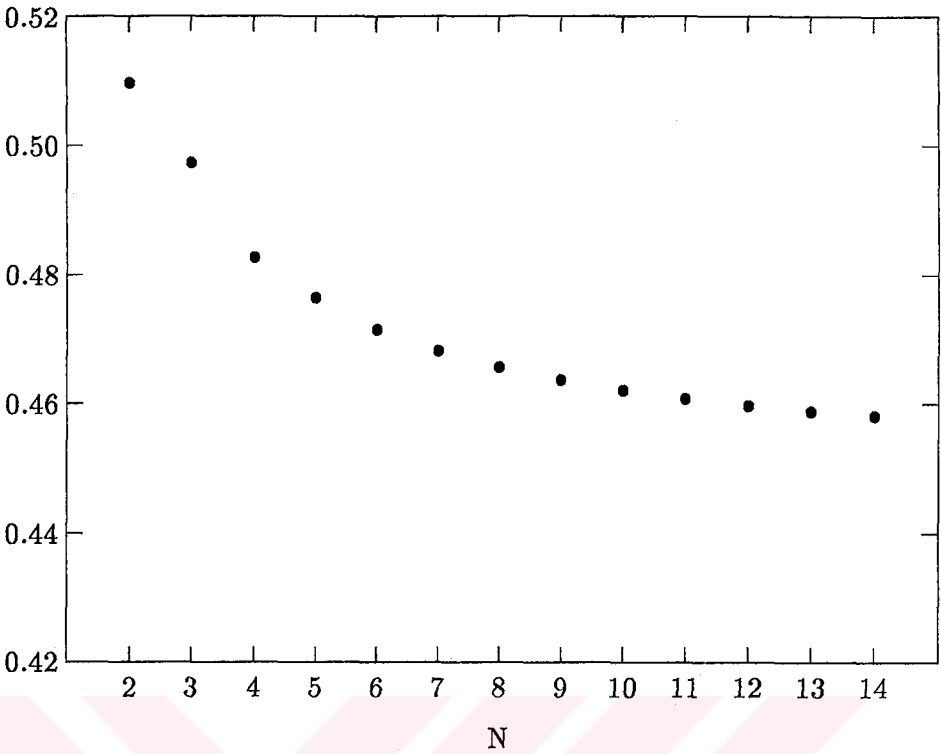


Figure 2.1: C_N vs. N for constrained BSC with $\epsilon=0.01$

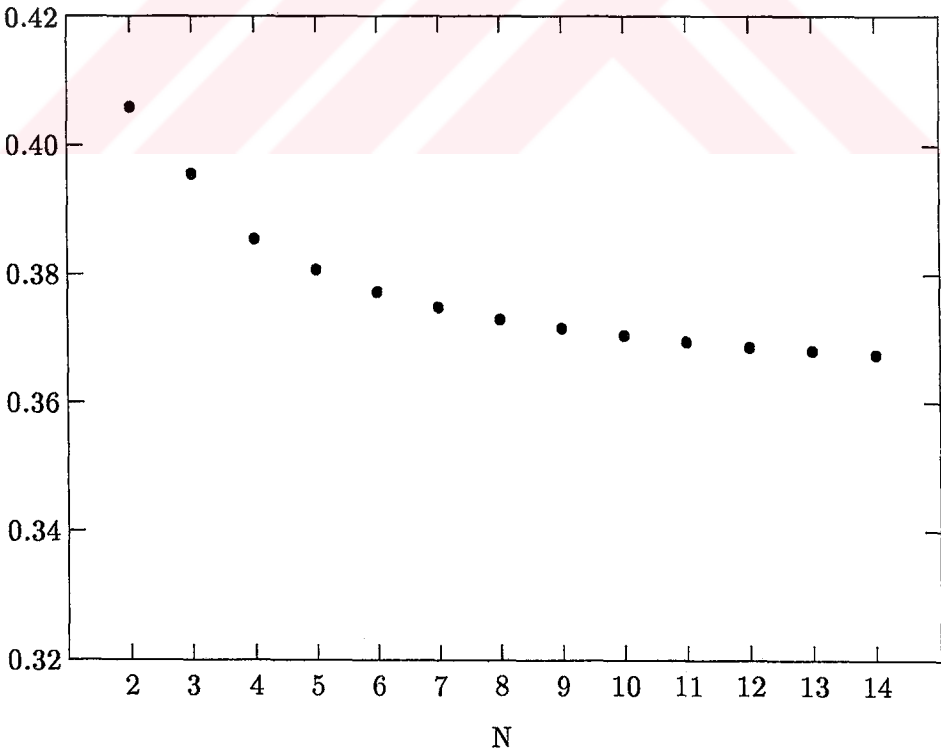
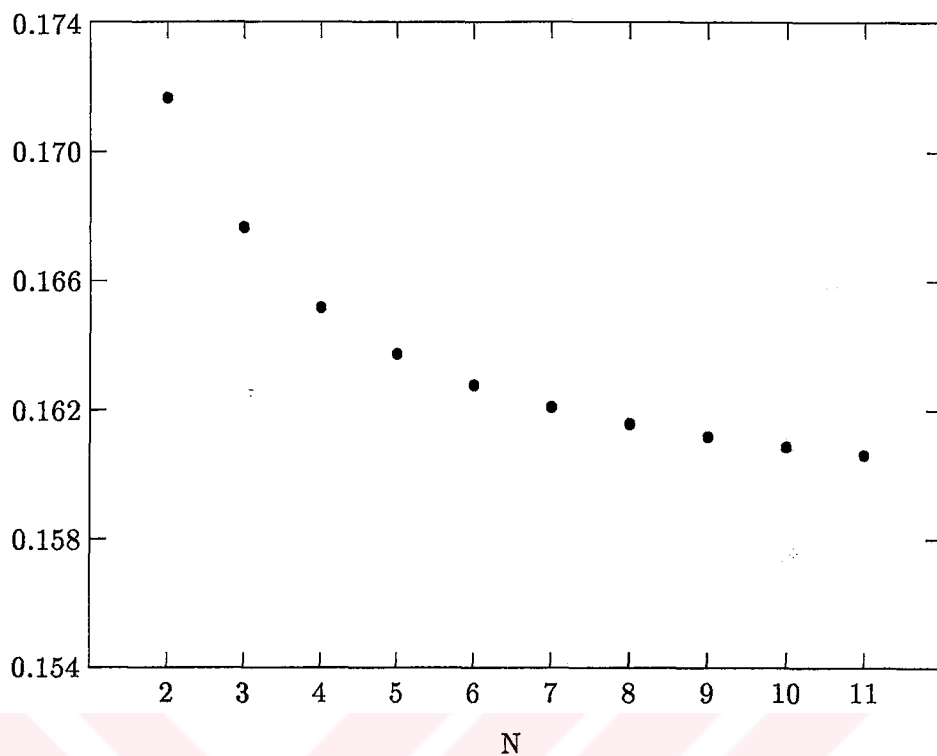
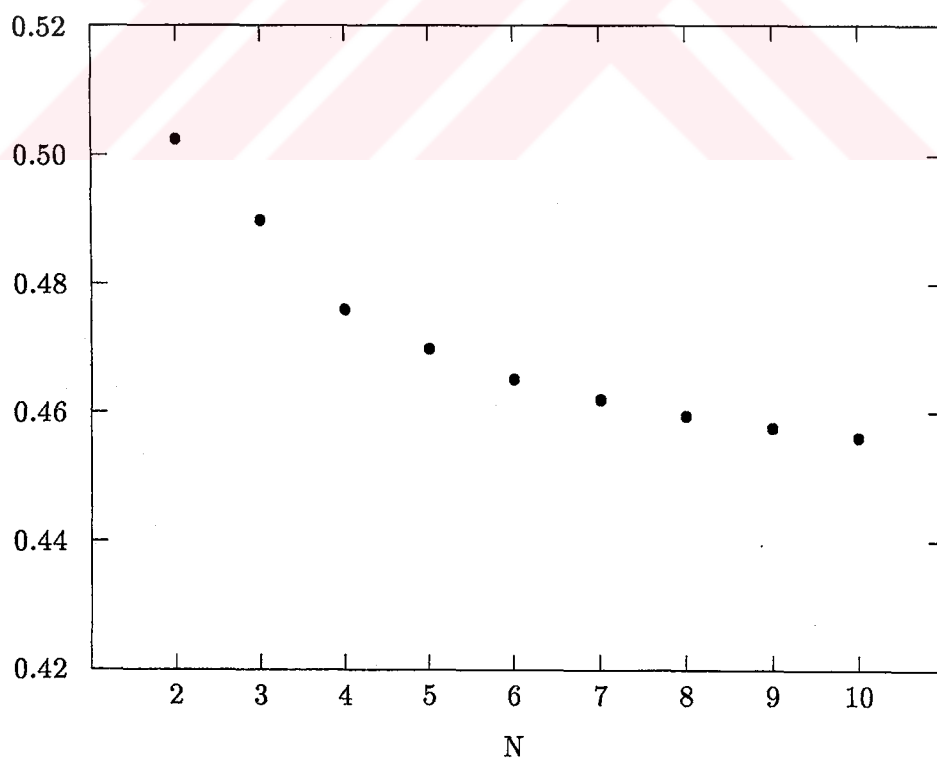


Figure 2.2: C_N vs. N for constrained BSC with $\epsilon=0.05$

Figure 2.3: C_N vs. N for constrained BSC with $\epsilon=0.1$ Figure 2.4: C_N vs. N for constrained BEC with $\epsilon=0.1$

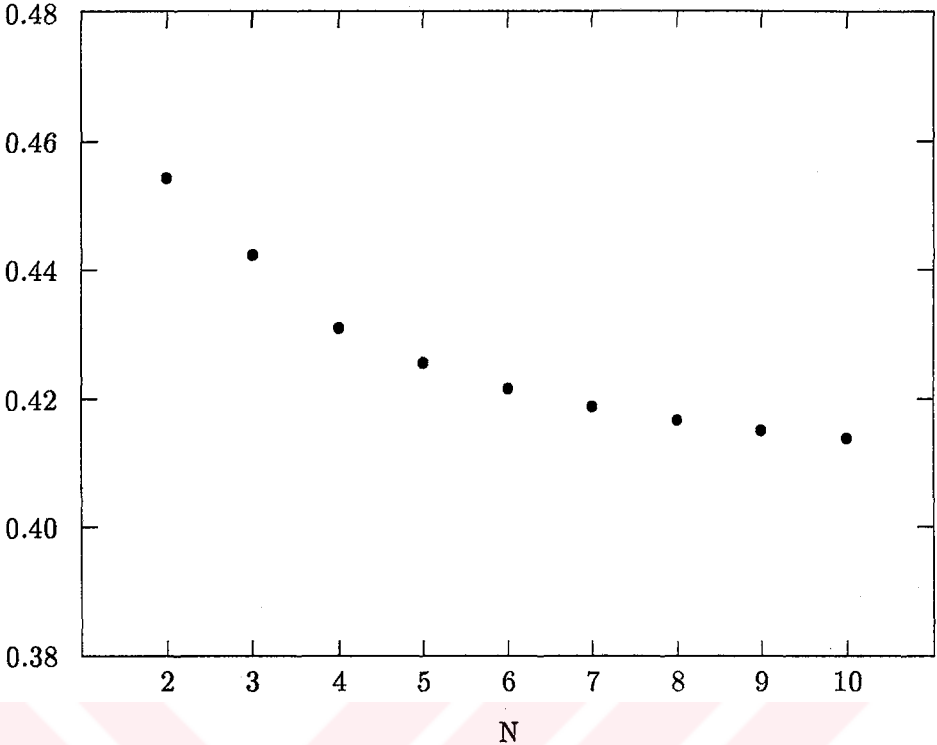


Figure 2.5: C_N vs. N for constrained BEC with $\epsilon=0.2$

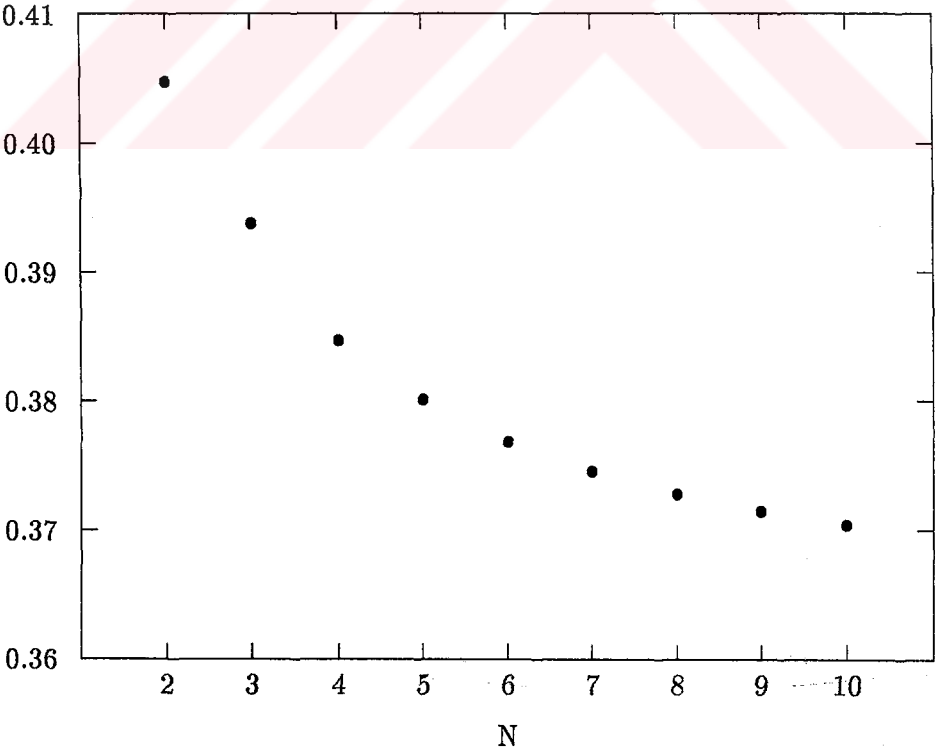


Figure 2.6: C_N vs. N for constrained BEC with $\epsilon=0.3$

Chapter 3

CONCLUSIONS

In this work, we were mainly concerned with the capacity of noisy, input-restricted, discrete, memoryless channels. No efficient algorithm is known for computing the capacity of such channels. In particular, we considered BSC and BEC with the input restriction given in Fig. 1.1. Our main result is that, unlike the noiseless case, one-step Markov distribution on the channel input alphabet does not achieve the channel capacity when noise is present. We showed that a second-order markovian distribution is strictly better than a first order distribution for the BSC and BEC with an input constraint in Fig.1.1.

The amount of incremental improvement in maximum rate by increasing order is also an interesting problem. Code designers search for better codes that achieve higher rates closer to channel capacity. Information about the input distributions on the constrained channel input sequences achieving higher rates is required for finding such codes. Our results show that one may have to use higher order markovian distributions in order to achieve higher rates. However, increasing the order results in an increase in the code complexity. Hence, there is a trade-off between rate and complexity. Future research may be directed at finding bounds on the incremental improvement in the maximum mutual information as one considers higher order markovian distributions on the input sequences of the constrained channel. Such a bound will give an information to the code designer about the trade-off between achievable rate and implementation complexity.

Appendix A

Proof that the limit in (1.1) exists

In this part, we present a proof of the existence of the limit in (1.1), i.e., we show that the capacity of a DMC (with an arbitrary input constraint) defined by (1.1) exists.

$$C = \lim_{N \rightarrow \infty} C_N$$

where

$$C_N = \max_{Q_N(X^N)} \frac{I(X^N; Y^N)}{N}$$

and the maximum is taken over the probability distributions on the channel input sequences satisfying channel input constraints. The following is a proof of the limit in (1.1) exists.

Let Q_N^* be the input distribution that achieves C_N . Let n and l be two positive integers with $n + l = N$ and let X_1 and X_2 be the ensembles of input sequences $\underline{x}_1 = (x_1, \dots, x_n)$ and $\underline{x}_2 = (x_{n+1}, \dots, x_N)$ and let Y_1 and Y_2 be the resulting channel output ensembles.

$$NC_N = I_{Q_N^*}(X_1 X_2; Y_1 Y_2) = I_{Q_N^*}(X_1; Y_1) + I_{Q_N^*}(X_2; Y_1 | X_1) + I_{Q_N^*}(X_1 X_2; Y_2 | Y_1) \quad (\text{A.1})$$

The first term in (A.1) satisfies

$$I_{Q_N^*}(X_1; Y_1) \leq nC_n \quad (\text{A.2})$$

Since conditional on X_1 , X_2 and Y_1 are independent,

$$I_{Q_N^*}(X_2; Y_1 | X_1) = 0 \quad (\text{A.3})$$

$$\begin{aligned} I_{Q_N^*}(X_1 X_2; Y_2 | Y_1) &= H(Y_2 | Y_1) - H(Y_2 | X_1 X_2 Y_1) \\ &\leq H(Y_2) - H(Y_2 | X_2) \\ &= I_{Q_N^*}(Y_2; X_2) \leq lC_l \end{aligned} \quad (\text{A.4})$$

where the first inequality follows from the fact that unconditioning cannot decrease entropy and conditional on X_2 , Y_2 is independent of X_1 and Y_1 . Combining (A.1), (A.2), (A.3) and (A.4), we obtain

$$NC_N \leq nC_n + lC_l \quad (\text{A.5})$$

Applying Lemma 4A.2 of [2, p.112] for the bounded sequence $\{C_N\}$ satisfying (A.5), it follows that C exists and is given by

$$C = \lim_{N \rightarrow \infty} C_N = \inf_N C_N \square. \quad (\text{A.6})$$

Clearly, $C_N \geq C$ for all N . Although the sequence $\{C_N\}$ is not monotonically decreasing, an immediate consequence of (A.5) is

$$C_N \geq C_{2N}$$

Appendix B

List of Upper Bounds on the Capacity

In this appendix, the complete list of upper bounds $\{C_N\}$ on the capacity of the input-restricted channel, with a constraint represented by the FSM in Fig.1.1. These bounds are computed by using Arimoto- Blahut algorithm.

ϵ	C_N	N	#of iter.
0.01	0.509731	2	5
0.01	0.497379	3	5
0.01	0.482767	4	6
0.01	0.476476	5	6
0.01	0.471519	6	6
0.01	0.468220	7	6
0.01	0.465667	8	6
0.01	0.463707	9	6
0.01	0.462131	10	6
0.01	0.460844	11	6
0.01	0.459771	12	6
0.01	0.458883	13	6
0.01	0.458085	14	6
0.05	0.405988	2	9
0.05	0.395571	3	10
0.05	0.385527	4	11
0.05	0.389788	5	10
0.05	0.377270	6	11
0.05	0.374859	7	12
0.05	0.373022	8	9
0.05	0.371601	9	12
0.05	0.370462	10	13
0.05	0.369531	11	13
0.05	0.368755	12	13
0.05	0.368098	13	13
0.05	0.367535	14	13

Table B.1: Upper bounds computed by Arimoto-Blahut algorithm(BSC)

ϵ	C_N	N	#of iter.
0.1	0.310650	2	11
0.1	0.302569	3	21
0.1	0.296059	4	29
0.1	0.292736	5	37
0.1	0.290377	6	43
0.1	0.288729	7	47
0.1	0.287423	8	49
0.1	0.286516	9	51
0.1	0.285743	10	52
0.1	0.285110	11	53
0.1	0.284582	12	54
0.1	0.284136	13	55
0.3	0.077271	2	160
0.3	0.075849	3	691
0.3	0.075173	4	3241
0.3	0.074723	5	2546
0.3	0.074429	6	3029
0.3	0.074222	7	3145
0.3	0.074066	8	3159
0.3	0.073944	9	3182
0.49	0.000200	2	107583
0.49	0.000200	3	150464
0.49	0.000200	4	174862
0.49	0.000200	5	190908
0.49	0.000199	6	201427
0.49	0.000199	7	207862

Table B.2: Upper bounds computed by Arimoto-Blahut algorithm(BSC)

ϵ	C_N	N	#of iter.
0.2	0.171696	2	41
0.2	0.167674	3	128
0.2	0.165174	4	2264
0.2	0.163748	5	649
0.2	0.162790	6	1013
0.2	0.162106	7	981
0.2	0.161593	8	994
0.2	0.161193	9	1000
0.2	0.160874	10	1005
0.2	0.160613	11	1009
0.4	0.019764	2	1330
0.4	0.019527	3	2051
0.4	0.019418	4	2003
0.4	0.019367	5	6591
0.4	0.019331	6	2406
0.4	0.019305	7	2778
0.4	0.019286	8	2525
0.4	0.019271	9	2891
0.45	0.004984	2	15441
0.45	0.004962	3	15707
0.45	0.004946	4	14243
0.45	0.004935	5	27628
0.45	0.004928	6	27547
0.45	0.004924	7	57441
0.45	0.004921	8	35010

Table B.3: Upper bounds computed by Arimoto-Blahut algorithm(BSC)

ϵ	C_N	N	#of iter.
0.1	0.502475	2	6
0.1	0.489871	3	7
0.1	0.476003	4	7
0.1	0.469837	5	8
0.1	0.465084	6	8
0.1	0.461885	7	8
0.1	0.459425	8	8
0.1	0.457531	9	8
0.1	0.456009	10	8
0.2	0.454333	2	9
0.2	0.442372	3	10
0.2	0.430993	4	10
0.2	0.425596	5	9
0.2	0.421601	6	9
0.2	0.418860	7	12
0.2	0.416772	8	12
0.2	0.415158	9	12
0.2	0.413863	10	13
0.3	0.404763	2	11
0.3	0.393825	3	15
0.3	0.384691	4	19
0.3	0.380102	5	22
0.3	0.376816	6	24
0.3	0.374529	7	24
0.3	0.372797	8	25
0.3	0.371454	9	26
0.3	0.370379	10	26

Table B.4: Upper bounds computed by Arimoto-Blahut algorithm(BEC)

REFERENCES

- [1] Shannon, C.E., "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol.27, pp.379-423(part 1), 1948.
- [2] Gallager, R.G., *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [3] Zehavi, E. and Wolf, J.K., "On runlength codes," *IT* 34(1988), pp.45-54.
- [4] Arimoto, S., "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IT* 18 (1972), pp.14-20.
- [5] Blahut, R., "Computation of channel capacity and rate-distortion functions," *IT* 18 (1972), pp.460-473.
- [6] McEliece, R.J., *The theory of information and coding*. Massachusetts: Addison-Wesley.

T. C.
Yükseköğretim Kurulu
Dokümantasyon Merkezi